

博士論文

構造変化を伴う知識領域における
イノベーション・マネジメントの
ための予測手法に関する研究

佐々木 一

概要

科学技術への適切な投資は、技術志向型の企業や各国政府にとって将来の競争優位性を獲得するための重要な意思決定である。歴史を遡れば人類の営みにおいて科学技術が活用されはじめて久しいが、情報化社会以降の科学技術情報量は指数関数的に増加しており、いまや人類が有する科学技術知識の大半を占めつつある。また科学技術が社会に導入されるまでのライフサイクルがますます早まってきており、多様化する社会需要による影響など、生産される知識の動的変化が著しい。知識構造が大きく変化し、社会や産業構造との相互影響が拡大するなかで、企業などの経済主体が早期に有望な科学技術知識を獲得することや、自社の知的資源や研究開発戦略と相性の良い科学技術分野を予測することや、さらにそれらの知識を用いて未踏の領域へ適切に進出することは難しくなってきた。

これまでも、サイエンスマップやパテントマップと呼ばれる科学技術知識を把握するためのツールが提案・活用され、それらによる網羅的な情報が重要であるという指摘はされてきた。しかしながら、研究開発戦略や科学技術イノベーション政策において将来動向の予測に寄与する情報の需要が高まると同時にそれらの一次情報が持つ複雑性も増しており、既存の手法のみでは対応が困難となってきた。大規模データ処理技術や機械学習などの手法の高度化は、それらの要求に応えることができると期待されている。特に機械学習による有益な点は、多様な特徴量により高い予測精度を実現することができる点にある。あるいは深層学習などを活用することでより抽象化された特徴量によって高い精度を実現できることが多くの領域において確認され始めている。しかしながら一方で、イノベーション・マネジメントに代表されるような、人間が意思決定をするための情報・分析が必要とされる場面では、解釈が困難な特徴量だけでは高い精度を実現できたとしても適切なアクションにつなぐににくいという懸念がある。

本研究では、イノベーション・マネジメントに代表されるような意思決定を前提とした予測モデルのあり方を議論する。特に予測精度のみならず、特徴量の解釈可能性を含有した分析の有効性を提案する。そのためのひとつの手段として、科学技術情報をもつデータ構造から分析対象に応じた本質的な情報を抽出することの有効性を提案する。通常、科学技術関連データは、それぞれ特定のデータ型で収集、格納、管理がされているが、データの構造情報に着目し特徴量を抽出した例は少ない。これらの構造情報はドメインの専門知識や目的の本質次第では有用な情報が得られる。本研究では専門家の知見を特徴量に落とす有効性を主張するとともにその中でもこれらのデータの構造情報から分析の目的に本質的な特徴量が得られると考えた。

本論文では、科学技術イノベーション戦略における意思決定に資する情報として、科学レイヤー、技術レイヤー、産業レイヤーのそれぞれのレイ

ヤーが有する大規模データを活用しイノベーションプロセスに必要な段階について包括的に議論する。これらのプロセスではそれぞれ予測として有効な視点が異なることから、本研究では次の3点に着目する。

- 1. 科学における有望な知識の予測
- 2. 技術知識同士の融合の予測
- 3. 知識を社会に実装する主体の行動の予測

本研究では科学レイヤーとして対象とする分野として、ナノカーボン分野、太陽光発電分野に関する学術領域に焦点を当てている。これらは既に豊富な研究がなされた分野であるとともに同時にサイエンスリンクエージの高い分野である。また、技術レイヤーの対象としては、炭素繊維強化プラスチック分野、機能性材料分野を対象とする。これらは我が国の貢献の高い分野であり、用途汎用性が高く分野横断的な活用が期待できる分野である。また、産業レイヤーにおける分析としては、経済主体である企業同士の取引関係に注目した予測を対象とする。

本研究の結果としては、ナノカーボン分野や太陽光発電分野において、これまでの萌芽予測手法では十分に議論のされてこなかった有望知識を領域の単位で抽出することの有効性を示した。また技術知識同士の有望な融合においては、特許情報が有する階層情報が技術融合の予測に寄与することを示した。経済主体同士の関係性の予測では、業種間の資材の流れを確率として取り込むことによって、予測精度並びに解釈可能性の両立性がはかれることを示すことができた。

結果から得られた本研究の貢献は次の三つである。第一に、解釈可能性を含有したうえで精度を実現することを示すことの重要性を示したことにある。また、本研究の第二の貢献としては、科学技術情報をもつデータ構造から分析対象に応じた本質的な情報を抽出することの有効性を示したことにある。第三の貢献として、科学、技術、産業それぞれの予測手法を統合することで、企業をはじめとする経済主体が未踏領域へ進出する際のイノベーション・マネジメント支援のフレームワークを提案した。本研究で提案するこれらの貢献によって、動的に変化する知識に対応した情報の獲得が可能となる。本研究では、動的に変化する知識領域への総合的なイノベーションマネジメントにおける予測の視点について総合的な手法を提案した。以降に各章の概要を示す。

第1章ではこの論文の背景・目的について記述する。特にイノベーション・マネジメントにおける意思決定のための予測タスクがこれまで有してきた課題やその解決手段としてこれまで提案されてきた手法の中でも、特に本論文が用いるようなデータ分析にもとづく定量的アプローチにおいて必要となる概念を中心に整理を行う。

第2章は、本論文で用いる科学技術対象について概要とデータセットの抽出条件を示し、同時にそれらの知識動向が動的に変化していることを時系列コミュニティイベントの可視化手法を用いて確認する。

第3章は、1. 科学知識の有望領域の予測 に対応する章である。具体的に

はナノカーボン分野および太陽光発電分野を対象に将来有望な萌芽領域の予測手法と解釈手法を提示し、それぞれの分野について議論を行う。

第4章は、2. 技術知識同士の融合の予測 に対応する章である。特許データが有する国際特許分類を対象に、共起情報を抽出した上で将来的に共起するであろう国際特許分類をリンク予測の観点によって予測する。特に、国際特許分類が有する概念の階層構造に着目し、異なる階層の情報が技術融合の予測に寄与することを示す。

第5章は、3. 知識を社会に実装する経済主体同士の関係性の予測 に対応する章である。機械学習を用いた企業活動における取引相手予測の視点では十分に議論されていなかった、モノの流れに関する情報の抽出方法に着目する。業種の条件付き確率を特徴量として導入することで解釈性と精度を両立した潜在的取引相手予測モデルの手法の提案を行う。

第6章では、第3章、第4章、第5章それぞれを用いて実際に得られた結果およびその解釈を整理する。また、これら3手法を統合的に用いることによる付加価値として新たなイノベーション・マネジメントのフレームワークを提案する。

第7章では、科学技術イノベーションプロセスにおけるそれぞれのレイヤーで、分野や産業の境界が曖昧になることを改めて主張する。新たな進出先を探索して自社のコア知識を獲得することが困難となるなかで、本研究の提案および視点が有効であることを主張し、本研究の結びとする。

目次

1	第1章 序論	1
1.1	背景	1
1.1.1	科学技術イノベーション戦略における意思決定に資する予測の重要性	2
1.1.2	科学技術情報に基づくイノベーション戦略意思決定における課題	4
1.2	本論文における予測の考え方	9
1.2.1	本研究における予測タスク	10
1.3	本研究における手法によって便益を得る事が想定できる主な対象	14
1.4	データに基づく意思決定における解釈性バイアス	15
1.5	本研究の特徴	15
1.6	本研究の構成	17
2	第2章：本研究で対象とする知識領域	19
2.1	知識コミュニティの遷移とイベント	19
2.2	本研究で対象とする知識領域	21
2.2.1	ナノカーボン分野	22
2.2.2	太陽光発電分野	26
2.2.3	炭素繊維強化プラスチック技術分野	30
2.2.4	傾斜機能材料技術分野	33
3	第3章：ナノカーボン分野と太陽光発電分野における萌芽領域予測	37
3.1	背景と目的	37
3.2	手法	40
3.2.1	モデル構築	40
3.2.2	特徴量抽出	41
3.2.3	モデル評価	42
3.2.4	将来予測	43
3.2.5	各クラスタにおけるトピック抽出	43
3.3	結果	45

3.3.1	ナノカーボン分野データセットにおける結果	45
3.3.2	太陽光発電分野データセットにおける結果	51
3.4	考察	61
3.4.1	ナノカーボン分野における考察	61
3.4.2	太陽光発電分野における考察	65
3.5	まとめ	66
4	第4章：国際特許分類の階層情報を用いた技術融合予測	69
4.1	背景と目的	69
4.1.1	技術融合	69
4.1.2	書誌情報に基づく知識発見とリンク予測	70
4.1.3	技術的スピンオフ	72
4.1.4	国際特許分類（IPC）の活用	74
4.2	手法	76
4.2.1	予測モデル	80
4.2.2	評価指標	82
4.3	結果	85
4.3.1	CFRP データセットにおける結果	85
4.3.2	FGM データセットにおける結果	105
4.4	考察	125
4.4.1	CFRP の技術融合に関する具体例	126
4.4.2	FGM の技術融合に関する具体例	127
4.5	まとめ	128
5	第5章：サプライチェーンにおける N グラム情報を活用した潜在的取引相手の予測	129
5.1	背景と目的	129
5.1.1	企業間関係メカニズム	129
5.1.2	潜在的なつながりの創造	130
5.1.3	リニアサプライチェーンとネットワークサプライチェーン	131
5.2	手法	134
5.2.1	データセット	135
5.2.2	モデル構築	136
5.2.3	モデルおよび評価基準	137

5.3	結果	139
5.4	考察	144
5.5	まとめ	145
6	第6章：結論	147
6.1	科学技術イノベーションにおける予測活動	147
6.2	有望領域の特定	147
6.3	有望関係の特定	147
6.4	経済主体の関係の特定	148
6.5	提案手法の組み合わせによる統合的予測手法によるイノベーション・マネジメント	148
7	第7章:まとめ	152
8	参考文献	154
9	研究業績	175
9.1	第1章に関連する研究	175
9.1.1	第1章に関連する査読付き論文	175
9.1.2	第1章に関連する査読付き国際学会プロシーディングス	176
9.2	第3章に関連する研究	177
9.2.1	第3章に関連する査読付き論文	177
9.2.2	第3章に関連する査読付き国際学会プロシーディングス	177
9.3	第4章に関連する研究	178
9.3.1	第4章に関連する査読付き国際学会プロシーディングス	178
9.3.2	第4章に関連する資金獲得実績	178
9.4	第5章に関連する研究	178
9.4.1	第5章に関連する査読付き国際学会プロシーディングス	178
9.5	その他の研究業績	180
9.5.1	その他の査読付き論文	180
9.5.2	その他の査読付き国際学会プロシーディングス	181
9.5.3	その他の査読付き国内学会プロシーディングス等	183
9.5.4	その他の査読なし国際会議発表	183
9.5.5	その他の寄稿論文	184
9.5.6	その他の書籍	184

9.5.7	その他の資金獲得実績	184
10	謝辞	185

図目次

1	コミュニティ遷移イベント (Palla et al.(2017))	20
2	ナノカーボン分野関連論文の出版数推移	23
3	ナノカーボン分野関連論文 (2003 年から 2015 年) における各 年上位 10 コミュニティのコミュニティイベント	24
4	ナノカーボン分野関連論文 (2003 年から 2015 年) の各年前後 の遷移行列	25
5	太陽光発電分野関連論文の出版数推移	26
6	太陽光発電分野 (2003 年から 2015 年) における各年上位 10 コ ミュニティのコミュニティイベント	28
7	太陽光発電分野 (2003 年から 2015 年) における各年前後の遷 移行列	29
8	炭素繊維強化プラスチック技術分野 (2004 年から 2016 年) に おける各年上位 10 コミュニティのコミュニティイベント	31
9	炭素繊維強化プラスチック技術分野 (2004 年から 2016 年) に おける各年前後の遷移行列	32
10	傾斜機能材料概念図	33
11	傾斜機能材料技術分野 (2005 年から 2017 年) の各年上位 10 コミュニティのコミュニティイベント	35
12	傾斜機能材料技術分野 (2005 年から 2017 年) における各年前 後の遷移行列	36
13	Model training and prediction	41
14	LDA によるトピックモデル (Blei et al.(2003))	44
15	サブクラスタ別 (第三レイヤー) 上位 1000 萌芽論文の頻度	48
16	ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 におけるト ピック分類概要	49
17	ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 における第 一トピック	49
18	ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 における第 二トピック	50

19	ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 における第三トピック	50
20	サブクラスタ別 (第三レイヤー) 上位 1000 萌芽論文の頻度 . . .	54
21	太陽光発電分野 2012 年出版論文クラスタ 2-3-1 におけるトピック分類概要	55
22	太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第一トピック	55
23	太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第二トピック	56
24	太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第三トピック	56
25	サブクラスタ別 (第三レイヤー) 上位 1000 萌芽論文の頻度 . . .	57
26	太陽光発電分野 2015 年出版論文クラスタ 2-3-1 におけるトピック分類概要	59
27	太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第一トピック	59
28	太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第二トピック	60
29	太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第三トピック	60
30	国際特許分類 IPC の階層別名前 (Source: (WIPO,2014)	76
31	階層概念	77
32	分析プロセス概要	78
33	ネットワークにおけるグローバル指標の推移 (CFRP データセット: サブクラス)	85
34	ネットワークにおけるグローバル指標の推移 (CFRP データセット: グループ)	86
35	ネットワークにおけるグローバル指標の推移 (CFRP データセット: サブグループ)	88
36	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	90

37	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	90
38	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	90
39	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセット全 IPC を対象]	92
40	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセット全 IPC を対象]	92
41	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセット全 IPC を対象]	92
42	グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	94
43	サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	94
44	サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	94
45	グループ (8digit) の予測に上層情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセット全 IPC の組み合わせ対象]	96

46	サブグループ (10digit) の予測に上層情報であるグループ (8digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセット全 IPC の組み合わせ対象]	96
47	サブグループ (10digit) の予測に下層情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [CFRP データセット全 IPC の組み合わせ対象]	96
48	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	98
49	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	98
50	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	98
51	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]	100
52	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]	100
53	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]	100
54	グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	102

55	サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	102
56	サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	102
57	グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]	104
58	サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]	104
59	サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) 下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]	104
60	ネットワークにおけるグローバル指標の推移 (FGM データセット : サブクラス)	105
61	ネットワークにおけるグローバル指標の推移 (FGM データセット : グループ)	106
62	ネットワークにおけるグローバル指標の推移 (FGM データセット : サブグループ)	106
63	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [fgm データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	110
64	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	110

65	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	110
66	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [fgm データセット全 IPC を対象]	112
67	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [fgm データセット全 IPC を対象]	112
68	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [fgm データセット全 IPC を対象]	112
69	グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	114
70	サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	114
71	サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	114
72	グループ (8digit) の予測に上層情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセット全 IPC の組み合わせ対象]	116
73	サブグループ (10digit) の予測に上層情報であるグループ (8digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセット全 IPC の組み合わせ対象]	116

74	サブグループ (10digit) の予測に下層情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) [FGM データセット全 IPC の組み合わせ対象]	116
75	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	118
76	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [fgm データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	118
77	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	118
78	サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]	120
79	グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]	120
80	サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象] .	120
81	グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	122
82	サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	122

83	サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]	122
84	グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]	124
85	サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]	124
86	サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) 下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]	124
87	主業の条件付き確率	134
88	3 ノードネットワークモチーフ (Source: Math Insight(2016)) .	138
89	各変数の Impurity	143
90	イノベーション・マネジメントプロセス (上図：中期的アプローチ。下図：短期的アプローチ。)	149
91	統合的予測手法に基づくイノベーション・マネジメント	150

表目次

1	NEDO による近年の技術ロードマップ例 (Source: NEDO(2017))	3
2	各モデルの学習年と評価年の組み合わせ	41
3	萌芽予測モデルに用いるネットワーク特徴量	43
4	各年における評価結果 (ナノカーボン分野)	45
5	各年のモデルにおける寄与度上位 5 つの特徴量 (ナノカーボン分野)	46
6	2011 年に出版された論文のうち萌芽論文となる可能性が高いと予測された論文上位 10 件	47
7	各年における評価結果 (太陽光発電分野)	51
8	各年のモデルにおける寄与度上位 5 つの特徴量 (太陽光発電分野)	52
9	2012 年に出版された太陽光発電分野の論文のうち萌芽論文となる可能性が高いと予測された論文 10 本の概要と 3 年後の実際の被引用数	53
10	2015 年に出版された論文のうち萌芽論文となる可能性が高いと予測された論文上位 10 件	58
11	学習期間窓およびテスト期間窓の組み合わせ	79
12	本研究における分析条件の整理	84
13	上位 5 ペア (CFRP データセット: サブクラス)	86
14	上位 5 ペア (CFRP データセット: グループ)	87
15	上位 5 ペア (CFRP データセット: サブグループ)	87
16	上位 5 ペア (FGM データセット: サブクラス)	107
17	上位 5 ペア (FGM データセット: グループ)	107
18	上位 5 ペア (FGM データセット: サブグループ)	108
19	CFRP の技術融合に関する具体例 (サブクラス)	126
20	FGM の技術融合に関する具体例 (サブグループ)	127
21	ネットワーク中心性	137
22	企業属性情報概要	139
23	各年における企業数および取引数	139

24	主業上位 10	140
25	各年における正例負例サンプルサイズ	140
26	出現頻度上位 10 組の 3gram の生起確率	141
27	各モデルの Precision	141
28	各モデルの Recall	142
29	各モデルの Fmeasure	142

1 第1章 序論

1.1 背景

科学技術への適切な投資は、技術志向型の企業はもとより各国政府にとっても将来の競争優位性を確立するための重要な意思決定である。インプットとしての投資金額とアウトプットとしての学術成果の関係に着目する研究 [1, 2, 3, 4, 5] や、R&D 投資がもたらす GDP や豊かさとの関係に関する研究 [6, 7] から、不確実性を内包する科学技術イノベーションを戦略的にマネジメントできるかどうかは大きな課題とされている。近年では、アジア諸国を対象とした研究開発の進歩と国力の関係性に関する議論 [8, 9] も活発であり、これまで経済発展の常識として考えられてきた後進国が先進国と同じ道を通じてキャッチアップする雁行型発展形態 [10, 11] は科学技術の文脈においては過去のものであるという指摘は Sakata and Sasaki(2013) によってなされている [12]。技術志向型企業における研究開発投資のアウトカムとしての特許や研究、またそれに付随する知識が企業の競争力に不可欠であることはすなわち、科学技術競争力を生み出す人材の教育投資が不可欠である事を意味する。このような背景のもと、STEM(Science, technology, engineering, and mathematics) 人材の教育にいち早く力を注ぎ始めている国として米国、オーストラリア、カナダ、トルコなどが挙げられる。

イノベーションという言葉はソーシャルイノベーション、ビジネスイノベーション、プロセス・イノベーション、テクノロジー・イノベーションなど多様な意味で用いられる言葉であり、昨今ではその言葉指す対象がきわめて広義である。その中でも科学がイノベーションの源泉となり、企業がその源泉をもとに技術を開発して、産業を生み出すというプロセスを指すことがある [13, 14]。このようなイノベーションにおいては適切な科学技術投資がアウトカムに不可欠であるとみなされる。このように特に科学を源泉とするようなイノベーションを科学技術イノベーションと呼び、本論文では断り無くイノベーションという言葉を用いた場合は科学技術イノベーションを指すものとする。本項では、背景として科学技術イノベーション戦略における意思決定についてその現状と課題を述べる。

1.1.1 科学技術イノベーション戦略における意思決定に資する予測の重要性

企業にとっての研究開発戦略、政府にとっての科学技術政策に代表されるような意思決定プロセスでは、対象となる分野の科学的知識や技術的知識の動向把握や将来の見通しを行い、業界や領域における自社・自国のポジションを明確にし、次の一手を打つための判断が行われる。本論文ではこれらのプロセスを総じて科学的技術イノベーション戦略とする。科学技術イノベーション戦略にはその意思決定を行うために豊富な知的情報が必要となる。その知的情報の形態のひとつに技術ロードマップがあげられる [15, 16, 17]。Branscomb and Keller(1998) は科学技術ロードマップを、当該技術に関する魅力的な将来に資するような科学的な情報に基づくビジョンの合意形成であると定義している [18]。また Galvin(1998) の言葉を借りれば、当該分野に変化をもたらすような有望なドライバに関する想像力や集合知によってえられる将来的見通しであるとも言える [19]。すなわち、技術ロードマップの策定目的は、企業や政府が今後の投資や現在の技術知識資産棚卸しを目的とした、時系列情報が付随した技術的見通しといえる。このプロセスでは、対象分野とそれらを構成する要素技術についてこれまでの技術革新やスピード、市場や社会の受け入れ可能となるタイミングを考慮することによって、特定の技術が得られるであろうマイルストーンを構築していく。

企業においても同様にこれらの技術ロードマップを構築・活用することが少なくない。特に企業が構築するこれらのロードマップの目的は、自社のビジネスと技術の統合に資することが前提となっており、また技術戦略に資する時系列的な製品と技術の間の相互関係を表現したものを指す [20]。これらの技術開発課題は、異なる他の目標に依存して決定されることが少なくない。2020年までに温室効果ガスを3.8%(2005年度比)削減したいという目標であったり、5年後に新製品を投入したいという目標を解決するための技術開発が必要であったりする。このように自社や政府が主体的にこのようにしていきたいという目標策定が先に存在するようなケースでは、フォーサイトやフォーキャストに対する言葉としてバックキャストと呼ぶ。フォーサイトやフォーキャストは前述の通り将来起こるであろうことをデータを元に演繹的に予測する方法である。一方でバックキャストとは、目標や方向性が所与であるとして、現在からその時点までの各マイルストーンにおいて達成すべき目標を規定するものである [21]。高炭素型による発展から抜け出せなくなってしまう状態にならないよう

表 1. NEDO による近年の技術ロードマップ例 (SOURCE: NEDO(2017))

領域/分野	平成 23 年度 (2011)	平成 24 年度 (2012)	平成 25 年度 (2013)	平成 26 年度 (2014)	平成 27 年度 (2015)
電子・情報通信	電子・情報技術分野に関する戦略策定調査事業	電子・情報技術分野の技術ロードマップ 2013 の策定に関する調査	電子・情報技術分野の技術ロードマップ 2014 の策定に関する調査	電子・情報技術分野の技術ロードマップ 2015 の策定に関する調査	IoT 社会で重要となるデータ処理・制御技術等に関する調査
ナノテクノロジー・材料			ナノテクノロジー・材料技術分野の技術ロードマップ 2014 の策定に関する調査	ナノテクノロジー・材料技術分野の技術ロードマップの策定に関する調査	ナノテクノロジー・材料技術分野の技術ロードマップ 2016 の策定に関する調査
システム・新製造	産業競争力の強化に向けた周辺状況調査	中長期的なロボット産業の競争力の強化に関する検討	我が国ロボット技術の社会実装の推進等に関する検討		
システム・新製造	機械システム分野に関する戦略策定調査	MEMS 分野の革新的デバイスに関する調査及び技術戦略マップ改訂と国際標準化に向けた検討	航空機分野における戦略策定調査		
システム・新製造	航空機分野の国内外技術動向および市場分析に係る情報収集				
バイオテクノロジー・医療				バイオテクノロジー・医療技術分野の技術ロードマップ 2014 の策定に関する調査	
環境	使用済み製品中のレアメタル等を対象としたリサイクル技術・システムに関する動向調査	グリーンサステイナブルケミストリー技術分野の戦略策定調査		3 R 分野の技術戦略に関する検討	
環境		3R 分野の技術戦略マップ改訂に関する調査			
エネルギー	二次電池の技術戦略マップローリングおよび中長期世界競争力に関する調査	二次電池技術開発ロードマップのローリングに関する調査		平成 26 年度超電導機器の市場形成に資する技術戦略策定に向けた調査	

に、低炭素社会の実現のために必要な具体的なアクションと眼前の目標を定めることはバックキャストアプローチによるものである。ビジョンを策定するにはトップダウン的なアプローチが不可欠であるが、実際にそのビジョンが実現できるかどうかといった視点は現在有しているデータに基づいて検証をしなければならない。その意味で、予測といった視点は、信頼性の高い技術ロードマップには不可欠である [22]。

1.1.2 科学技術情報に基づくイノベーション戦略意思決定における課題

科学技術情報に基づくイノベーション戦略を行う際には信頼のおける一次情報を拠り所にする必要があるが、何を一次情報とするかはその手法によって異なる。本項目では多くの情報において共通的に発生している普遍的課題であり、かつ科学技術情報の代表とされる学術論文や技術情報において顕著である課題として次の6つを指摘する。すなわち、情報の量的問題と質的問題、専門知識の細分化、サイエンスリンケージ、イノベーションの加速、知識の外部性である。

1.1.2.1 情報の量的問題

情報化社会以降顕在されてきたひとつの課題を表す言葉として、情報の指数関数的な増加が挙げられる。例えば1751年のフランス百科全書に掲載された記事の数は71,709程度、ページ数16,142の28巻であったとされている [23]。これは当時、明示的に記録されていた全ての情報とほぼ同義であるといえる。一方比較対象として現在の百科事典のWikipediaを挙げるとすると、収録されているフランス語の記事の数は1,838,612、英語の記事に至っては5,329,023(2017,2月時点)とされている。ましてやこれらは、現在明示的に記録されている情報のごく一部である。知識を記録するという行為に必要なコストが大幅に減り、広く公開することの機会は今後より一層増えていく。科学技術情報においても同様のことが起きていることは多く指摘されている [24, 25]。Price(1961,1965)は、科学技術情報が10年から15年ごとに倍の情報量となっていることを1960年代の時点で指摘していた [26, 27]。その後、2000年代に入り、van Raan(2000)により世界における科学知識は2012年時点で、9年間でおおよそ倍になっている事を指摘している [28]。また、マックスプランク研究所のLutz Bornmann and Mutz(2015)によれば、1980年から2012年までの世界における論文出版数は $y(t) = b_0 * \exp(b_1 * (t - 1980))$ のモデル

で 96% 程度説明ができるという [29]。まさに指数関数的な増加そのものであることが定量的に示されている。このように情報が指数関数的に増えることにより、その情報を用いた意思決定をより困難にしている。

1.1.2.2 情報の質的問題

情報化社会以前では、出版という行為は一定の知見のある第三者もしくは編集者によって一定のフィルタリングがされた上で商業的に刊行されることが前提となっていた。情報化社会以降は、ブログやインターネット記事を含む情報コンテンツが指数関数的に増加してきたことは前述のとおりである。これらは知見のある第三者に十分に精査されることのない情報であることが多い。情報量の一方的な増加は相対的に情報の質や信頼性が下がり続けていくことが避けられない。

科学技術においても同様の事象が指摘されており、Larsen and Ins(2010) や Statzner and Resh(2010) は学术论文の増加と同時に起きていることとして引用数の低下を示している [30, 31]。特にオープンアクセスジャーナルが増えることによってその傾向は加速している。通常、学術雑誌のビジネスモデルは読者が料金を払うことによって成り立っている。一方でオープンアクセスジャーナルは掲載することによって収入を得る仕組みである以上、質の担保は難しい。このように世界的な潮流としては、量が増えることによって相対的な質が低下することが指摘されているが、残念ながら我が国においては量と質の両方の低下が見られるという懸念がされて久しい [32]。

1.1.2.3 専門知識の細分化

科学研究費において、分科細目表の細分化が進んでいる。1983 年（昭和 58 年）には 191 であった細目数は、2014 年（平成 26 年度）には 321 の細目数となってきた [33]。この 30 年間にしておよそ 1.7 倍となったことになる。このような状況下において、研究者は自らの研究をいずれかの細目に一意に決定しなければならないこととなり、新しい研究分野や異分野融合の研究は応募しにくいのではないかという課題があがるなどしている。このような専門性の細分化が、基礎研究の質的低下をもたらしている一つの要因であるという懸念は 1985 年には既に指摘されていた [34]。このようななかで、我が国の科学研究費助成事業に限れば、新たな審査区分や審査方式などの導入による見直し（科研費審査システム改革 2018）が行われる見通しである。しかしながら、科学研究

費助成申請は研究者における研究活動の一部でありこれだけで解決できる問題とは考えにくい。むしろ本質は論文投稿の際の対象にある。学会や雑誌が多様になりつつあるなか、一つの研究を見方によれば複数のアウトプットにすることは、容易になってきてしまっている。すなわち、研究者は自らのアウトプットをできるだけ薄く切り取り、“salami sliced publishing”していくことが合理的な判断となってしまう [29]。結果として前述した知識量の増加につながっている。このようななか、ピアレビューの仕組みについても課題が存在してくる。ピアレビューは、投稿された学術論文の査読の基本的な仕組みである。本来であれば、研究者にとっては新たな科学的知見を得られる喜ばしいプロセスである。一方で、多くの研究者、科学者がこの貢献への対価を支払われず、ボランティアによって成り立っている。サラミスライスされた論文の量産によって、新しい科学の理解にほとんど貢献をしない論文へのますます時間を費やさなくなっていくことは想定される [35]。

1.1.2.4 サイエンスリンケージ

Mansfield(1991) は、学術研究の貢献がなければ新たな産業イノベーションの1割はその登場が遅れたであろうと推測している [36]。大学などにおける研究が企業を中心とした産業が行う研究開発に与える影響は多くの分野で上昇している。技術に科学が与えている影響を示唆する指標のひとつとしてサイエンスリンケージがあげられる [37, 38]。サイエンスリンケージは、特許が引用する論文の平均的な数で分野ごとに算出されるのが一般的である。学術的な研究成果と産業における研究開発が密接につながっている分野であるほどこの数値は高く、欧米と我が国を比較すると圧倒的に米国におけるサイエンスリンケージが高いことが分かる。科学論文を引用している特許のほうが、より多くの引用を獲得していること [39, 40, 41] から、学術論文は科学技術イノベーションの加速に必要不可欠であることが分かる。玉田 (2010) はキリンホールディングスの各企業研究所をサーベイし、バイオ技術分野において科学と技術の結びつきについて考察を行っている [42]。玉田 (2010) によれば、バイオテクノロジー以外にサイエンスリンケージが高い分野として、ナノテクノロジーがあげられる。一方で IT 関連や環境保全関連技術についてはサイエンスリンケージが低い傾向が認められている [42]。Looy et al.(2003) は、10 の技術領域に対して欧州 8 カ国の特許データを元にサイエンスリンケージを算出した結果、バイオテクノロジー、薬学、有機化学、半導体、農学食品学、光学、制御工

学、情報工学、材料化学、通信の順でサイエンスリンケージが高いことを示した [38]。分野間の差はあるものの、サイエンスリンケージの強さはそれぞれの分野でより高くなっている。サイエンスリンケージが高まっていくということは、新たな技術開発において科学的知識を拠り所にする有効性が高くなることを意味する。

これは、科学技術からのボトムアップ的な要因もさることながら、社会的ニーズに応えるための必然的な現象であるとも言える。例えば、医療における目標のひとつには健康寿命の延伸や患者の QOL の向上があげられる。これらを満たすニーズは複数あり、低侵襲治療、健康管理、予防、再生医療による機能代替などがあげられる。これらの解決技術として DDS、診断キット、MEMS ナノ粒子、インフォマティクスなどがあげられる。これらを実装するには、さらに関連する技術として、ナノテク、MEMS 技術、ロボット技術、情報通信技術といった複数の分野を有機的に活用しなければ実現が困難である。これまで十分に科学的な拠り所に依存する必要がなかったような分野においても、今後より科学的知見が必要になる分野は増えていくことは多くの分野で発生する現象であると想定できる。

1.1.2.5 イノベーションの加速

科学として生まれた知識が実際に社会にインストールされるまでのスピードが速まっている。S カーブ [43] に代表されるようなテクノロジーライフサイクルを描くまでの時間がますます短くなっている事を意味する。基礎研究、応用研究、開発を経て、製品の展開となるように、科学によって生じた知識が製品技術として線形に適用するようなモデルをイノベーションのリニアモデルと呼ぶ [44] が、現実においてこのようなりニアモデルをとらないケースも少なくない。それは科学から技術へという一方向のみに影響を与えるのではなく、科学と技術は相互に影響をあたえ合うことによって相互に新たな知識が生み出されるプロセスが行われやすくなることを意味する。これは前述のサイエンスリンケージが高い分野であるほど発生しやすいと考えることができる。また、科学と技術の 2 者間のみならず、技術と社会、社会と科学が相互に影響し合うことでそれぞれの知識が高度化していく。知識が複雑になるほど、異なるレイヤーとのインタラクションが増え、そのことがますます知識を複雑なものにしていく。イノベーションのスピードについての研究は幾つかある [45, 46] が、Markman et al.(2005) は新しい技術や技術革新はより一層ライフサイクルのスピードを

加速させ、イノベーションを誘発するとしている [47]。科学技術の発展速度が増していることから、意思決定に求められるリードタイムはよりスピードが要求される。Collins and Evans(2002) の指摘のように、既存の手段による科学技術の合意形成の確立を待っているようでは、社会が抱えている課題や需要に追いつくことが困難となってしまう [48]。

1.1.2.6 知識の外部性

情報量が増加し多様性が拡大していくなか、自国や自社が有している内部情報のみで課題解決を行おうとする自前主義は非合理的なアプローチとなりつつある。経済主体がアウトソースを行う判断基準のひとつは、自社で行うことよりも総合的にコストが低いと判断したことにある。サプライヤを選定するこの行為は、部品や資材の調達と言った形でかつてから行われてきたものである。いまは部品や材料の調達のみならず、知識をとりこむということに大きなコストがかかり始めている。これは既に述べた、情報の増加、細分化、サイエンスリンケージの増大、イノベーションの速度といった要因によるものである。今後の産業構造変化においては、オープンイノベーション [49] 以上に大きな知的構造変化が起きるであろうことは想像に難くない。例えば、あらゆる産業に影響を与えるであろう社会的構造変化の一つに、産業のスマート化があげられる。スマート化とは、情報システムや装置に高度な処理を可能にする機能を与えることをいう。スマートグリッド、スマートホーム、スマートシティ、農業や向上のスマート化、家具家電のスマート化と多様な形態があり、IoT(Internet of Things) もスマート化の一つの形態である。実態的には演算処理とセンサを有しネットワークに繋ぐことによる管理による効率化ともいえる。内閣府による科学技術イノベーション総合戦略でも、3つの視点の一つとして掲げられている [50]。このスマート化の動きは、サプライチェーン全体の高度化が必要となる。これまで農業に従事する主体がこれまで情報通信技術関連の主体と取引するということは稀であった。スマート化の動きは必然的に農業における情報通信技術の導入を促すこととなる。今後はより一層、農業を農業という枠組みだけで、製造業を製造業という枠組みだけで議論できなくなるであろう。全く異なる分野から知識を取り入れなければならないことは、大きな情報収集コストを掛けなければならないことを意味する。

1.2 本論文における予測の考え方

有史以来人類は、手元の情報を元に将来起こるであろうことを予測する方法の開拓に取り組んできた。予測モデルの目的は将来起こるであろうことを演繹的に推測する点にあるが、科学の歴史において発見されてきた多くのモデルは当初帰納的に発見されたものである。観察データを元にした帰納的推論によってケプラーが導き出した得られたケプラーの法則は、惑星の運動の法則化という観点で将来予測である。さらに後にニュートンの万有引力からケプラーの法則を説明する演繹的モデルに繋がった。すなわち科学そのものが、過去から現在までに確定しているデータをもとに将来起こることを予測するモデリングプロセスであるといえる。このような科学によってモデル化された理論をもとに、より人類の便益に資する仕組みを構築する手段として技術が生まれた。科学と技術、総じて科学技術は人類が火を手に入れて以降、文明の高度化、社会の先進化、産業の発展に寄与してきた。同時に、依然として人類が有している知は限定的であり、未開拓とされるフロンティアは無数に存在することも事実である。そのような中、近代においては科学技術における意思決定そのものに対して、将来予測をするという動機にその注目が集まっている。将来予測をするためのプロセスとしての科学ならびにその実現手段としての技術が向かうべき方向自体を将来予測するというモチベーションが生まれたこと自体は皮肉的ではあるが、同時にこれまでの人類が積み重ねてきた知識が豊富であるが故のものであると考えることができる。

人類が予測というものに取り組む初めてからの歴史は長い、こと科学技術の動向予測という観点では近代になってより注目を集めつつある。科学技術政策の観点においてもこの需要は大きく、科学技術資源の投入配分はその国が向っていく科学技術の方向付けそのものであり、国力に大きく影響を与える重要な意思決定である。技術先進の企業においてもその需要は変わらない。自社が有している資源をもとに研究開発戦略を策定することは、自社の方向性を形付ける行為と同義である。グローバリゼーションが進み、こと特に企業活動においては国境の概念が希薄しつつある中で、自社が注力する技術が世界ではどういった位置づけにあるか、そしてその技術をより先進化させるためにはどういった科学知識が必要かという情報を得ることは経営課題の中核になりうる重要な活動である。

科学技術予測を行う手法の一つの事例としては、ドイツと日本において 25 の

科学技術分野における将来展望を描く試みがなされた際に用いられた手法としてデルファイ法があげられる。我が国ではこの手法に基づき 1971 年より 5 年おきに科学技術予測調査を行っている。しかしながら、時代とともにその手法の限界も徐々に明らかになってきた。実際、我が国で行われてきた科学技術予測調査においてデルファイ法のみに基づいてきたのは 1996 年までである。複数の要因があるなかでも、本論文では複雑な科学技術情報を扱うにあたって個人が処理・理解であることができる限界を指摘する。専門家が処理しきれない膨大な情報処理能力は機械の得意とするものである。したがって、専門家の知見が持つ欠点を補間するという観点で、機械による意思決定支援をもとに人間（専門家）が意思決定を行うというハイブリッドな仕組みが有効である。こと重大な意思決定であるほど、エビデンスベースによる意思決定が求められているなか、指摘するような限界をを補助する道具としてのデータ分析アプローチによる意思決定支援手法について議論をしたい。

情報を情報として残すことのコストや広く公開することへのコストが格段に下がることで、近代における情報は情報化社会以前にまして玉石混交となってきた。一方的に増加し続け、互いに繋がり続ける情報の有機化はヒトによる理解をより困難にしている。科学技術情報も同様に複雑化かつ増大化している中で、特定の分野について専門的かつ網羅的な見解を実施することはもはや困難となっているという指摘は多くなされている。そのようななかで、計算機の性能向上や情報処理技術の高度化は科学技術予測に一定の寄与をするであろう期待がされている。実際、多くの産業領域では、適切なリテラシーと知識に基づくことで、複雑かつ膨大な情報から有益な知識を生み出すという手段が現実のものとなっている。多次元かつ大規模なデータからのパターン認識および予測を可能とする統計的機械学習に基づく分析手法では、その目的とデータ構造に応じた手法が複数提案されつつあり、科学技術予測に資する手法として用いられることも増えてきた。本論文では、そのような中でイノベーションの視点として 3 つのプロセスに分類しそれぞれに対応した分析を総合し、イノベーション・マネジメントに一定の視点を投じるものである。

1.2.1 本研究における予測タスク

構造変化を伴う知識領域において、上流の知識構造が動的に変化に応じて下流においても変化が求められる。例えば、二次電池（ニッケル水素蓄電池・リチウムイオン電池）のようなサイエンスリンケージの高く構造変化の速度が早

いような科学分野において考える。二次電池（ニッケル水素蓄電池・リチウムイオン電池）は技術レイヤーにおける適用先もひろく、当初想定されてなかったような用いられ方が発生してもおかしくない。事実、電気自動車にむけた車載用電池の開発にといった融合形態が存在することとなった。さて、トヨタが電気自動車のようなこれまでと異なるパラダイムの製品を開発する際、関連する知識を持ち合わせていない。結果としてパナソニックが有するような二次電池の技術的知識知見が不可欠であり、結果として両者の取引関係が発生する。本研究では、イノベーションプロセスに必要な上流から下流までの全行程として科学、技術、産業における情報を網羅的に対象とし、下記の3種類の予測タスクについて議論を行う。

- 1. 科学における有望な知識の予測
- 2. 技術知識同士の融合の予測
- 3. 知識を社会に実装する主体の行動の予測

下記にそれら概要とその意義を示す。

1.2.1.1 科学における有望な知識の予測

科学技術の予測におけるモチベーションの一つに、将来有望となるであろう科学知識を事前に特定したいという動機が挙げられる。本論文ではこれを一つの予測対象とする。将来どのような科学知識が注目に値するかといった知見をあらかじめ有することは、競合他社に先駆けて研究開発投資を行うために有益である。あるいは競合他社がどのような戦略を行おうとするかを予見するという目的においても有益である。科学や技術の情報は国境の概念と無関係に取得、活用、管理されるべき対象となりつつあり、特に企業活動においてはいまや国という枠組みの中でのみの情報で行える技術経営活動は過去のものである。そのような中、世界が当該分野についてどのように考えどの方向に向かってきたかという過去から現在までの情報だけでなく、どの研究が有望となりうるかという未来の情報をあらかじめ予見する方法に注目が集まってきている。本論文では具体的にはそのような手法のうち新たに提案された手法を具体的分野として太陽光発電分野とナノカーボン分野の双方に適用することでその結果を元にそれら分野の注目研究について議論を行う。

1.2.1.2 技術知識同士の融合の予測

二つ目の予測対象として、これまで異なる領域ものと考えられていた複数の科学知識や技術知識が融合することで新たな価値を創出する事象を予見したいというモチベーションを提示する。イノベーションと呼ばれる事象についての一般的な説明では、新たなものがゼロから生み出されたように説明される場合が少なくない。しかしながら実際に知識の創出においてそのような現象に起因することは極めて少ない。イノベーションの訳語に新結合 (neue Kombination) という言葉が当てられることから分かる通り、本来イノベーションとは既存の知識同士を結びつけることによって生まれる新たな価値を指す。シュンペーターはあらゆる生産要素の結合の仕方をイノベーションのひとつの事象としてあげているが、無論この生産要素には知識も含まれる。科学技術における知識の結合はあらたな領域の誕生をもたらすことがある。一つの産業が一つの技術分野のみに大きく依存してきた時代はもはや過去のものであり、現代において特定の需要を満たす製品やサービスを実現するために必要な科学技術分野は多岐にわたる。これまで無関係と考えられてきた科学技術同士の結合があらたな価値を生み出すことが有るばかりでなく、新たな分野の創出に繋がることも少なくない。このような技術融合はまごうこと無くイノベーションの本質的形態である。本論文ではこのような技術における知識の融合を予測する手法の提案を行う。特に、技術の階層構造に着目し、予測手法としての新規性と有用性を議論した上で、実際の分析結果をもとに対象分野の知識結合について議論を行う。

1.2.1.3 知識を社会に実装する主体の行動の予測

三つ目の予測対象として、これらの科学技術知識を源泉にすることで実際に事業を行う主体である企業の取引行動の予測について提示する。技術を競争力の源泉としている企業にとって、自社が有している技術やノウハウは、現在の顧客が抱える課題への解決手段である。言い換えれば、現在の顧客が抱える課題解決に最適化した形で製品やサービスに落とし込まれているに過ぎない。しかしその一方でそれらの技術やノウハウや知識は、現在まだ見ぬ潜在的な顧客が抱える課題を解決する可能性を少なからずはらんでいるという点で経営資源でもある。痛くない注射は深絞り加工を得意とする金属板加工技術の企業から生まれたものであり、当該企業はもともと医療産業を顧客としていたわけでは

ないことはよく知られている。ステープラーの代名詞であるホチキスを産んだ企業の前身もゼロ戦の尾翼加工技術を得意としていた企業であり、当時の金属加工技術がステープラーを産んだ技術の源泉となっている。このように自社が属している産業以外に潜在的な需要があるということは珍しくない。自社の産業以外への情報に積極的にアクセスしない限り、新たな市場を切り開くことは困難であるが、多くの場合、その情報の取得コストや処理コストは高いがゆえに行われにくい。企業にとって、売上に直接繋がるかどうかわからない不確実性の高い情報のために多額のコストを投入することは多くの企業にとってためられる選択であることは容易に理解できる。一定の研究開発予算を投入することができるような体力のある研究開発型企業であっても、他産業に関する知識の処理コストは高い。ここ数年で、アイデアソンやハッカソンと呼ばれるイベントを通じて、自社では生まれなかったあらたな発想や製品プロトタイプを得る試みがなされてきてはいるものの、適切にマネジメントするには一定の知見が必要であることもあり、依然として一般的に広まるまでには課題が残る。すなわち、企業取引は必然的に情報コストによっても大きく左右されてしまっている。取引にかかるコストによって組織の境界を決定する考え方を組織間関係論においてはコストパースペクティブと呼ぶが、このコストには情報コストも含まれる。この情報コストを下げることは、他産業へのアクセスを容易にすることを意味する。本論文では、潜在的な企業取引予測というタスクにおいて、サプライチェーン上の位置付けが自社と同じような役割をしている他社が有する情報の有用性を検証する。この手法によって、自社が行っているビジネスの新たな展開を推薦することが可能となる。

以上三つの予測対象について説明した。前者二つの予測については、科学技術書誌情報を用いた分析とする。具体的には科学知識を記述した文書の代表として学術論文群を分析対象データの一つとして扱う。また、技術知識を記述した文書の代表として特許公報群を分析対象の一つとして扱う。本来、科学および技術に係る知識は、学術論文や特許公報以外にも企業が独自に公開している技報、企業内のみで蓄積されている論文なども存在する。また、明示的に文書に記述されていないようなノウハウも科学技術知識の範疇であるが、本論文においてこれらの情報源は分析の対象外としている。もう一つのフレームワークについては、日本企業の既存の取引データを用いた分析とする。具体的には、帝国データバンク株式会社が提供する企業取引データであり、サプライヤからカスタマへのモノの流通を記録したデータベースである。

1.3 本研究における手法によって便益を得る事が想定できる主な対象

事業転換（ピボット）によって成長を遂げてきた企業は少なくないが、科学技術を事業の中核とする企業において、新たな分野への進出は今後いっそう不可欠な要素となる。特に知識構造の変化が激しい科学技術分野における知識に基づくビジネスでは、常に自社の事業の中核が変化しうることを前提にイノベーション・マネジメントを行う必要がある。また、経営者の世代交代によって、先代とは異なるビジネスに舵を切ることもし少なくない。特に我が国では、人口の多い世代（1947年から1949年生まれ）の経営者が70歳になる時期を迎えており、事業継承の2017年問題と呼ばれることもある。このような時代では、これまでの自社のコア技術、コア製品だけでなくそれらの知見を元にした新たな研究開発や未踏市場への進出が必要となってくる。

本研究における手法によって便益を得る対象は、主に未踏分野への進出を必要としている企業の経営者や研究開発責任者などを想定している。特に、第三章で議論する萌芽予測は、これから進出する未踏分野において当該分野の専門家の中で、近い将来伸びることの合意が作られつつ有る分野を早期に特定することを可能とする。ここでいう近い将来とは、イノベーション・マネジメントにおいて現実的かつ実用的な期間、具体的には3年程度を想定する。この期間で急激な注目が集まるであろう有望な知識領域について本研究ではあえて萌芽領域という言葉を用いる。なお、イノベーション・マネジメントの時間軸については第6章においても議論する。

1.4 データに基づく意思決定における解釈性バイアス

近年、データに基づく経営における意思決定の潮流は増々浸透しつつ有る。1990年代に提唱された根拠に基づく医療 (EBM: Evidence-based medicine) による影響も少なくない。EBMは臨床研究の結果が日常の診療に生かされない状況の改善を目指したものである。Evidence-basedという言葉は我が国でも1995年ごろから徐々に認知されてきたが、近年では政策立案における意思決定はもとより、企業経営における意思決定においてもこの考え方が進んできている。ここまで、意思決定という言葉を通り無く用いてきたが、これは判断という言葉と区別して用いる必要がある。判断は機械によって可能なものであるが、意思決定を行うのはあくまで人間である。すなわちデータに基づく意思決定を現実用いる際は、人間が行う意思決定を前提としているがゆえ一定のバイアスが生じ、それが有効なアクションの妨げと成ることがある。大岩 (2017) は人工知能技術の判断時に発生する三種類の代表的なバイアスとして、先入観バイアス、失望バイアス、解釈性バイアスについて整理をした [51]。本研究ではイノベーション・マネジメントに資する予測分析に基づく意思決定を阻害する要因の一つとして解釈性バイアスに着目する。データ分析に基づく意思決定では、アルゴリズムや特徴量の透明性や解釈性はアルゴリズムそのものの性能に加えて重要な評価要因である [52]。EUでは、人工知能を始めとする技術による意思決定がユーザへの影響が大きい場合、導出した結果の根拠をユーザが求める権利を要求する規制が成立している [53]。このように、データ分析における解釈性は、データに基づく意思決定において無視できない。判断を目的とするアルゴリズムであれば深層学習を始めとする手法の幾つかは有効かもしれない。しかしながら、イノベーション・マネジメントに代表されるような人間による意思決定に資する分析手法としては、アルゴリズムをそのまま用いるだけでは解釈性といった点でやはり制約がある。そこで本研究では、解釈性を担保するための工夫として、用いるデータの構造情報に着目し、目的に応じて本質的な特徴量を抽出することの重要性に着目する。

1.5 本研究の特徴

意思決定にデータ分析が用いられて久しく、科学技術イノベーションに関連する意思決定においても例外ではなくなっている。近年では、機械学習を

用いた分析も高度化しておりその範囲はより拡大しつつ有る。機械学習による有益な点は、複数の特徴量により高い予測精度を実現することができる点にある。深層学習などを活用することでより抽象化された特徴量によって高い精度をもたらすことが確認され始めている。これは一般的に、機械学習の前提は機械が「判断」をするモデルの構築であり、特徴量やモデルそのものの解釈可能性が低くとも（すなわち抽象度が極めて高くとも）高い予測精度の実現が命題であることによる。

一方で、人間が意思決定をする材料としての分析の場合、解釈が困難な特徴量だけでは、高い精度のモデルを実現できたとしても適切なアクションにつながりにくいという欠点があげられる。ここに、意思決定における分析の課題が存在する。精度のみを追求するあまりに、利用シーンにおける乖離が生じてしまう例は実利用において少なくない。イノベーションマネジメントに資する分析も例外ではない。これらのシーンに特徴的な点として、大きく3点挙げられる。一つは、これらのシーンに必要な最終的なタスクは、判断をするというタスクではなく意思決定を行うというタスクであるという点である。特定の基準に応じて判断をするだけであれば機械のみに委ねることができるが、最終的に行うのは人間による意思決定である。2つ目にアカウントビリティが挙げられる。企業における研究開発戦略や、政府による科学技術予算編成に関わるような意思決定には大規模なリソースの投入や行政コストが必要となる。大きな影響を与える意思決定である以上、他者へ説明可能であることが前提となる。動的に変化する知識構造の業界において、意思決定のスピードは不可欠である。この点においても情報伝達コストが低くなければならない。3つ目として、意思決定における分析は複数の材料の一つであるという点があげられる。イノベーション・マネジメントの文脈において、単一の分析結果のみにおいて意思決定が行われるということは極めて稀である。こと、企業における研究開発戦略や、政府に寄る科学技術予算編成に関わるような大規模な意思決定であればより多角的な視点において意思決定がなされる。以上のように、イノベーション・マネジメントにおいてデータ分析を用いた予測は意思決定支援のための材料であり、そのためには予測モデルの精度のみの議論は不十分である。本研究は、人間の意思決定を前提としたイノベーション・マネジメントに資する予測分析を用いた手法と視点の提案である。特にモデルの予測精度のみならず解釈可能性を含有した分析に着目する。本研究の特徴としては、科学技術情報をもつデータ構造から分析対象に応じた本質的な情報を抽出することで、解釈可能

性と精度の両方を実現するための手法を提案することにある。

本論文の社会的貢献は科学技術イノベーション戦略意思決定において情報のに資する指針の提供である。また本論文の学術的貢献は、対象のネットワークから一定の着想を元に構築した特徴量を元にした機械学習的手法の適用による、予測モデルのオリジナリティである。

1.6 本研究の構成

以上に基づき、本論文の構成は下記の通りとする。

第1章ではこの論文の背景・目的について記述する。特にイノベーション・マネジメントにおける意思決定のための予測タスクがこれまで有してきた課題やその解決手段としてこれまで提案されてきた手法の中でも、特に本論文が用いるようなデータ分析にもとづく定量的アプローチにおいて必要となる概念を中心に整理を行う。

第2章は、本論文で用いる科学技術対象について概要とデータセットの抽出条件を示し、同時にそれらの知識動向が動的に変化していることを時系列コミュニティイベントの可視化手法を用いて確認する。

第3章は、科学知識の有望領域の予測に対応する章である。具体的にはナノカーボン分野および太陽光発電分野を対象に将来有望な萌芽領域の予測手法と解釈手法を提示し、それぞれの分野について議論を行う。

第4章は、技術知識同士の融合の予測に対応する章である。特許データが有する国際特許分類を対象に、共起情報を抽出した上で将来的に共起するであろう国際特許分類をリンク予測の観点によって予測する。特に、国際特許分類が有する概念の階層構造に着目し、異なる階層の情報が技術融合の予測に寄与することを示す。

第5章は、知識を社会に実装する経済主体同士の関係性の予測に対応する章である。機械学習を用いた企業活動における取引相手予測の視点では十分に議論されていなかった、モノの流れに関する情報の抽出方法に着目する。業種の条件付き確率を特徴量として導入することで解釈性と精度を両立した潜在的取引相手予測モデルの手法の提案を行う。

第6章では、第3章、第4章、第5章それぞれを用いて実際に得られた結果およびその解釈を整理する。また、これら3手法を統合的に用いることによる付加価値として新たなイノベーション・マネジメントのフレームワークを提案する。

第7章では、科学技術イノベーションプロセスそれぞれのレイヤーで、分野や産業の境界が曖昧になることを改めて主張する。新たな進出先を探索して自社のコア知識を獲得することが困難となるなかで、本研究の提案および視点が有効であることを主張し、本研究の結びとする。

2 第2章：本研究で対象とする知識領域

この章では、おもに本研究で対象とする知識領域について説明を行う。本論文の3章および4章では、特に科学技術の知識構造自体について注目した予測手法の提案を行うものである。それぞれ2つずつのデータセットを元に、知識の発見と融合の発見という予測課題に取り組む。3章で注目する科学レイヤーの対象としては、ナノカーボン分野と太陽光発電分野の論文群を、4章で注目する技術レイヤーの対象として炭素繊維強化プラスチック技術分野と傾斜機能材料技術分野の特許群を分析対象として用いる。これらの科学技術分野は近年注目を浴びており、我が国の研究者や技術者による知的貢献のプレゼンスも大きい分野でもある。本章ではそれぞれのデータセットの概要を示すとともに、これら分野の知識変遷が動的に大きく変化していることを確認することを目的とする。

2.1 知識コミュニティの遷移とイベント

俯瞰的分析の一貫として、引用ネットワークにおける時系列遷移を可視化し、分野の進化現象に着目する。本項では、ネットワークコミュニティの進化に関する考え方を整理する。Palla et al.(2007)によれば、時系列コミュニティ遷移におけるイベントは図1に示す6種類に分類することができる[54]。この分類はその後複数のコミュニティイベント検出に関わる研究に影響を与えている[55, 56, 57]。以降、知識コミュニティの遷移の可視化にはNetgram[57]を用い、このイベント検出は下記定義によって行うものとする。

- Birth (コミュニティの誕生) : 時刻 $t-1$ の時点においては不可視であったノードによって構成される、時刻 t におけるコミュニティ C_{new}^t が発生する現象をコミュニティの誕生とする。すなわち時刻 $t-1$ におけるコミュニティ C^{t-1} の全てのコミュニティ C_i について、 $C_{new}^t \cap C_i^{t-1} = \emptyset$ を満たすコミュニティ C_{new}^t を誕生コミュニティと定義する。
- Death (コミュニティの死滅) : 時刻 t において、1つもしくはそれ以上のコミュニティが消滅する現象をコミュニティの死滅とする。すなわち、時刻 t における C^t に属するコミュニティ C_j^t のうち、 $\frac{|C_i^{t-1} \cap C_j^t|}{|C_i^{t-1} \cup C_j^t|} < \theta^{t-1}$ を満たすコミュニティを死滅コミュニティと定義する。

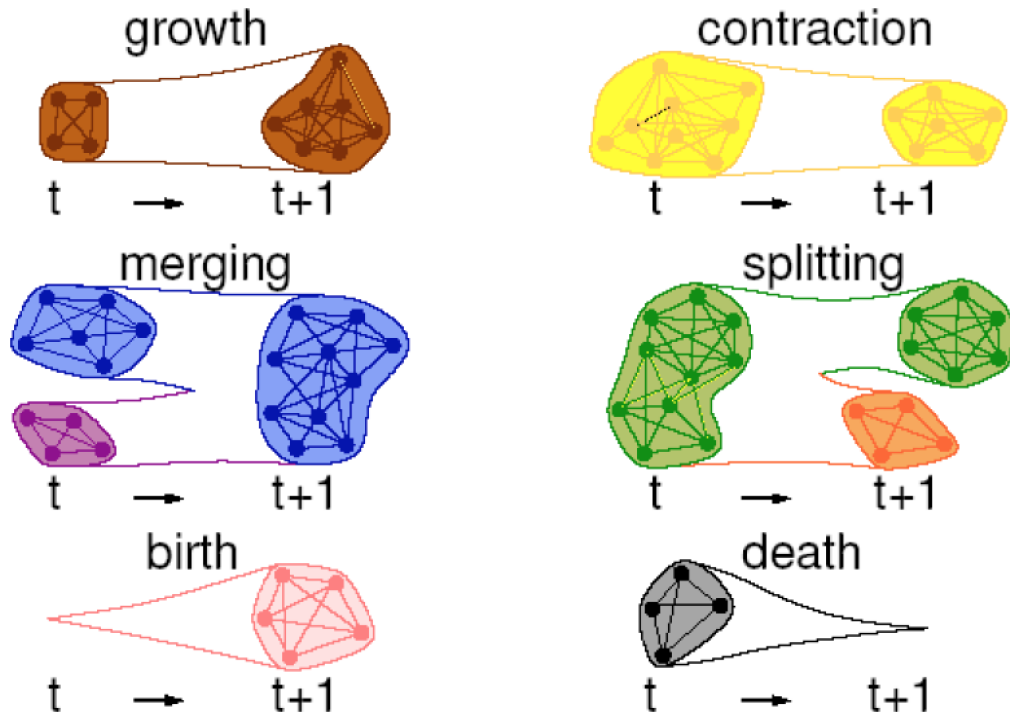


図 1. コミュニティ遷移イベント (Palla et al.(2017))

- Growth (コミュニティの成長) : $t-1$ におけるコミュニティ C_i^{t-1} が、次の時刻 t においても生き残っている状況をコミュニティの継続とする。すなわち、時刻 t における任意のコミュニティ C_j^t が、 $\frac{|C_i^{t-1} \cap C_j^t|}{|C_i^{t-1} \cup C_j^t|} \geq \theta^{t-1}$ を相互の時間方向に維持する場合に、継続コミュニティと定義する。後述するコミュニティの結合とは分けて議論する必要がある。コミュニティ C_i^{t-1} の構造が変化せず、時刻 t においても単一コミュニティ C_j^{t-1} として継続している場合に限り、コミュニティの継続と考える。
- Merging (コミュニティの結合) : 時刻 $t-1$ における 2 つ以上のコミュニティ C_h^{t-1} および C_i^{t-1} の大部分のノードが時刻 t において結合し、新たなひとつのコミュニティ C_j^t を形成した場合、コミュニティの結合とする。すなわち、 $\frac{|C_h^{t-1} \cap C_j^t|}{|C_h^{t-1} \cup C_j^t|} \geq \theta^{t-1}$ かつ、 $\frac{|C_i^{t-1} \cap C_j^t|}{|C_i^{t-1} \cup C_j^t|} \geq \theta^{t-1}$ を満たす場合、 C_j^t を結合コミュニティと定義する。これらのコミュニティは、時刻 t 以降は共通したタイムラインを共有することになる。
- Splitting (コミュニティの分離) : 時刻 t において、ひとつ前の時刻のコミュニティ C_i^{t-1} を構成していた大部分のノードが 2 つのコミュニティ C_j^t と C_k^t に別れる状況を、コミュニティの分離と定義する。すなわち、

$$\frac{|C_i^{t-1} \cap C_j^t|}{|C_i^{t-1} \cup C_j^t|} \geq \theta^{t-1} \text{ かつ } \frac{|C_i^{t-1} \cap C_k^t|}{|C_i^{t-1} \cup C_k^t|} \geq \theta^{t-1}$$

- Contraction (コミュニティの衰退) : 時刻 t におけるコミュニティ C_j^t に属するノードが、時刻 $t-1$ における対応するコミュニティ C_i^{t-1} と比べて急激に減少した場合をコミュニティの衰退と定義する。(例えば 20% 程度などのしきい値に基づく)。パラメータ θ^{t-1} は各タイムスタンプにおいて定義される。

これらのコミュニティイベントを可視化する時系列ネットワーク可視化手法を適用し、それぞれの知識領域(学術分野/技術分野)における動的な知識構造の変遷を確認する。学術分野のデータセットでは、学术论文をノードとし、相互の直接引用をリンクとした時間拡張ネットワークを作成する。一方、特許公報における直接引用のみのネットワークは一般的に疎であることが知られており、十分な知識のネットワークを構築することが難しい。したがって、技術分野のデータセットでは、特許公報をノードとし、直接引用、共引用、書誌結合、の全てをリンクとした時間拡張ネットワークを作成する。これらの時間拡張ネットワークに対して各年次におけるクラスタリングを実施しコミュニティを特定する。クラスタリング手法は Louvain 法を用いる。

それぞれの知識領域において、コミュニティイベントの可視化および各年前後遷移行列の可視化といった2つの可視化手法によりコミュニティイベントを特定する。コミュニティイベントの可視化では、各年上位 10 コミュニティ同士におけるコミュニティがそのコミュニティサイズに応じた大きさの円として図示される。各年前後の遷移行列の可視化においてはそれぞれのタイムスタンプ前後におけるコミュニティ同士の類似度をヒートマップとして表している。縦軸は T 年におけるコミュニティに相当し、横軸を $T+1$ 年におけるコミュニティに相当する。 T 年と $T+1$ 年におけるコミュニティの類似度の度合いをそれらの交点のセルの色によって示す。類似度が近いコミュニティ同士が暖色(黄色)に、類似度が低いコミュニティは寒色(青色)としている。

2.2 本研究で対象とする知識領域

本項では、本論文全体において用いるデータセットとなる学術分野、技術分野における書誌情報についてその概要や歴史的背景について理解する。具体的には、ナノカーボン分野、太陽光発電分野、炭素繊維強化プラスチック分野、傾斜機能材料分野の4つの分野を用いる。

2.2.1 ナノカーボン分野

ナノカーบอนはグラファイト（黒鉛）を母材とする物質であり、その機械的、熱伝導的、電気的な特徴から近年は、半導体、燃料電池、光学機器、構造材料といった多様な用途に実用化が進みつつある。ナノカーボン材料は、カーボンナノチューブ、グラフェン、およびフラーレンで構成された、グラファイトから作られた材料を指す語である。カーボンの歴史は長きに渡るにもかかわらず、これらの構造は短期間で連続して現れた。まず、1985年に、Kroto et al.(1985)がC60分子（フラーレン）を発見した [58]。1991年に飯島はナノチューブを発見し、2004年にNovoselovらはグラフェンを発見した [59]。これらは、個別の特徴を有する0次元、1次元、および2次元のナノカーボン物質であり、材料の様々な用途が期待される。ナノカーボン材料は、その優れた機械的、電気的、および熱的特性のために、半導体、燃料電池、光学デバイス、および構造材料などの多様な用途に使用されている。例えば、エネルギー分野 [60, 61, 62, 63, 64] や宇宙エレベータにナノカーボン材料を使用する可能性について議論されている [65, 66, 67]。

2.2.1.1 データ抽出

Web of Science より、1900年01月01日から2015年07月27日に出版されたジャーナル論文のうち、タイトルもしくはアブストラクトを対象にクエリ ((carbon and (nano* OR micro*)) or fullerene or Buckminsterfullerene or Buckminster-fullerene or C60 or C-60 or graphene or (filament* and carbon)) に該当する論文 411,084 本を抽出した。最も古い論文は***年に出版されたものであった。図 2 に 1900 年以降の各年の出版数のプロットを示す。この図から、1991 年からは出版数が急増していることが見て取れる。2014 年には一年間に 45,000 本以上の論文が出版されていることが分かる。

2.2.1.2 コミュニティの動的遷移の可視化

図 3 に、ナノカーボン分野の 2003 年から 2015 年までのコミュニティ遷移を可視化する。また、図 4 には、各年前後の遷移を可視化したものを示す。ここ数年で新たなコミュニティが (New C11 等) 生まれつつあることが確認できる。

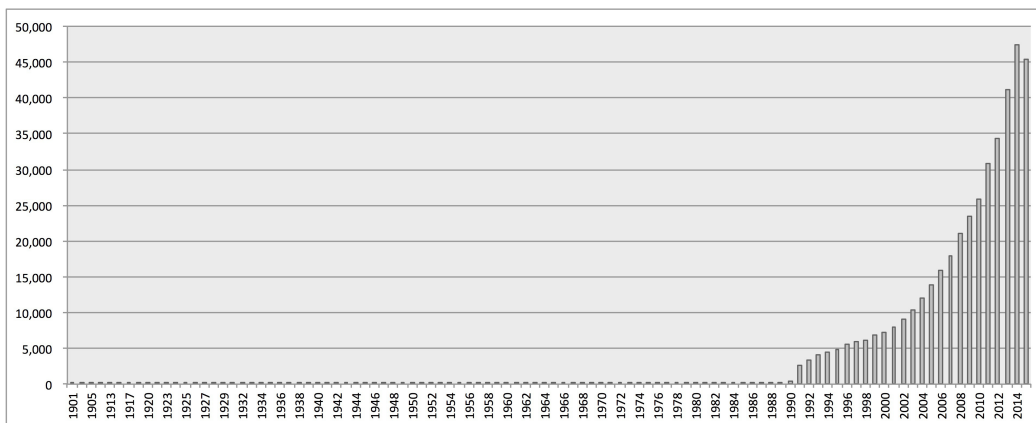


図 2. ナノカーボン分野関連論文の出版数推移

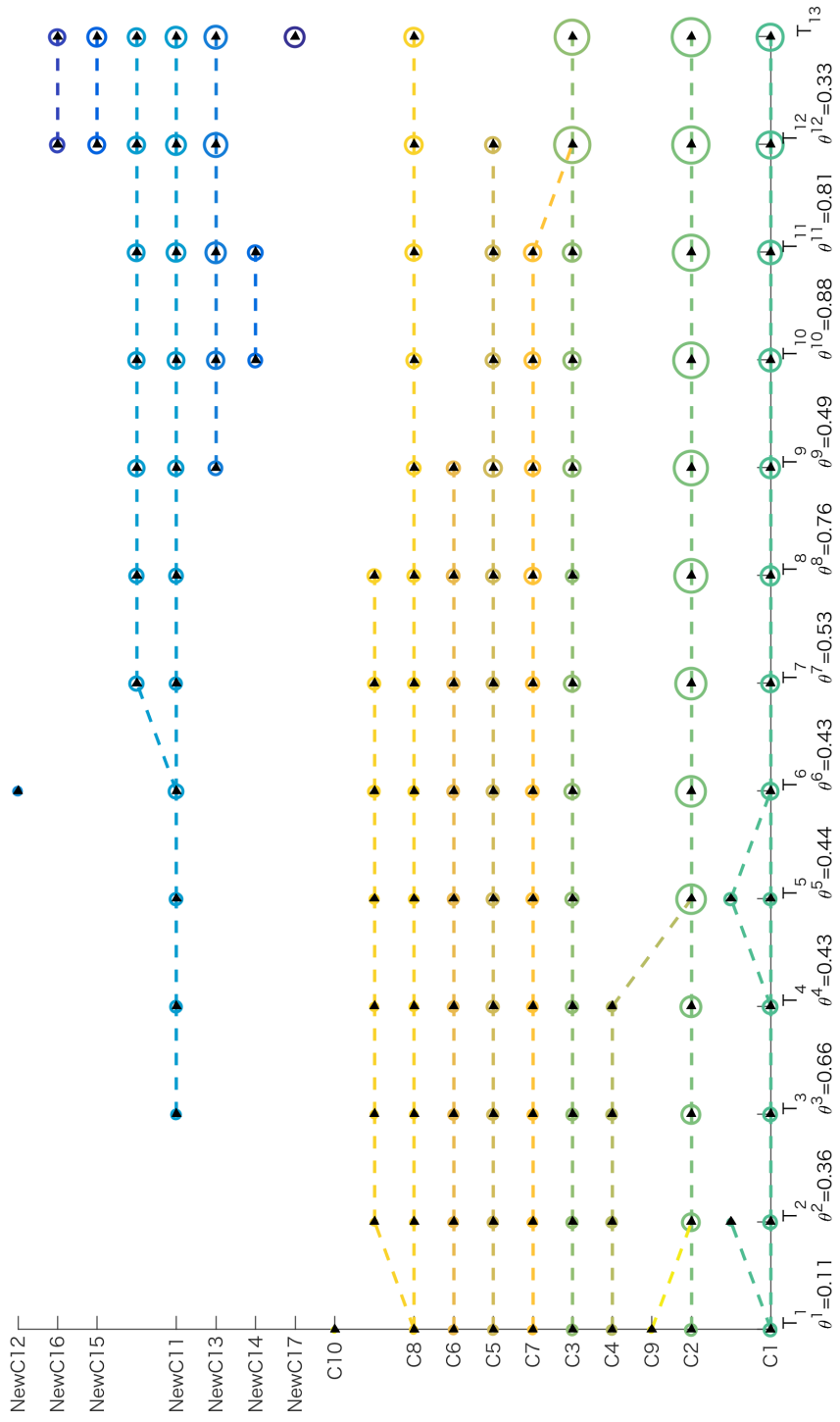


図 3. ナノカーボン分野関連論文 (2003 年から 2015 年) における各年上位 10 コミュニティのコミュニティイベント

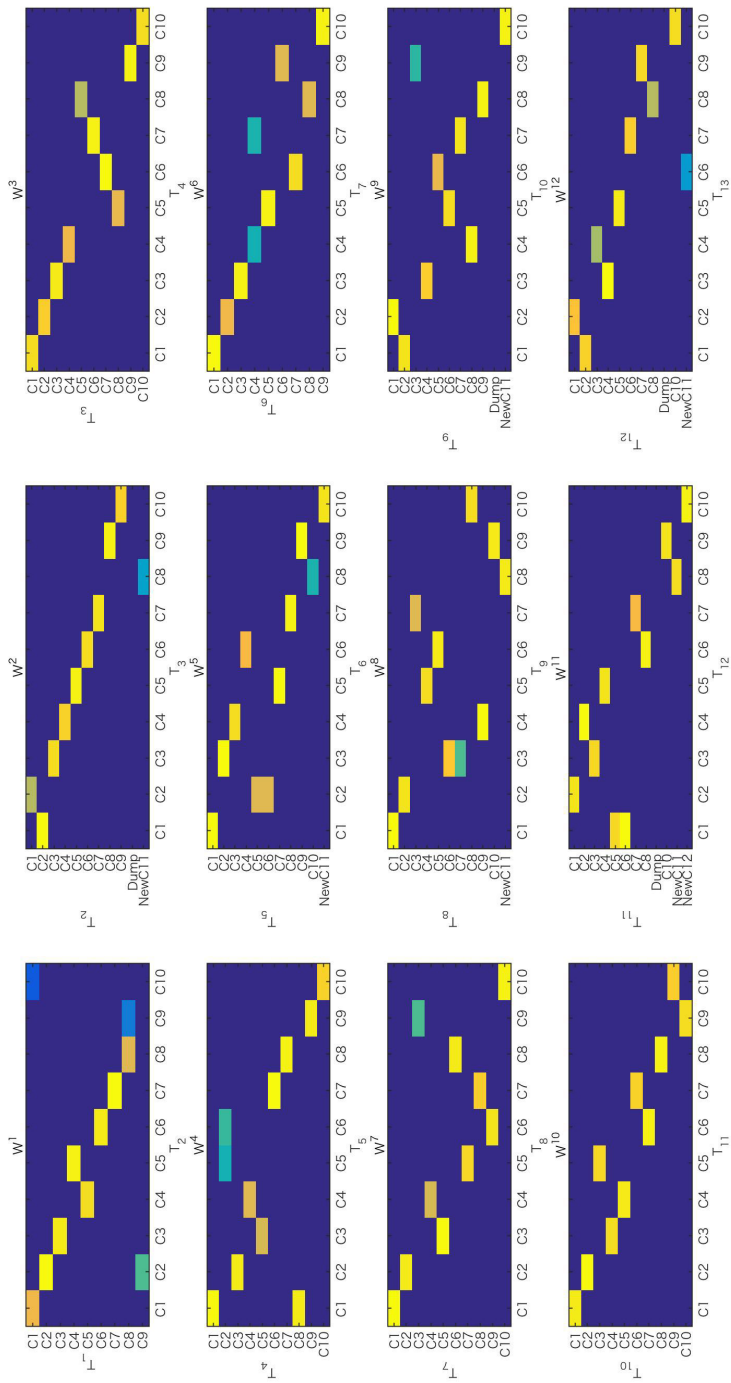


図 4. ナノカーボン分野関連論文 (2003 年から 2015 年) の各年前後の遷移行列

2.2.2 太陽光発電分野

光起電力効果は Edmond Becquerel に最初の実証された効果であり、その後、Chares Fritts によりセレン半導体に金の薄膜をコーティングして接合部を形成したものが最初の個体太陽電池とされた。太陽光発電分野太陽電池の研究は、光起電力効果、光熱効果および光電気化学太陽電池エネルギー変換に関連する研究のみならず、太陽電池セルからなるモジュール、モジュールから構成されるパネルを組み合わせ、そこにアイソレータ、ヒューズ、バッテリー、メータなどを組み合わせることでシステムを対象とするものまで幅広い。学術分野において太陽光発電分野を対象とした研究はセルの構成や電極周辺の基礎的研究から応用研究まで幅広い。

2.2.2.1 データ抽出

Web of Science より、1900 年 01 月 01 日から 2015 年 07 月 27 日に出版された Journal 論文のうち、タイトルもしくはアブストラクトに "solar cell" もしくは "photovolt" が含まれる論文 121,393 本を抽出した。図 5 に 1900 年以降の各年の出版数の推移を示す。最も古い論文は 1906 年に出版されたものであった。この図から、1990 年代からは毎年指数関数的に出版数が急増していることが見て取れる。2015 年では一年間に 18,000 本以上の論文が出版されていることが分かる。

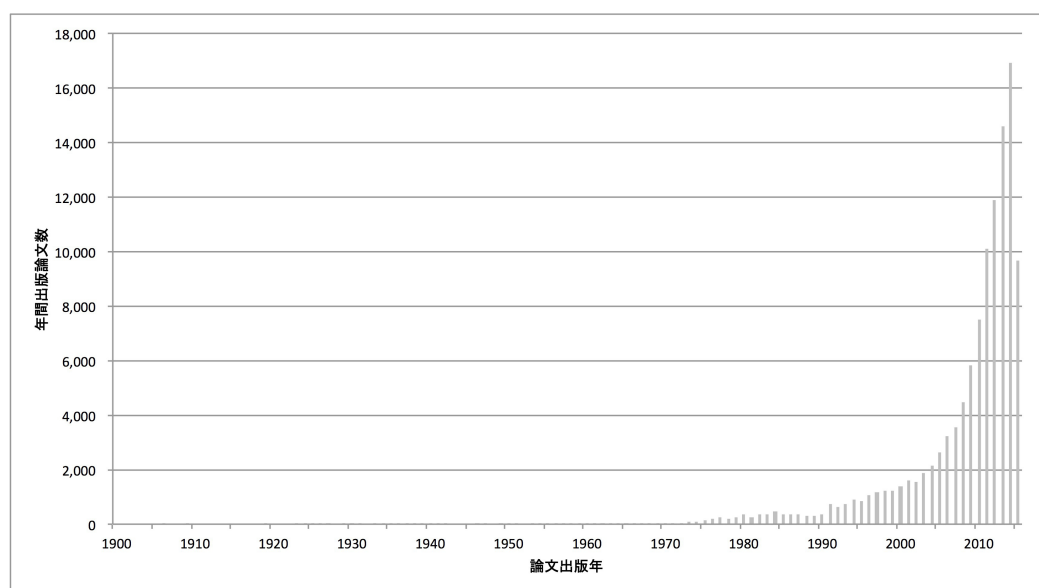


図 5. 太陽光発電分野関連論文の出版数推移

2.2.2.2 コミュニティの動的遷移の可視化

図 6 に、太陽光発電分野の 2003 年から 2015 年までのコミュニティ遷移を可視化する。また、図 7 には、各年前後の遷移を可視化したものを示す。たとえば 2014 年 (T12) と 2015 年 (T13) の間は極端に変化していないが、2006 年 (T4) と 2007 年 (T5) の前後では動的に大きく変化している様子が見て取れる。

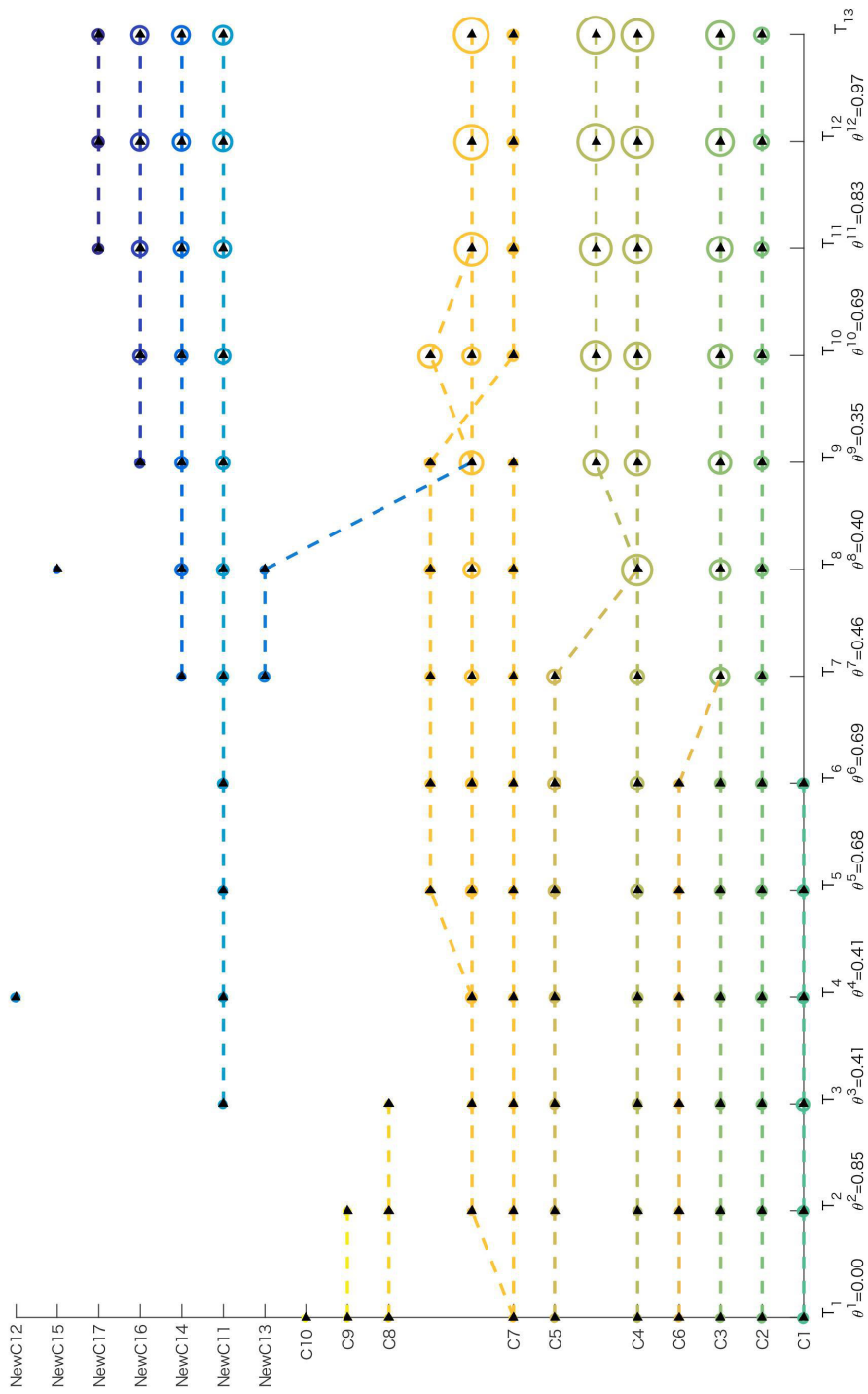


図 6. 太陽光発電分野 (2003 年から 2015 年) における各年上位 10 コミュニティのイベント

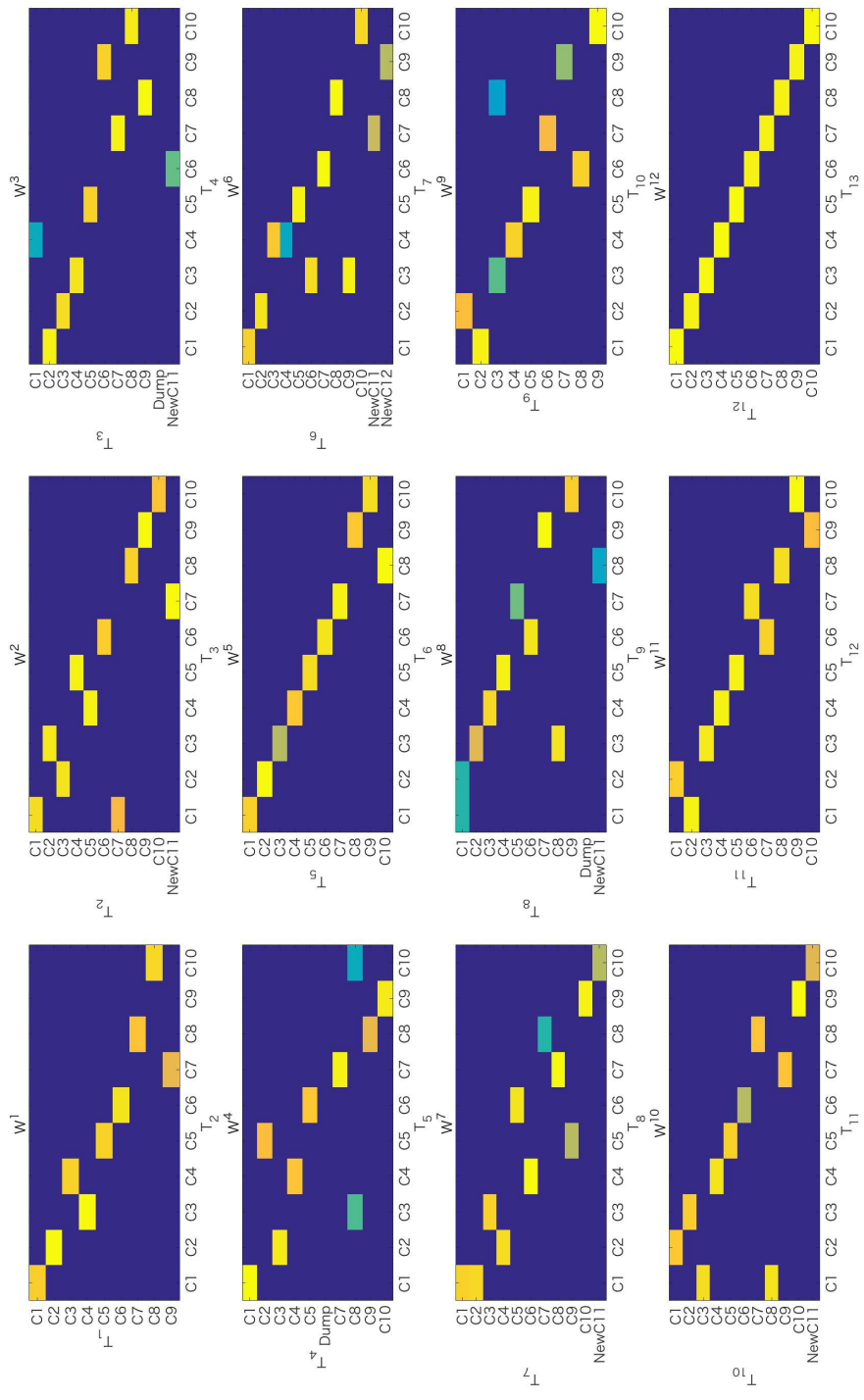


図 7. 太陽光発電分野 (2003 年から 2015 年) における各年前後の遷移行列

2.2.3 炭素繊維強化プラスチック技術分野

炭素繊維強化プラスチック（CFRP）は炭素繊維を含む複合材料であり低密度、高強度、校比強度、高比弾性率および高剛性の面で優れた特性を有するプラスチック材料である。特に繊維の方向に高い弾性率を有する特徴を持っており繊維方向の割合を変えることによって調整することもできる。このような特性を有することで、設計や工業デザイン上これまで単一の材料では困難であったような形状やニーズにも対応できる材料である。CFRP は当初、大規模生産が可能ではなかったこともあり広く流通されるものではなく、ゴルフクラブや釣り竿のようなスポーツ用品のうち高級なものに限定して用いられていた。1990 年ごろより徐々に航空機産業において主翼の材料として用いられ、自動車産業においてもボディや内装の材料に用いられるようになり、今や建築や橋梁の耐震補強材としても用いられているものである。トーマス・エジソンとジョセフ・ウィルソンスワンは、19 世紀に綿と竹を燃やしてフィラメントを生成していた。その観点では彼らは炭素繊維のパイオニアでもある。しかしながらタングステンフィラメントの出現の後、このような炭素繊維の存在は人々の生活から忘れられてしまった。1950 年代頃になり、非常に高い耐熱性が要求されるロケットノズルの材料に適用されたことで、人類は再度この炭素繊維の可能性に気づいたという経緯をもつ材料である。

2.2.3.1 データ抽出

本研究では、米国、欧州、日本の 3 極機関に出願された炭素繊維強化プラスチックに関する特許を対象とする。Thomson Innovation より、1970 年 01 月 01 日から 2016 年 12 月 31 日に出版された特許公報のうち”carbon fibre” もしくは ”carbon fiber” を含みかつ、Polymer もしくは resin もしくは plastic を含みかつ composit* もしくは cfrp をテキスト情報にもつ特許をトムソンイノベーションより DWPI の形式で抽出した結果 18,924 件の特許公報を得られた。

2.2.3.2 コミュニティの動的遷移の可視化

図 8 に、炭素繊維強化プラスチック技術分野の 2004 年 (T1) から 2016 年 (T13) までのコミュニティ遷移を可視化する。また、図 9 には、各年前後の遷移を可視化したものを示す。

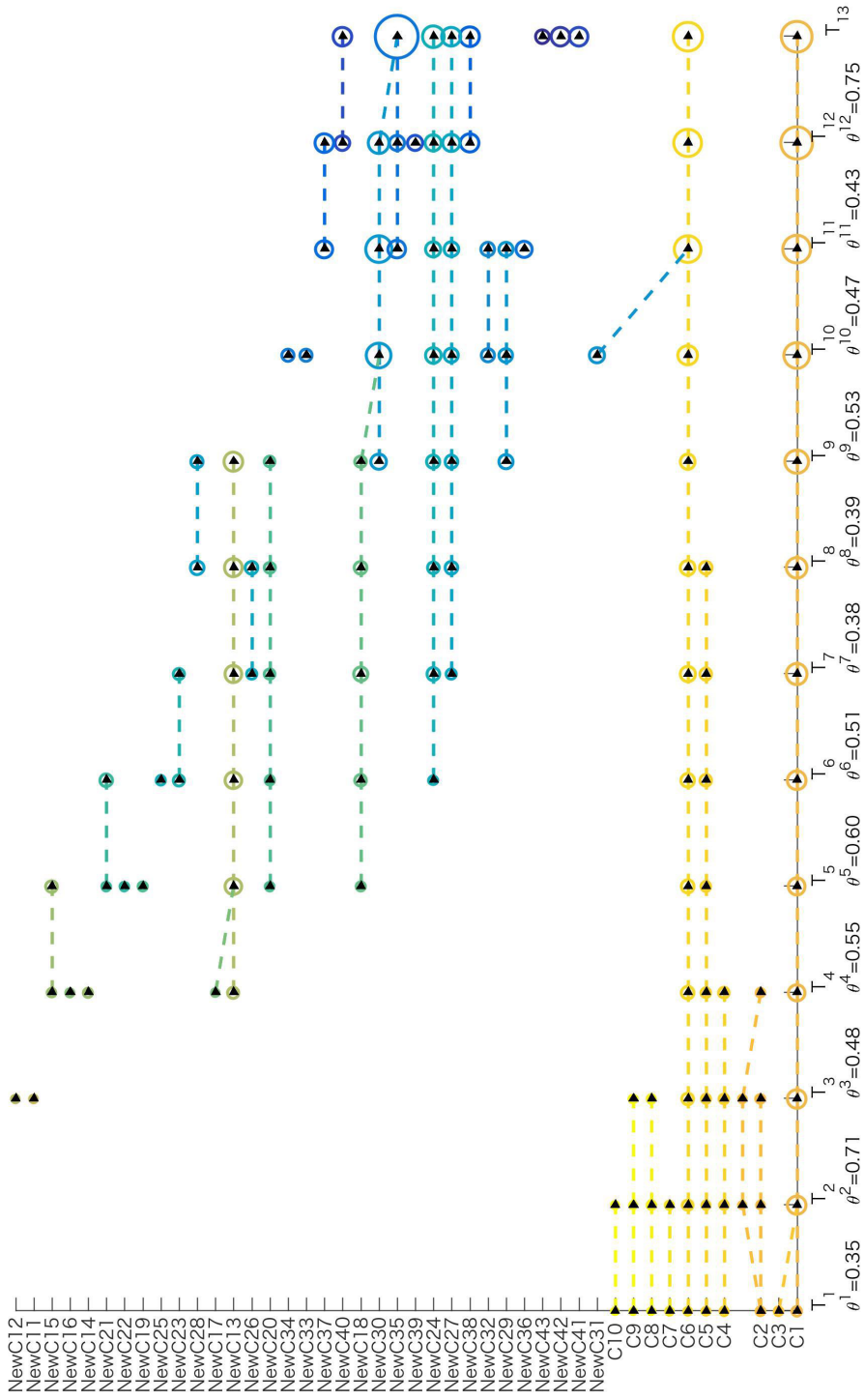


図 8. 炭素繊維強化プラスチック技術分野 (2004 年から 2016 年) における各年上位 10 コミュニティのイベント

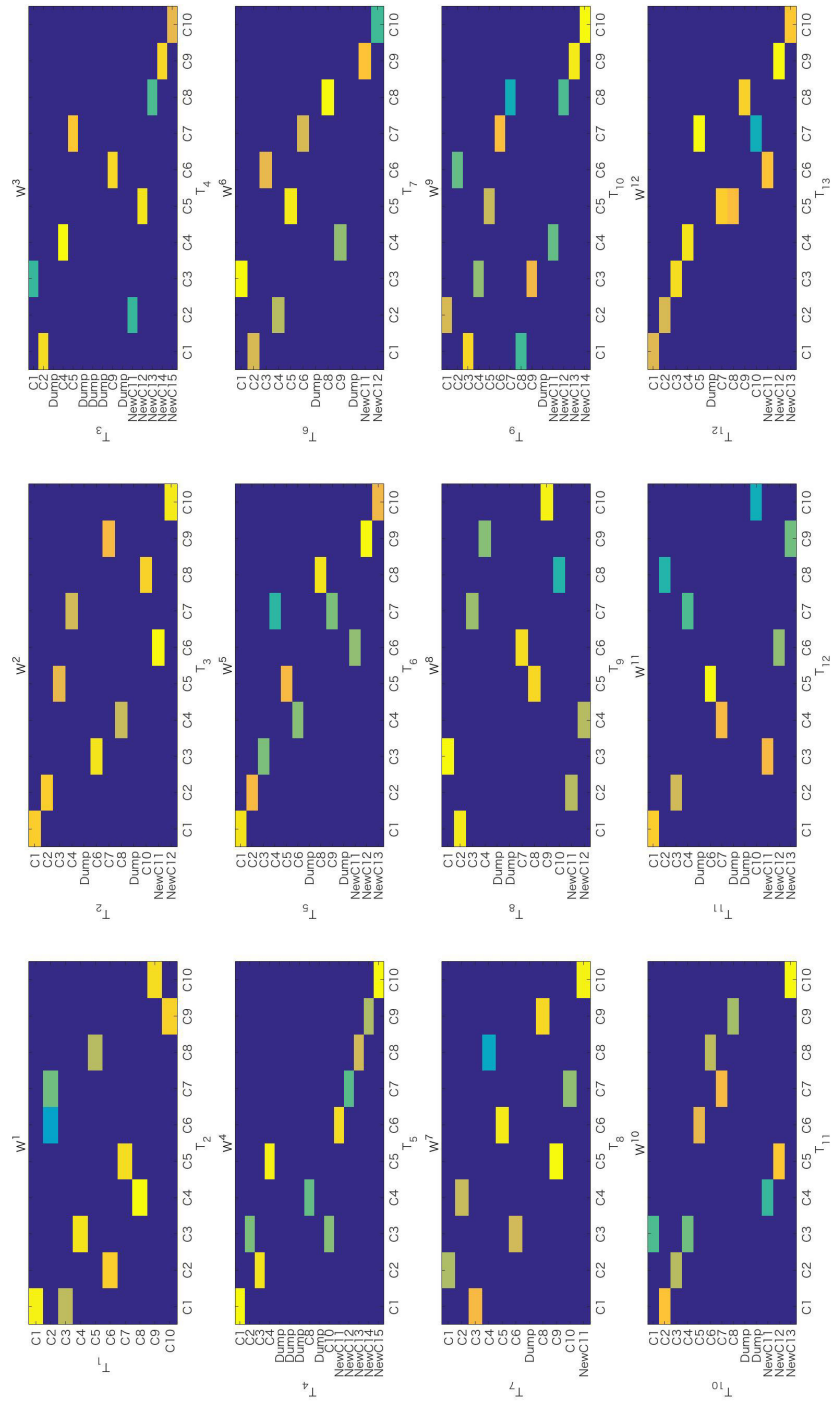


図 9. 炭素繊維強化プラスチック技術分野（2004 年から 2016 年）における各年前後の遷移行列

2.2.4 傾斜機能材料技術分野

傾斜機能材料 (FGM) は宇宙工学分野において日本で発明された材料の一つである。スペースプレーンは宇宙からの再突入時にノーズコーンや翼のリーディングエッジは断熱圧縮により生じる極めて高い空力加熱を受ける。機体表面は非常に高い温度 (摂氏約 1,700 度) にさらされる。内側との温度差が 1,000 度という過酷な条件に耐える必要があるが、このような条件に耐える材料は単体では存在しない。平井敏雄らは高温にさらされる面を耐熱性のあるセラミックスに、内側の面を熱伝導度が大きい材料に貼り合わせることを考え、また接着面の材料を混ぜり合わせることで境界をなくすことを提案した。このようにして考えだされた材料が傾斜機能材料である。図 10 に、傾斜機能材料の例の概念的モード図を示す。この例においては、面 X においては強度が高いが耐熱性は低い。一方で面 Y においてはその特性が逆転する。さらに面 X と面 Y の間に、熱疲労に対する特性が高い箇所を設けるような材料である。

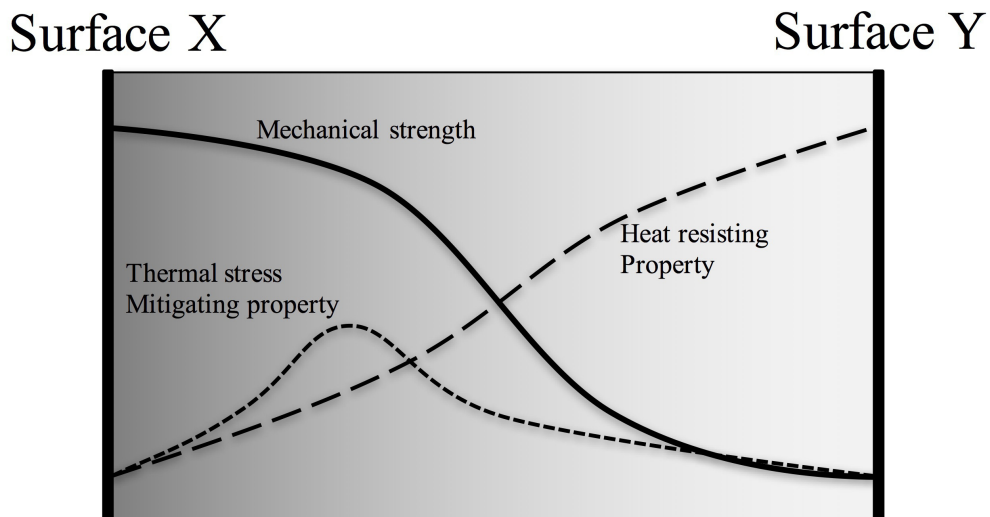


図 10. 傾斜機能材料概念図

2.2.4.1 データ抽出

本研究においては、1978 から 2015 年における特許情報のうち米国、欧州、日本の 3 極機関に出願された傾斜機能材料に関する特許を対象とする。具体的には、Thomson Innovation よりクエリ”func* NEAR grad* NEAR material*”によって抽出された特許公報とする。

2.2.4.2 コミュニティの動的遷移の可視化

図 11 に、炭素繊維強化プラスチック技術分野の 2003 年から 2015 年までのコミュニティ遷移を可視化する。また、図 12 には、各年前後の遷移を可視化したものを示す。

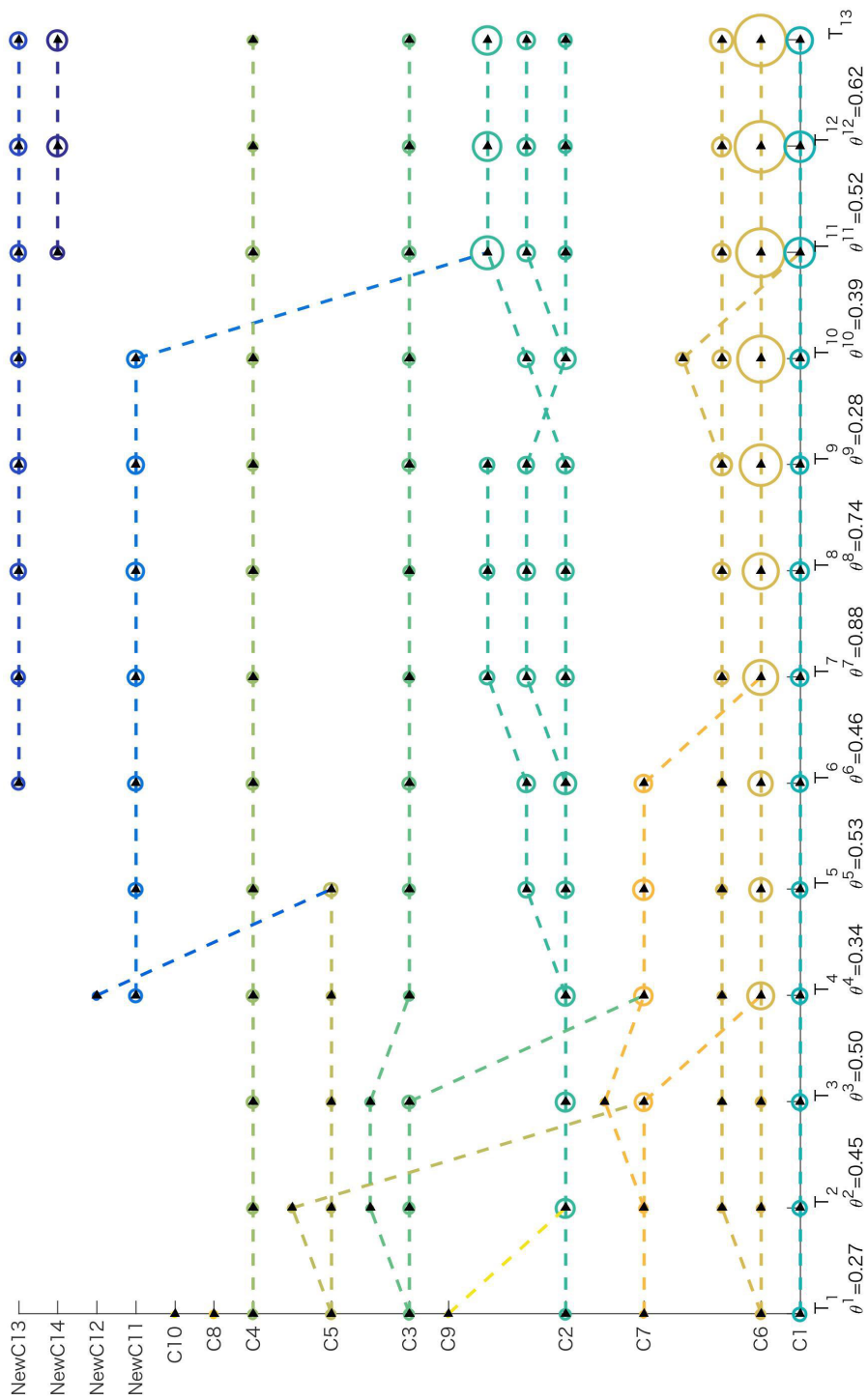


図 11. 傾斜機能材料技術分野（2005 年から 2017 年）の各年上位 10 コミュニティイベント

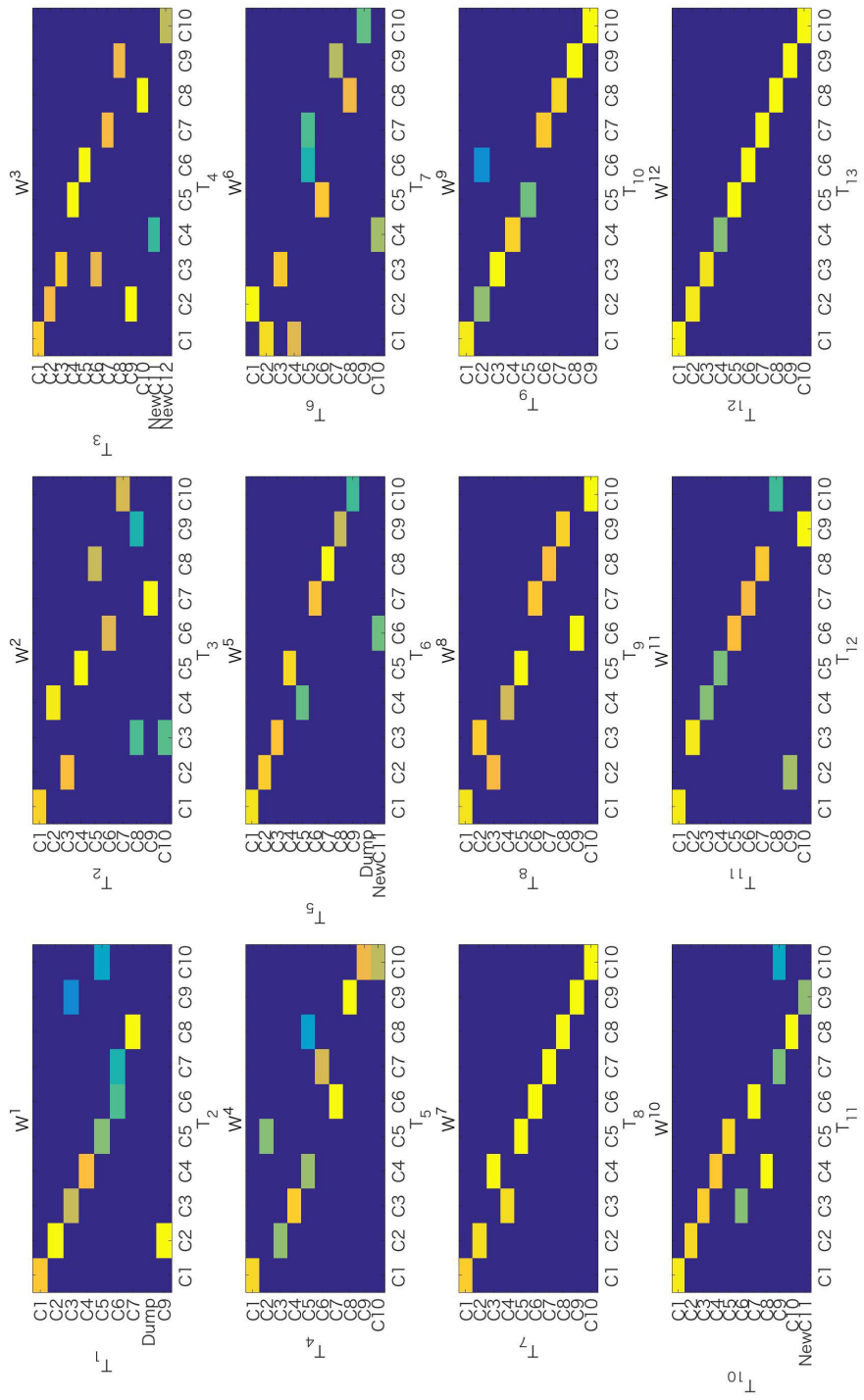


図 12. 傾斜機能材料技術分野（2005 年から 2017 年）における各年前後の遷移行列

3 第3章：ナノカーボン分野と太陽光発電分野における萌芽領域予測

3.1 背景と目的

企業が今後の技術開発の方針を策定するには、学術研究の動向は大きな助けとなることが少なくない。特に太陽光発電分野やナノカーボン分野のようなサイエンスリンクエッジの高い分野においてより有用であることは前述のとおりである。技術分野動向の策定において行われる手法のひとつとしてはデルファイ法のような専門家の知識を集約し合意形成を得ることで技術予測を行う手法がよく用いられるが、こういった手法には近年課題が指摘されている [68, 69]。特に専門知識が属人的になりつつあり、細分化が進んだ分野においては参加者の人選次第で結果が変わってしまいかねない。同時に、多くの自然分野に関する情報の増加速度が指数関数的に増えつつあり、当該領域の全体像を理解することが日に日に困難になってきているということである [70]。このようにデータが大規模になりつつあるなか、科学技術の全体像を把握しようとする試みが行われている [71, 72, 73, 74, 75, 76, 77]。一方で、将来どのような研究が注目されるかという視点に一定の注目が集まっている [76, 77, 78, 79]。近年では、統計的機械学習に代表されるような予測アルゴリズムの高度化ならびに計算能力の向上に伴い、過去のデータから将来を予測するに資するパターン抽出を行うことが可能になりつつあり、より一層科学技術予測の実現への挑戦が盛んである [79, 80]。本研究では、出版後に被引用数を稼ぐ可能性の高い論文を萌芽研究と呼称し、そのような論文を出版時点で特定するような手法を具体的な分野、すなわち太陽光発電分野とナノカーボン分野に適用することで将来的に注目されるであろう研究成果について議論を行うものである。

学術や技術における萌芽領域の予測あるいは特定に関する先行研究は、豊富に存在する。萌芽論文の予測は従来、計量書誌学あるいは図書館情報学で研究がなされてきた。近年はビッグデータの潮流から、大規模な論文データを対象に、コンピュータ科学、特に、データマイニング、情報検索、などの分野においても研究がなされてきている。以下に、直近の代表的な研究を紹介する。Winnink and Tijssen(2015) は、グラフェンに関する分野におけるノーベル賞論文に対して、書誌情報に基づいた萌芽予測可能性を検証した [81]。Adams(2005) は生命科学や物理の研究分野において、論文出版後3年から10年における被

引用件数は当該論文出版後 1 年から 2 年の被引用件数と相関があることを示した [82]。情報伝播を疫病の伝搬としてモデル化するというアイデアとしては (Goffman and Newill, 1964) が著名である [83]。また、Bettencourt et al.(2008). はある分野において新たに現れた分野が伝搬していく現象を、疫病の伝搬モデルである SIR モデルを用いて記述した [84]。Chen et al.(2009). は、学術論文における共引用ネットワークと共同研究ネットワークを用いて、ネットワークの構造的空隙にまたがる研究に着目した [85]。Li and Tong(2015) は論文引用数の予測を最適化問題として定式化した。著者らはコンピュータ科学系の 50 万論文を対象とし、論文出版後 3 年の情報を元に 10 年後の引用数を予測した [86]。この研究では、出版後 3 年の間にいかに引用されるかが予測に重要との考察が得られた。Dong et al.(2015) は、論文出版後 5 年後の著書の h-index を予測した [87]。この研究では、論文のインパクトを著者、内容、出版先、引用、共著、時系列の 6 つの要因で定義している。データセットとしてコンピュータ科学系の 200 万論文データを使用している。Davletov et al.(2014) は、論文出版後数年の引用数の時系列情報と引用ネットワークの構造情報を用いて 5,10 年後の引用数を予測した研究である [88]。データセットとして、arXiv のエネルギー物理分野の 27,000 論文データ、コンピュータ科学系の 150 万論文データ (ArnetMiner) ならびに 200 万論文データ (CiteSeerX) を使用している。著者らによれば、出版後 2 年間の引用数の時系列情報が予測に重要である [89]。Chakraborty et al.(2014) は、論文出版後数年の引用数の時系列情報を 6 つに類型化し、著者、学会、キーワードの特徴量と合わせて、5 年間以内の引用数を予測した研究である。データセットとしては、コンピュータ科学系の 150 万論文データを使用した [90]。著者らによれば、論文著者本人の論文数や出版後 1 年の引用数が予測に重要である。Wang et al.(2013) は、著者らは論文の引用数の冪乗則に着目し、出版後 5 年の引用数の時系列情報から将来の論文の引用数予測を定式化した [91]。データセットには Physical Review B, PNAS, Cell の 3 つの論文誌の論文データを使用し論文の 25 年後の引用数について、9 割の論文が予測結果と一致した結果が得られている。森ら (2014) はネットワーク情報とテキスト情報、クラスタ情報などを用いて、被引用数の増加の観点から萌芽論文の予測を試みている研究として人工知能分野に関連する学術論文を対象にしている [92]。同様の手法を用いて、Sasaki et al.(2016a) では太陽光発電分野を対象に、Sasaki et al.(2016b) ではナノカーボン分野を対象にした萌芽的論文を抽出する試みを行っている。[93, 94] これらの研究の多

くにおいて、引用ネットワークを用いていることが共通している。引用ネットワークは知識の繋がりを意味するものであり、すくなくとも学術においては知識のうえに知識が成り立っているという前提である以上、より多くの引用を集める学術論文は新たに生まれた多くの知識に影響を与えた論文にほかならない。しかしながら引用関係には時間遅れが生じることも指摘されている。そのような短所を補う目的として、引用関係を用いずに書誌情報から科学技術単語を抽出しトピック情報を用いて分野の全体像を把握しようとする試みも行われている。いずれにせよこのような手法によって、全体像の把握は、過去から現在まで当該領域において行われてきたことを俯瞰するという点については一定の指針を提供することができる。Kajikawa et al.(2007) は太陽光発電に関する論文を収集し、学術知識の俯瞰的状況を構築し、当該分野が大きくわけて4つのクラスタに分類されることを示した [95]。Lizin et al.(2013) は有機系太陽電池に関する特許データを元に、知識の俯瞰的状況の構築をしている [96]。Sakata and Sasaki(2013) は太陽光発電分野に関する論文群から、各国の出版動向を分析し、アジア各国がこの分野において急激なキャッチアップを見せていることを示した [12]。Shibata et. al.(2010) は太陽光発電分野を対象にした学術論文と特許公報の書誌データを用いて、研究はされているが特許として十分に成長していない分野、を特定することで今後の技術動向に洞察を得るなどを行っている [97]。ナノカーボン（グラファイト（黒鉛）を母材とする物質であり、その機械的、熱伝導的、電気的な特徴から近年は、半導体、燃料電池、光学機器、構造材料といった多様な用途に実用化が進みつつある。カーボンナノチューブ高分子複合材料における滲出に関するレビュー論文 [98]、カーボンナノチューブの化学蒸着におけるレビュー論文 [99]、リチウムイオン電池の電極材料としてカーボンナノチューブを適用する領域のレビュー論文 [100] など、ナノカーボンに関するレビュー論文は枚挙に暇がない。これらの論文はいずれも将来動向に関する記述があれど、ナノカーボンに関する全てを対象にして議論しているものではない。太陽光発電分野や、ナノカーボン分野に代表されるような、技術と科学の距離を表すサイエンスリンケージが高い分野では、今後の技術動向を捉えるためには学術論文の動向が大きな指針となる。

これまでの萌芽論文の予測、特に論文の引用数あるいは影響度（インパクト）の予測や研究は、基本的に論文出版後数年の引用数の時系列情報に基づき予測を行う事が多い。本研究では、早期に萌芽領域を特定するために、論文が出版された直後少なくとも1年未満の情報のみから、近い将来具体的には3年後

の当該論文の引用数の伸びの大きい論文を特定することを目的とする。本研究では論文が発表された直後の時点での特徴量から、数年先の引用数の伸びを予測している点も既存研究と異なっている。本論文の目的は、萌芽論文を早期に検知する手法にもとづき、太陽光発電分野とナノカーボン分野に適用することでそれらの分野の動向における知見を抽出し、イノベーション・マネジメントの観点に役立てることである。本手法は、3.2.1 から 3.2.4 までは、Sasaki et al.(2016a) および Sasaki et al.(2016b) による手法 [93, 94] と共通しており、3.2.5 以降においてそれら情報をより面的にとらえた分析を行うものである。

3.2 手法

3.2.1 モデル構築

得られた論文の書誌情報ならびに引用ネットワークから、各論文の特徴量を抽出する。ここで言う特徴量とは、萌芽論文を予測するための学習データであり説明変数として用いられる。特徴量の詳細は後段に記述する。本研究では引用ネットワークの最大連結成分に含まれる論文に対して全ての特徴量を算出し説明変数として扱う。また、当該論文が萌芽論文であるか否かを（被説明変数）とする。本研究における萌芽論文の定義は、当該論文が出版された年から3年後における引用数の増加が、データセット全体の上位5%に含まれる論文と定義した。実際に被引用数増加率上位5%に含まれる萌芽論文であれば正例としをフラグとして付与し、50%以下を負例として被説明変数として扱う。すなわち、本研究における萌芽予測問題は、出版から3年以内に萌芽論文になるかならないかといった2クラス分類の問題として捉える。分類機には線形分類器であるロジスティック回帰を採用し実装には LIBLINEAR を用いた。負例に含まれるデータのうち、正例と同じデータ量をランダムに8回抽出し、各年で8種類のデータセットを構築した。また、それぞれのモデルで、5-fold cross validation を行うことで過学習を回避する。予測モデルは、出版から3年経過した時点で萌芽論文となったかそうでないかを学習データとし、実際には学習データの出版年の4年後に出版された論文群を予測対象として適用することとなる。すなわち、 $t_1(=t_0+4)$ を予測対象論文の出版年とした場合、学習窓である t_0+3 年の学習データをもって $t_1(=t_0+4)$ の論文に適用する。また、このモデルは予測対象モデルの出版年からさらに3年経過後 (t_1+3) に評価することができる。この期間を評価窓と定義する。学習窓と評価窓の関係の模式図を図13に示す。

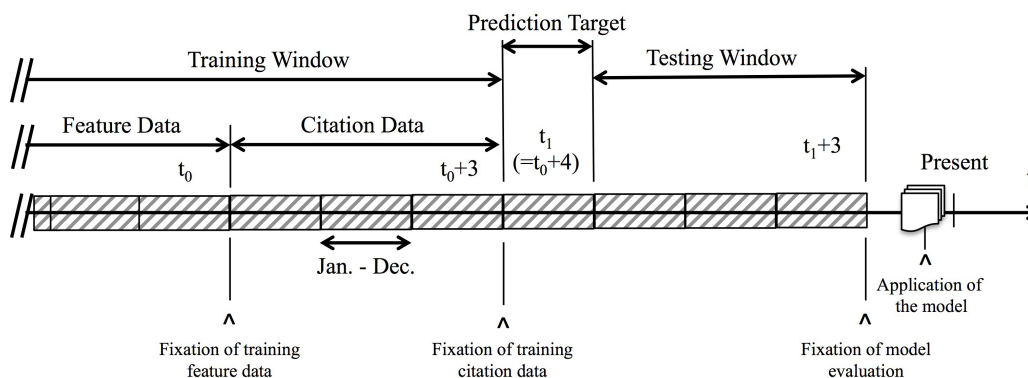


図 13. Model training and prediction

例えば、2012 年 (t_1) に出版論文を予測対象論文とした場合、予測モデルは 2008 年 (t_0) に出版された論文の 2011 年 (t_0+3) 時点における被引用数伸率をもって構築される。このモデルを便宜上 2008 年モデルと呼称する。この 2008 年モデルを 2012 年 ($t_1=t_0+4$) 出版の論文群に適用することで、2012 年を対象とした予測を行う。この 2008 年モデルを 2012 年出版論文データに適用した結果の評価は、さらに 3 年後の 2015 年 (t_1+3) に可能となる。表 2 に各モデルの学習と評価可能年次を対応させた表を示す。

表 2. 各モデルの学習年と評価年の組み合わせ

モデル年 (t_0)	学習窓 正解データ取得年 ($t_0 + 3$)	評価窓 予測対象年 (t_1)	モデル評価可能年次 ($t_1 + 3$)
2003	2006	2007	2010
2004	2007	2008	2011
2005	2008	2009	2012
2006	2009	2010	2013
2007	2010	2011	2014
2008	2011	2012	2015

3.2.2 特徴量抽出

本研究における特徴量は 4 つのカテゴリ (ネットワークマクロ特徴量、クラスター特徴量、ネットワーク中心性特徴量、引用論文特徴量) に分類することができる。ネットワークマクロ特徴量は、対象となる引用ネットワークの代表的な特徴量であり、具体的には、ネットワークに含まれる論文数 (NW_NODES)、

ネットワークに含まれる総引用数 (NW_EDGES)、ネットワークに含まれるクラスタのモジュラリティ (Q 値) の最大値 (NW_MAXQ) である。ここで言うクラスタとは、引用ネットワークにおけるモジュールのことを意味し、モジュラリティ最大化手法によって分割された引用関係が密に凝集した論文群のことを意味する [101]。モジュラリティ (Q 値) は、クラスタ内とクラスタ間の結合度合いの評価関数であり、下記で与えられる。

$$Q = \sum_i (e_{ii} - a_i^2) \quad (1)$$

この時、 e_{ii} は同じクラスタ i に属するノードが繋がるリンク数のネットワーク全体のリンク数に占める割合であり、 a_i^2 は e_{ii} の全リンク数に占める割合の期待値である。クラスター特徴量は、対象論文が所属するクラスタに属する特徴量である。具体的には、対象論文が所属するクラスタの Q 値の最大値 (CL_QMAX)、対象論文が所属するクラスタのノード数 (CL_NODES)、対象論文が所属するクラスタの順位 (CL_RANK) を用いる。ネットワーク中心性特徴量は、任意の論文が引用ネットワークにおいてどのような中心的位置付けに存在するかを表したものである。具体的には、次数中心性 (CNT_DEGRE)[102]、媒介中心性 (CNT_BETWE)[102]、近接中心性 (CNT_CLOSE)[102]、固有ベクトル中心性 (CNT_EIGEN)[103]、ネットワーク制約 (CNT_NETWORK)[104]、クラスタリング係数 (CNT_CLUST)[105]、ページランク (CNT_PAGER)[106]、ハブスコア (CNT_HUBSC)[107]、オーソリティスコア (CNT_AUTHOR)[107] を用いる。引用論文特徴量は、対象論文が引用する先の論文に対して上記の特徴量を算出したうえで、最大値、最小値、平均値、合計値といった統計代表値を特徴量としたものである。表 3 に、上記 4 カテゴリ全 63 種類の特徴量を示す。対象分野の引用ネットワークのうち最大連結成分を対象に算出したこれらの特徴量は、予測モデルの説明変数として用いられるものである。

3.2.3 モデル評価

構築された分析モデルの評価には F 値を用いる。Fmeasure は精度と適合率と再現率の調和平均で得られる指標である。適合率は、萌芽であると予測された論文のうち、本当に萌芽であった論文の割合である。再現率 (recall) は実際に萌芽である論文のなかで、萌芽であると予測された論文の割合である。Fmeasure は予測モデルの評価指標として一般的に用いられるもののひとつである。

表 3. 萌芽予測モデルに用いるネットワーク特徴量

特徴量の種類	特徴量名	説明	Ref.
ネットワーク		対象データセット、対象年度のネットワーク特徴量	
	NW_NODES	ネットワークに含まれる論文数	
	NW_EDGES	ネットワークに含まれる引用リンク数	
	NW_MAXQ	ネットワークに含まれるクラスタの Q 値の最大値	[101]
クラスタ		対象論文が所属するクラスタに属する特徴量	
	CL_QMAX	対象論文が所属するクラスタの Q 値の最大値	[101]
	CL_NODES CL_RANK	対象論文が所属するクラスタのノード数 対象論文が所属するクラスタの順位	
中心性		対象論文のネットワーク中心性	
	CNT_DEGRE	次数中心性	[102]
	CNT_BETWE	媒介中心性	[108]
	CNT_CLOSE	近接中心性	[102]
	CNT_EIGEN	固有ベクトル中心性	[103]
	CNT_NETWO	ネットワーク制約	[104]
	CNT_CLUST	クラスタリング係数	[105]
	CNT_PAGER	ページランク	[106]
	CNT_HUBSC	ハブスコア	[109]
	CNT_AUTHOR	オーソリティスコア	[109]
引用関係		対象論文が引用している論文群について、各特徴量を集計して特徴量としたもの	
	CITING_MAX-[feature]	引用論文群における当該特徴量の最大値	
	CITING_MIN-[feature]	引用論文群における当該特徴量の最小値	
	CITING_AVG-[feature]	引用論文群における当該特徴量の平均値	
	CITING_SUM-[feature]	引用論文群における当該特徴量の合計値	

3.2.4 将来予測

現時点で構築可能な最新モデルに対して、現在我々が入手可能な最新の論文データ（2015年1月1日から2015年7月27日までに発行された論文データ）を入力し、今後3年後の2018年に被引用数増加率が上位5%以内に入ると予測された論文を抽出する。抽出された論文のうち、上位10論文について考察を行う。また、萌芽論文上位n件を抽出しそれぞれの論文が所属するサブクラスタ（第三レイヤー単位）の頻度分布を元に、萌芽論文数割合が高いサブクラスタに着目し、次に説明する手段によりこういったトピックによって構成されているサブクラスタであるかを抽出する。本研究ではn=1,000とする。

3.2.5 各クラスタにおけるトピック抽出

各クラスタに所属する論文それぞれが有するトピック内容についてはLatent Dirichlet Allocation(LDA)を用いて推定を行う。LDAはトピックモデルと呼ばれる手法の一つであり、対象となる文書（群）がどのような内容について記述されているかを推定する事を目的とした確率的言語モデルである。LDAの前提は一つの文書（群）は複数のトピックから成り立つということを仮定した

モデルであることから、対象を引用ネットワークにおけるクラスタ単位として分析するという目的にかなう。例えば、ある論文群がシリコン系太陽光発電、薄膜型太陽光発電、色素増感型太陽光発電のトピックによって成り立っているとする。その際の各トピックの生成確率(シリコン, 薄膜, 色素増感)=(0.1, 0.3, 0.6)になる確率が 0.3, (シリコン, 薄膜, 色素増感)=(0.6, 0.2, 0.2)になる確率が 0.6 のように確率分布の確率を定める。グラフィカルモデルとして示すと図 14 のように表される。ここで、 α はトピックの選択確率を得るためのパラメータ、 β はトピックに応じて単語の生成確率を得るためのパラメータである。 M 個の文書集合に対してそれぞれの文書が持つ N 個の単語が与えられた際に、もっともらしいこれらのパラメータを推定する推論である [110]。

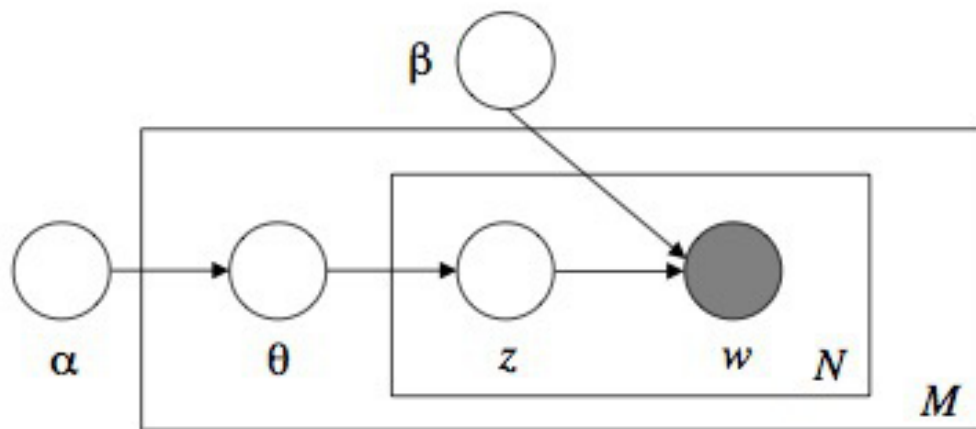


図 14. LDA によるトピックモデル (Blei et al.(2003))

科学技術書誌情報に LDA を用いた研究では [111, 112, 113, 114] などがあげられる。例えば、Jiang et. al.,(2016) は、水力発電分野に関連する 1,726 論文を対象に、トピックモデルを用いて “fish”, “species”, “climate”, “emission”, “lake”, “sediment”, “Turkey” といったホットキーワードを抽出した [114]。これらは特定の年次に出版された論文の全てを母数として、対象分野がどのようなトピックで構成されているかを理解することを助けるものである。LDA の可視化には LDAvis を用いる [115]。任意のトピック t における単語 w の顕著度 $Saliency_{term_w}$ は下記で定義する。

$$Saliency_{term_w} = frequency(w) * [sum_t p(t|w) * log(p(t|w)/p(t))] \quad (2)$$

また、それぞれのクラスタが有するトピック数の推定には (Deveaud et al.2014) による指標を用いる [116]。

3.3 結果

3.3.1 ナノカーボン分野データセットにおける結果

2章で示したナノカーボン分野のデータセットから、直接引用にもとづき引用関係ネットワークを構築した結果、379,044本の論文が最大連結成分に属する結果となった。最大連結成分に所属する論文全てに対し、表3で示した特徴量を計算した。負例の作成は、同年に出版された論文のうち被引用数の増加数が下位50%以下である論文のなかからランダム抽出を行った。ランダム抽出は8回実施する。すなわち、一回の実験において8つのモデルを構築し、モデル評価では8つのモデルの平均値を用いた。評価結果は表4に示すとおりである。

表4. 各年における評価結果 (ナノカーボン分野)

予測対象年 (t1)	モデル構築年 (t0)	予測対象論文数 (平均値)	予測萌芽論文数 (平均値)	Fmeasure (平均値)
2006	2002	2,990	1,495	67.5
2007	2003	3,598	1,779	63.8
2008	2004	3,990	1,995	74.3
2009	2005	4,664	2,332	55.5
2010	2006	4,994	2,497	86.2
2011	2007	5,830	2,915	85.3

それぞれの年次において構築されたモデルについて、予測貢献度が高い特徴量を表5に示す。

表6は、2011年に出版された論文のうち萌芽スコアが上位である論文の上位10本について、3年後の2014年の時点の被引用数を示したものである。表6の論文のうち、1, および3から10位に位置付けられた9本の論文は2014年に萌芽論文の条件を満たした。すなわち、表5に挙げられている10の論文の90%が2014年の時点で被引用数増加数上位5%にあった事を意味する。

表 5. 各年のモデルにおける寄与度上位 5 つの特徴量 (ナノカーボン分野)

モデル構築年 $t_0=2002$ 年 予測対象年 $t_1=2006$ 年	モデル構築年 $t_0=2003$ 年 予測対象年 $t_1=2007$ 年	モデル構築年 $t_0=2004$ 年 予測対象年 $t_1=2008$ 年
CNT_PAGER 20.5	CNT_PAGER 22.3	CNT_PAGER 27.1
CNT_AUTHO 9.4	CNT_AUTHO 10.3	CNT_AUTHO 11.2
CITING_MAX- CNT_DEGRE 5.3	CNT_DEGRE 8	CNT_DEGRE 9
CNT_DEGRE 5.3	CITING_MAX- CNT_DEGRE 5.4	CNT_CLOSE 5.5
CITING_SUM- CL_RANK 4.2	CNT_CLOSE 4.3	CITING_AVG- CNT_CLOSE 4.5

モデル構築年 $t_0=2005$ 年 予測対象年 $t_1=2009$ 年	モデル構築年 $t_0=2006$ 年 予測対象年 $t_1=2010$ 年	モデル構築年 $t_0=2007$ 年 予測対象年 $t_1=2011$ 年
CNT_PAGER 23.3	CNT_PAGER 25.8	CNT_PAGER 33.1
CNT_AUTHO 9.7	CNT_AUTHO 18.3	CNT_AUTHO 14.9
CNT_DEGRE 6.1	CNT_DEGRE 8.2	CNT_CLOSE 9.3
CITING_SUM- CL_RANK 3.6	CNT_CLOSE 5.7	CNT_DEGRE 8.9
CITING_SUM- CL.QMAX 3.5	CITING_SUM- CL_RANK 4.6	CITING_AVG- CNT_CLOSE 5.2

表 6. 2011 年に出版された論文のうち萌芽論文となる可能性が高いと予測された論文上位 10 件

Authors	Title	Journal	No. cited (2011)	No. cited (2014)	Ref.
Zhang, Q, et al.	Carbon Nanotube Mass Production: Principles and Processes	Chemsuschem	0	84	[117]
Lan, YC, et al.	Physics and applications of aligned carbon nanotubes	Advances in Physics	0	35	[118]
Lee, SH, et al.	Tailored Assembly of Carbon Nanotubes and Graphene	Advanced Functional Materials	6	82	[119]
Das Sarma, S, et al.	Electronic transport in two-dimensional graphene	Reviews of Modern Physics	51	664	[120]
Huang, X, et al.	Graphene-Based Materials: Synthesis, Characterization, Properties, and Applications	Small	26	587	[121]
Saito, R, et al.	Raman spectroscopy of graphene and carbon nanotubes	Advances in Physics	0	98	[122]
Li, YD, et al.	Methane decomposition to CO _x -free hydrogen and nano-carbon material on group 8-10 base metal catalysts: A review	Catalysis Today	5	46	[123]
Yan, LA, et al.	Low-toxic and safe nanomaterials by surface-chemical design, carbon nanotubes, fullerenes, metallofullerenes, and graphenes	Nanoscale	5	67	[124]
Singh, V, et al.	Graphene based materials: Past, present and future	Progress in Materials Science	7	506	[125]
Leary, R and Westwood, A	Carbonaceous nanomaterials for enhancement of TiO ₂ photocatalysis	Carbon	11	223	[126]

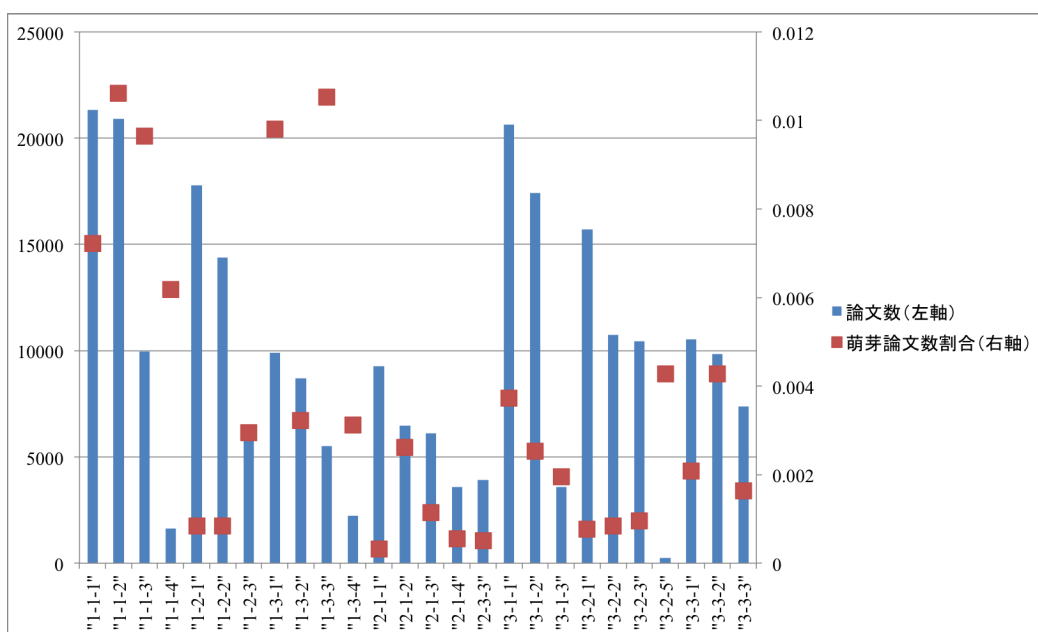


図 15. サブクラスター別（第三レイヤー）上位 1000 萌芽論文の頻度

萌芽スコアとして上位 1,000 論文を対象に、第 3 レイヤーまでのクラスタリング別集計を行った結果を図 15 に示す。なお、第 1 レイヤーおよび第 2 レイヤーは上位 3 クラスターのみを対象とし、第 3 レイヤーは全てのクラスターを対象としている。この図から、クラスター 1-1-2 および、1-3-3 が最もクラスター単位の萌芽度が高い論文が集中していることがわかる。ここでは、論文数の少ない 1-3-3 に注目する。

クラスター 1-3-3 に含まれる萌芽論文のみを対象とした LDA の結果を図 16、図 17、図 18、図 19 に示す。図 16 は、ナノカーボン分野 2011 年出版論文クラスター 1-3-3 におけるトピック分類概要を示す。右図は、クラスター 1-3-3 に所属した論文のアブストラクトから抽出された顕著用語の上位 20 語の頻度分布である。左図は、クラスター 1-3-3 を 8 つのトピックに分類した際の主成分分析の可視化図である。図 17 は、ナノカーボン分野 2011 年出版論文クラスター 1-3-3 における第一トピックの結果である。図 18 は、ナノカーボン分野 2011 年出版論文クラスター 1-3-3 における第二トピックの結果である。図 19 は、ナノカーボン分野 2011 年出版論文クラスター 1-3-3 における第三トピックの結果である。いずれの図においても、図の左側に、上位 10 トピックの大きさ、相対的距離による他のトピックとの近さをあらわしており、図の右側はハイライトされたトピックにおける特徴後をあらわしている。横棒グラフにおい

て水色で示されたものは、ハイライトされたトピックにおける単語の頻度をあらわし、赤でしめしたものはハイライトされたトピックないの推定頻度を示すものである。図 17 において、特に推定頻度の割合が高い単語に注目すると、”electronic”,”band”,”zigzag”,”gap”,”armchair”などが目立つ。これらに関する考察は後ほど考察の項目において言及する。

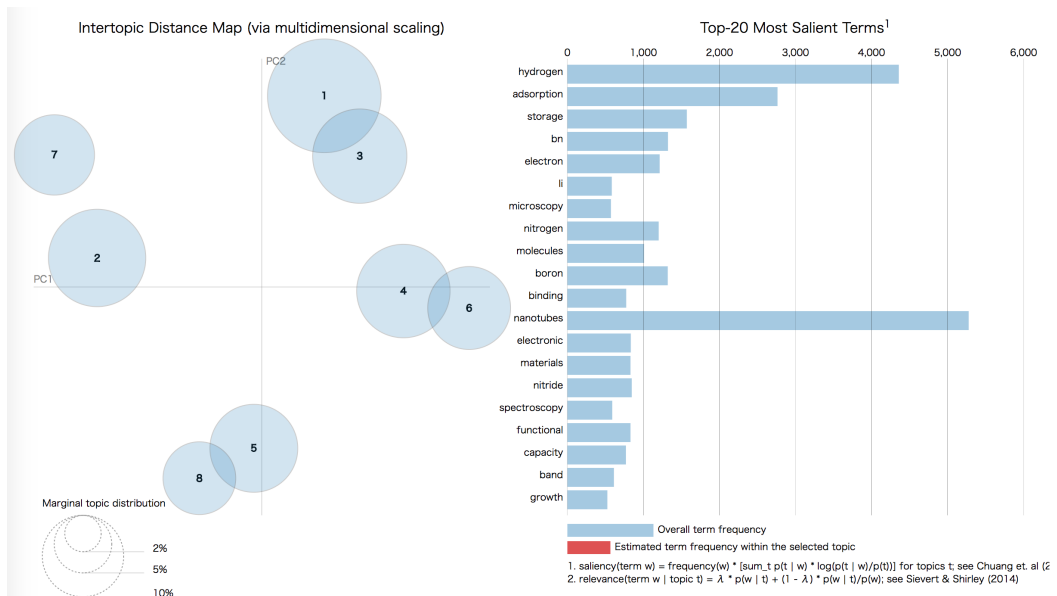


図 16. ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 におけるトピック分類概要

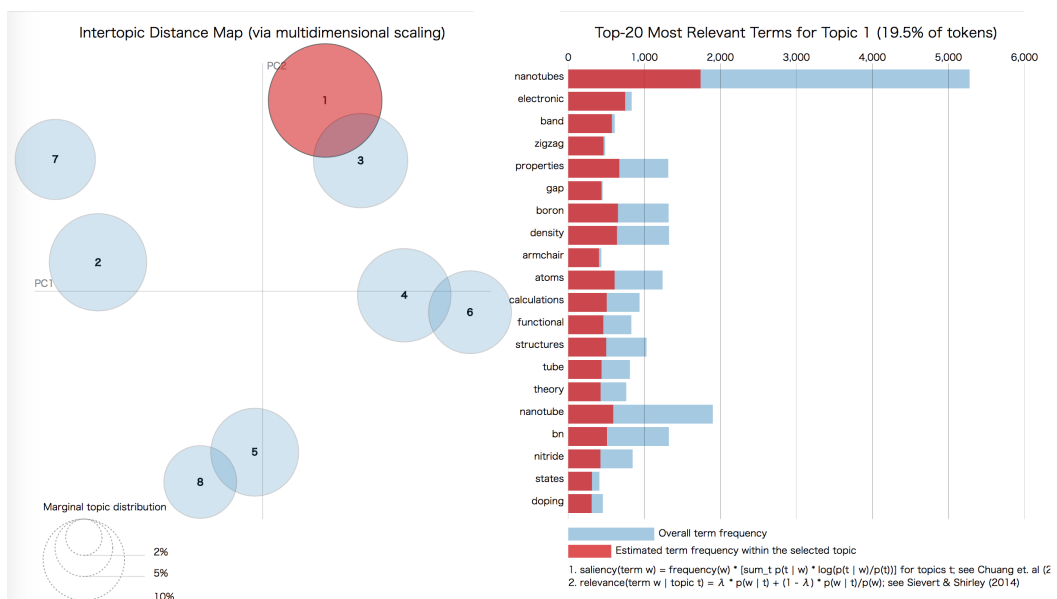


図 17. ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 における第一トピック

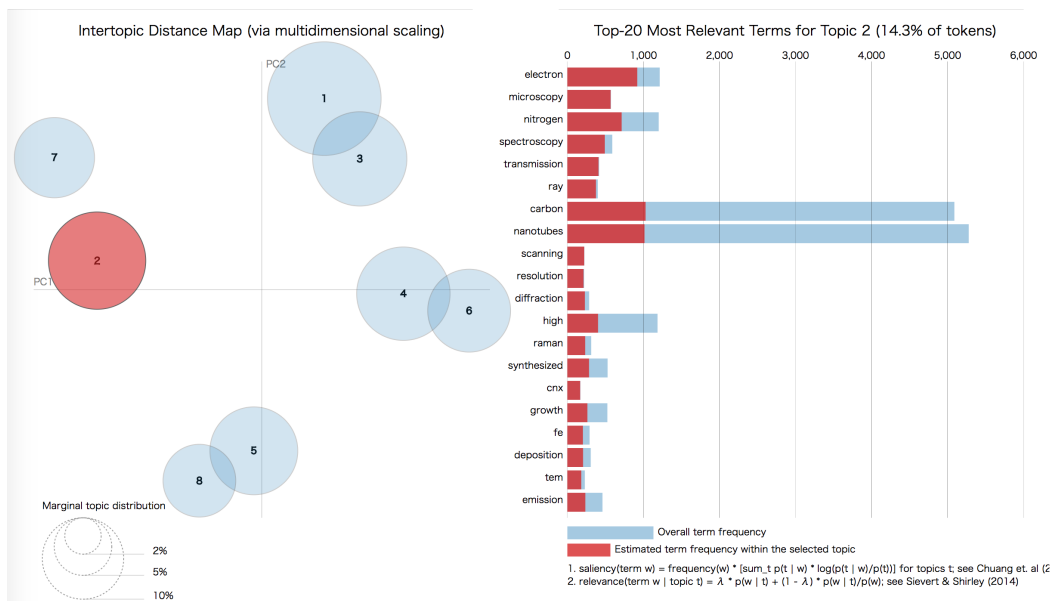


図 18. ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 における第二トピック

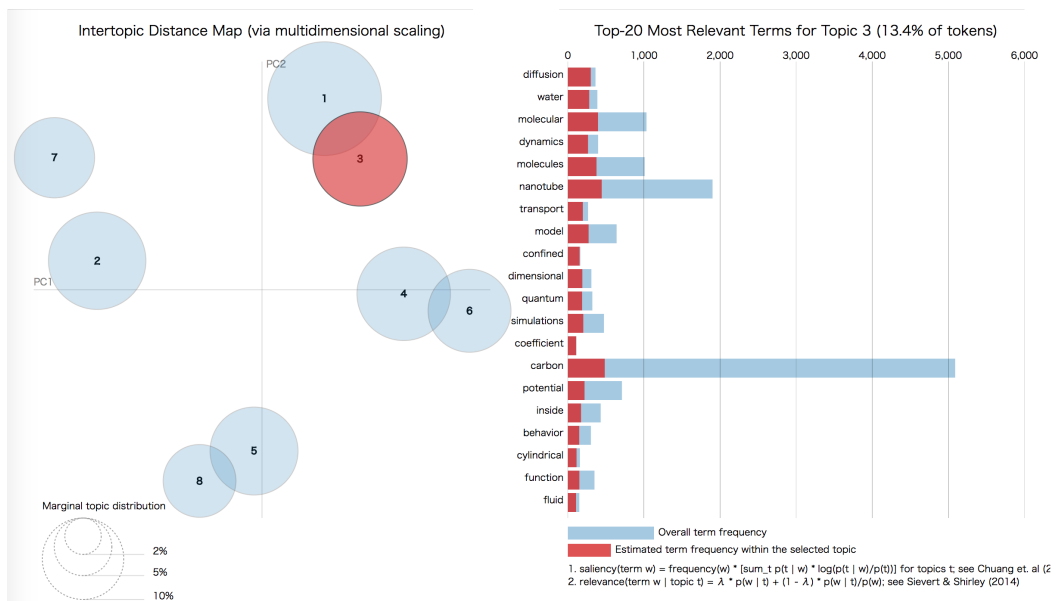


図 19. ナノカーボン分野 2011 年出版論文クラスタ 1-3-3 における第三トピック

3.3.2 太陽光発電分野データセットにおける結果

得られたデータセットに対して直接引用にもとづき引用関係ネットワークを構築したところ 112,430 本の論文が最大連結成分に属する結果となった。最大連結成分に所属する論文全てに対し、表 3 で示した特徴量を計算した。また、ネットワーク中における被引用数をすべての論文に対して計算した。負例の作成は、同年に出版された論文のうち被引用数の増加数が下位 50% 以下である論文のなかからランダム抽出を行った。ランダム抽出は 8 回実施する。すなわち、一回の実験において 8 つのモデルを構築し、モデル評価では 8 つのモデルの平均値を用いた。評価結果は表 7 に示すとおりである。

表 7. 各年における評価結果 (太陽光発電分野)

予測対象年 (t1)	モデル構築年 (t0)	予測対象論文数 (平均値)	萌芽論文数 (平均値)	予測萌芽論文数 (平均値)	Fmeasure (平均値)
2007	2003	688	344	334.3	82.2
2008	2004	824	412	401.2	83.3
2009	2005	1,030	515	493.6	82.9
2010	2006	1,358	679	655.9	83.9
2011	2007	1,818	909	900.5	76.8
2012	2008	2,246	1,123	1,118.70	74.9

それぞれの年次において構築されたモデルについて、予測貢献度が高い特徴量を表 8 に示す。

表 8 の中で最も寄与の高い特徴量はページランク (CNT_PAGER) であった。ページはウェブページの重要性を引用関係に基づいて評価しするものであり、本研究では学術論文を評価するアルゴリズムとして計算した。この特徴量は、引用を多くされている論文を引用することで向上する中心性の一つである。相互引用などを含むローカルなコミュニティ間での引用を有する論文の相対的重要性を減少させる。

次に重要な特徴は次数中心性 (CNT_DEGRE) ある。本研究では、特徴量の計算コストの観点からも引用関係ネットワークを無向ネットワークとして中心性を算出している。厳密には引用と被引用の合算ではあるが、出版年における被引用は極めて少なく多くは引用に基づくものと考えて良いと判断した。従って次数中心性は引用論文数とほぼ同義であると考えることができ、多くの論文を引用している論文に高いスコアが振り当てられる特徴量である。

表 9 は、2012 年に出版された論文を対象にから、萌芽論文になるであろうと予測された上位 10 の論文が、実際にどのように被引用数を獲得したかを示し

表 8. 各年のモデルにおける寄与度上位 5 つの特徴量 (太陽光発電分野)

モデル構築年 $t_0=2003$ 年 予測対象年 $t_1=2007$ 年	モデル構築年 $t_0=2004$ 年 予測対象年 $t_1=2008$ 年	モデル構築年 $t_0=2005$ 年 予測対象年 $t_1=2009$ 年
CNT_DEGRE 4.8	CNT_PAGER 6.2	CNT_PAGER 6.9
CNT_PAGER 4.7	CNT_DEGRE 4.3	CNT_DEGRE 3.4
CNT_AUTHO 2.4	CITING_MAX- CNT_DEGRE 3.2	CITING_MAX- CNT_DEGRE 3
CITING_MAX- CNT_HUBSC 2.4	CNT_AUTHO 2.5	CNT_AUTHO 2.6
CITING_MIN- CNT_CLUST 2.2	CITING_MAX- CNT_CLOSE 2.1	CITING_SUM- CNT_DEGRE 2.2
モデル構築年 $t_0=2006$ 年 予測対象年 $t_1=2010$ 年	モデル構築年 $t_0=2007$ 年 予測対象年 $t_1=2011$ 年	モデル構築年 $t_0=2008$ 年 予測対象年 $t_1=2012$ 年
CNT_PAGER 5.8	CNT_PAGER 6.5	CNT_PAGER 10.2
CNT_DEGRE 5.5	CNT_AUTHO 5	CNT_DEGRE 7.6
CNT_AUTHO 4.6	CNT_DEGRE 4.2	CNT_AUTHO 6.9
CITING_MAX- CNT_DEGRE 2.1	CITING_MAX- CNT_DEGRE 2.2	CITING_AVG- CNT_CLOSE 3.5
CITING_SUM- CNT_DEGRE 2	CITING_SUM- CNT_DEGRE 2.1	CITING_MIN- CNT_CLOSE 3.4

ている。表 9 の論文 1、3、4、6、7、8 および 10 は、萌芽論文の定義を満たしたことがみてとれる。

表 9. 2012 年に出版された太陽光発電分野の論文のうち萌芽論文となる可能性が高いと予測された論文 10 本の概要と 3 年後の実際の被引用数

Authors	Title	Journal	No. cited (2012)	No. cited (2015)	Ref.
Yip, H. L., Jen, A. K. Y.	Recent advances in solution-processed interfacial materials for efficient and stable polymer solar cells	Energy & Environmental Science	19	320	[127]
Chen, et al.	Morphology characterization in organic and hybrid solar cells	Energy & Environmental Science	5	180	[128]
Kumar, P., Chand, S.	Recent progress and future aspects of organic solar cells	Progress in Photovoltaics	1	68	[129]
Boule, J., Ackermann, J.	Solid-state dye-sensitized and bulk heterojunction solar cells using TiO ₂ and ZnO nanostructures: recent progress and new concepts at the borderline	Polymer International	13	50	[130]
Zhou, et al.	Rational design of high performance conjugated polymers for organic solar cells	Macromolecules	52	584	[131]
Ooyama, Y., Harima, Y.	Photophysical and electrochemical properties, and molecular structures of organic dyes for dye-sensitized solar cells	Chemphyschem	0	104	[132]
Mishra, A., Bauerle, P.	Small molecule organic semiconductors on the move: promises for future solar energy technology	Angewandte Chemie International Edition	42	529	[133]
Li, et al.	Characterisation of electron transport and charge recombination using temporally resolved and frequency-domain techniques for dye-sensitised solar cells	International Reviews in Physical Chemistry	1	35	[134]
Dou, et al.	Tandem polymer solar cells featuring a spectrally matched low-bandgap polymer	Nature Photonics	0	865	[135]
Berger, et al.	The electrochemistry of nanostructured titanium dioxide electrodes	Chemphyschem	3	36	[136]

萌芽スコアとして上位 1,000 論文を対象に、第 3 レイヤーまでのクラスタリング別集計を行った結果を図 20 に示す。なお、第 1 レイヤーおよび第 2 レイヤーは上位 3 クラスタのみを対象とし、第 3 レイヤーは全てのクラスタを対象としている。この図から、クラスタ 2-3-1 が最もクラスタ単位での萌芽論文数

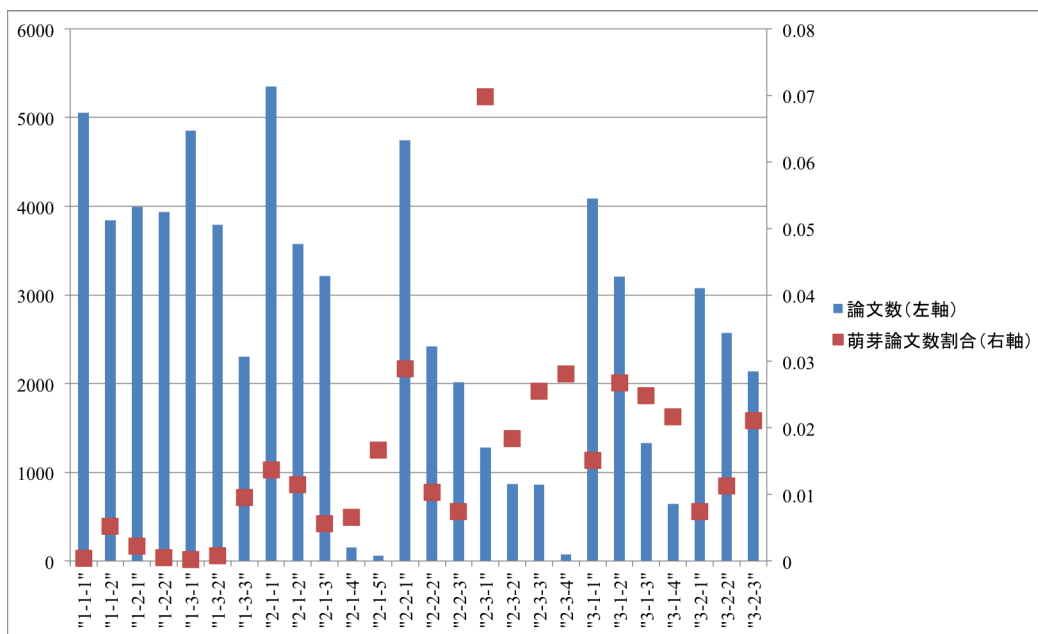


図 20. サブクラスタ別 (第三レイヤー) 上位 1000 萌芽論文の頻度

割合が高いことがわかる。

クラスタ 2-3-1 に含まれる萌芽論文のみを対象とした LDA の結果を図 21、図 22、図 23、図 24 に示す。図 21 は、太陽光発電分野 2012 年出版論文クラスタ 2-3-1 におけるトピック分類概要を示す。右図は、クラスタ 2-3-1 に所属した論文のabstractから抽出された顕著用語の上位 20 語の頻度分布である。左図は、クラスタ 2-3-1 を 6 つのトピックに分類した際の主成分分析の可視化図である。図 22 は、太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第一トピックの結果である。図 23 は、太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第二トピックの結果である。図 24 は、太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第三トピックの結果である。

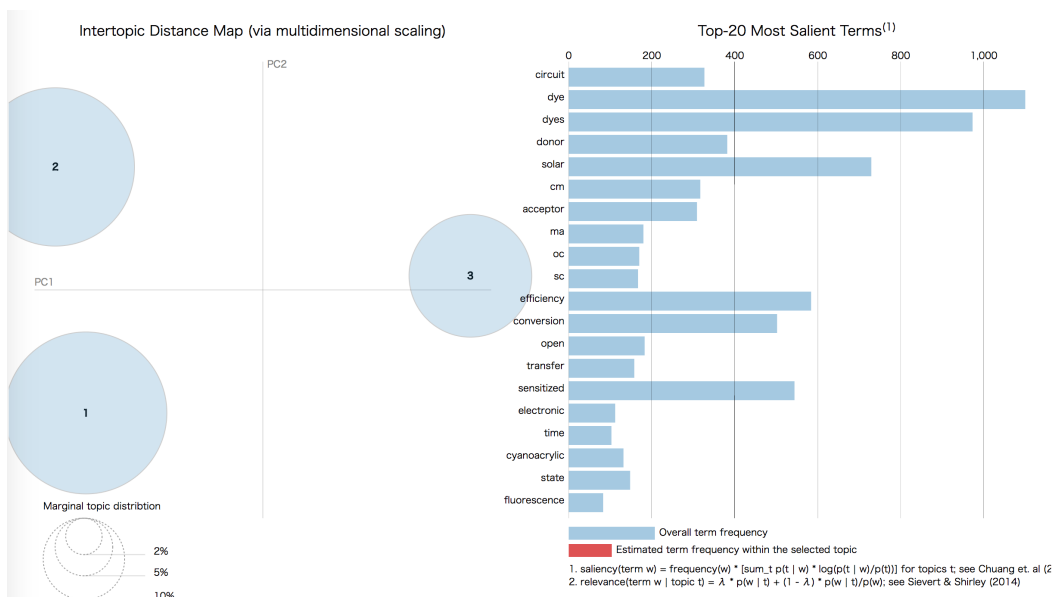


図 21. 太陽光発電分野 2012 年出版論文クラスター 2-3-1 におけるトピック分類概要

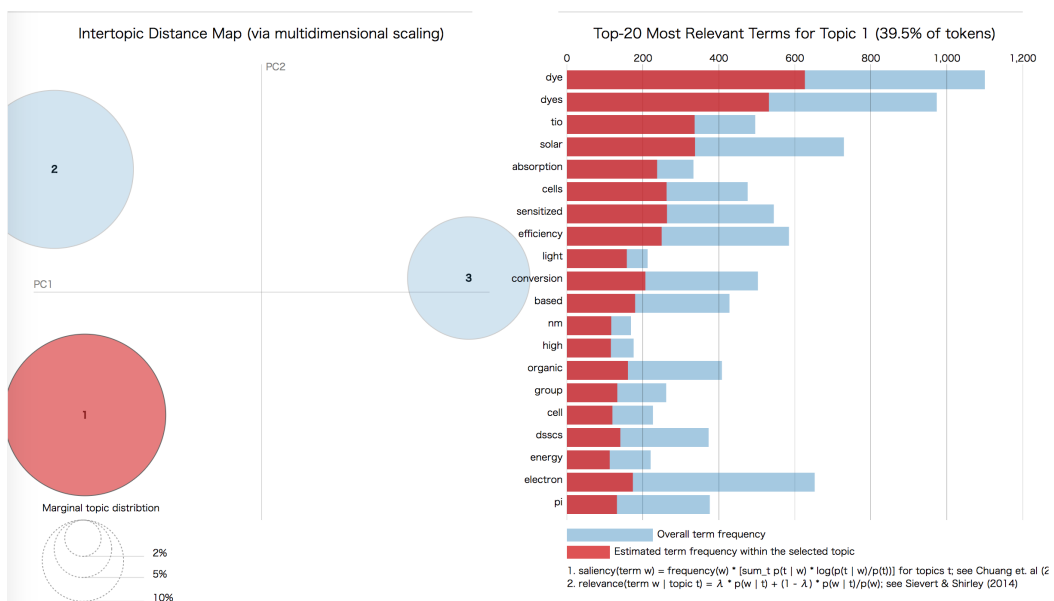


図 22. 太陽光発電分野 2012 年出版論文クラスター 2-3-1 における第一トピック

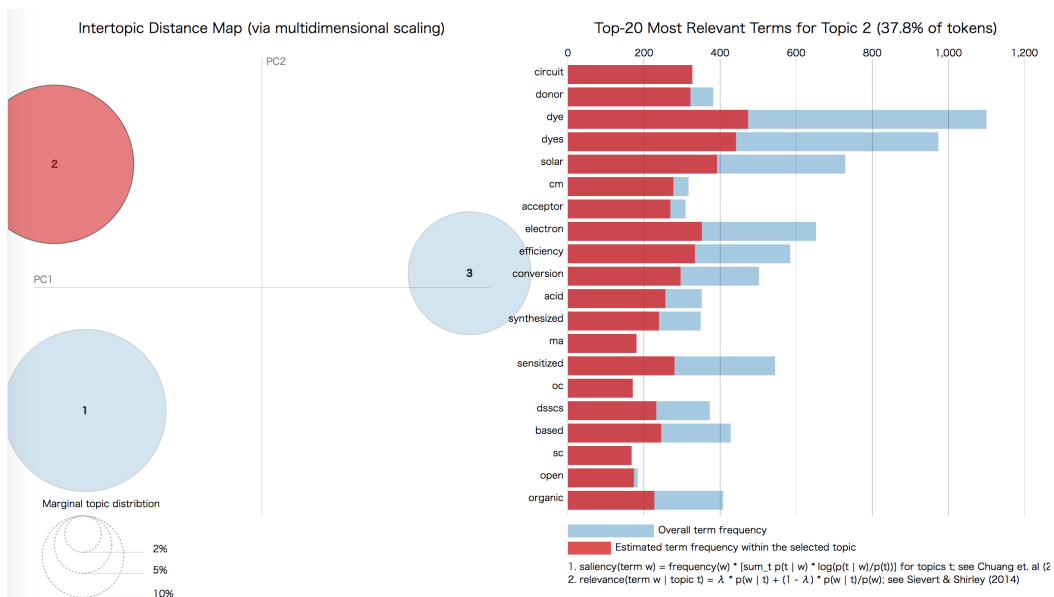


図 23. 太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第二トピック

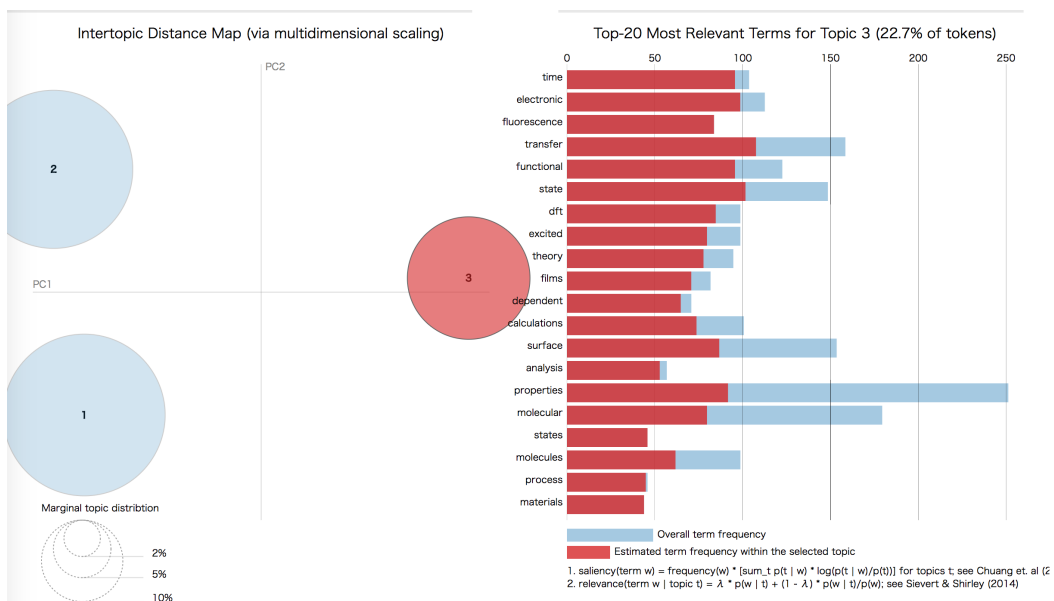


図 24. 太陽光発電分野 2012 年出版論文クラスタ 2-3-1 における第三トピック

3.3.2.1 将来予測

最後に、2015 年に出版された論文を予測モデルに入力し、今後被引用数が増加すると予測されたタイトル 10 本を表 10 に示す。表 10 に示したこれらの 10 本の論文が、実際に我々の定義する萌芽論文になるかどうかは 3 年後に評価が可能である。

萌芽スコアとして上位 1,000 論文を対象に、第 3 レイヤーまでのクラスタリング別集計を行った結果を図 25 に示す。なお、第 1 レイヤーおよび第 2 レイヤーは上位 3 クラスタのみを対象とし、第 3 レイヤーは全てのクラスタを対象としている。この図から、クラスタ 2-3-1 が最もクラスタ単位での萌芽論文数

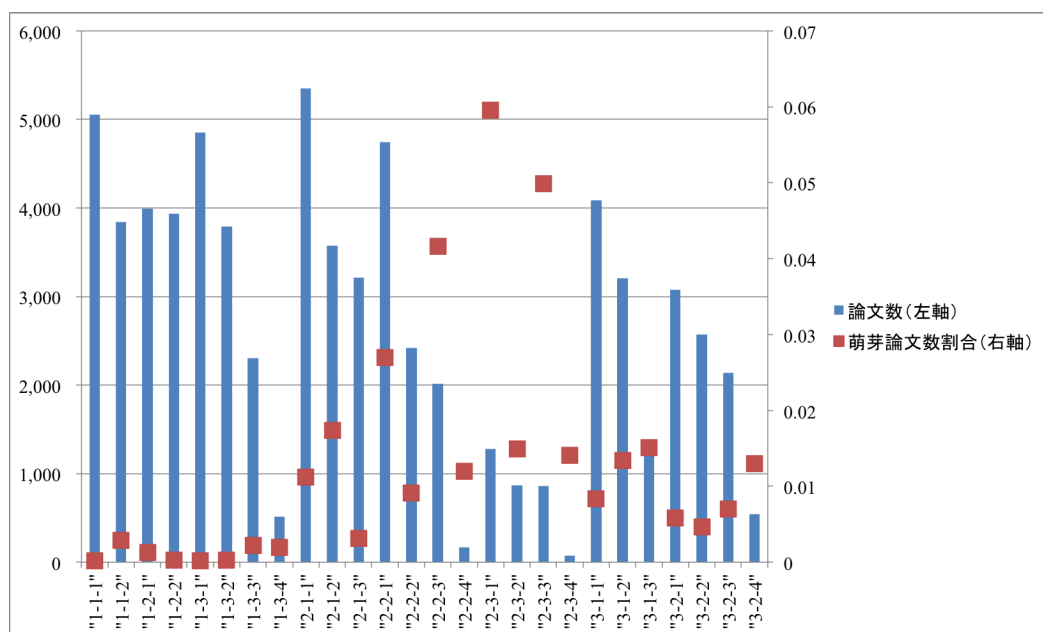


図 25. サブクラスタ別 (第三レイヤー) 上位 1000 萌芽論文の頻度

割合が高いことがわかる。

クラスタ 2-3-1 に含まれる萌芽論文のみを対象とした LDA の結果を図 26、図 27、図 28、図 29 に示す。図 26 は、太陽光発電分野 2015 年出版論文クラスタ 2-3-1 におけるトピック分類概要を示す。右図は、クラスタ 2-3-1 に所属した論文のアブストラクトから抽出された顕著用語の上位 20 語の頻度分布である。左図は、クラスタ 2-3-1 を 6 つのトピックに分類した際の主成分分析の可視化図である。図 27 は、太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第一トピックの結果である。図 28 は、太陽光発電分野 2015 年出版論文ク

表 10. 2015 年に出版された論文のうち萌芽論文となる可能性が高いと予測された論文上位 10 件

Authors	Title	Journal	Ref.
Calogero, et al.	Vegetable-based dye-sensitized solar cells	Chemical Society Reviews	[137]
Wu, et al.	Electrolytes in dye-sensitized solar cells	Chemical Reviews	[138]
Lu, et al.	Recent advances in bulk heterojunction polymer solar cells	Chemical Reviews	[139]
Bella, et al.	Aqueous dye-sensitized solar cells	Chemical Society Reviews	[140]
Cheng, et al.	Versatile third components for efficient and stable organic solar cells	Materials Horizons	[141]
Chueh, et al.	Recent progress and perspective in solution-processed Interfacial materials for efficient and stable polymer and organometal perovskite solar cells	Energy & Environmental Science	[142]
Liu, et al.	Functionalized graphene and other two-dimensional materials for photovoltaic devices: device design and processing	Chemical Society Reviews	[143]
Singh, et al.	Graphene-based dye-sensitized solar cells: a review	Science of Advanced Materials	[144]
Liang, et al.	ZnO cathode buffer layers for inverted polymer solar cells	Energy & Environmental Science	[145]
Albero, et al.	Efficiency records in mesoscopic dye-sensitized solar cells	Chemical Record	[146]

ラスト 2-3-1 における第二トピックの結果である。図 29 は、太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第三トピックの結果である。

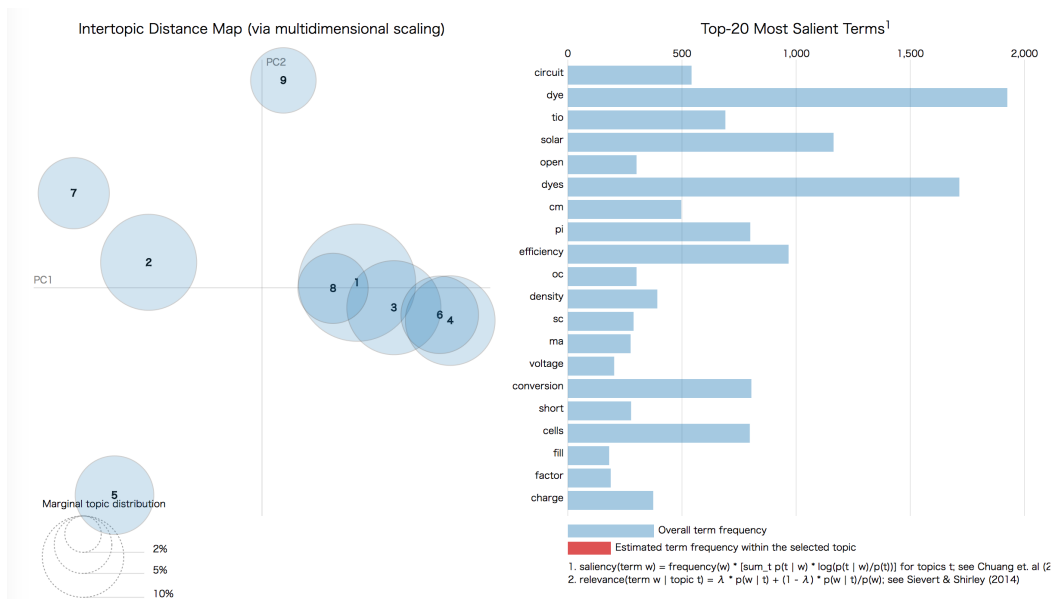


図 26. 太陽光発電分野 2015 年出版論文クラスタ 2-3-1 におけるトピック分類概要

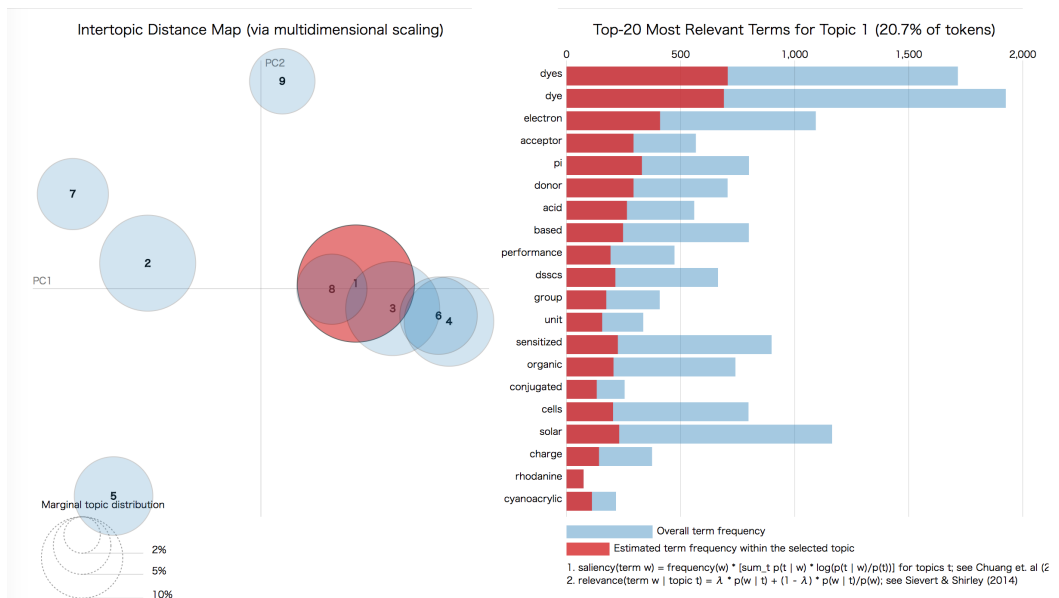


図 27. 太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第一トピック

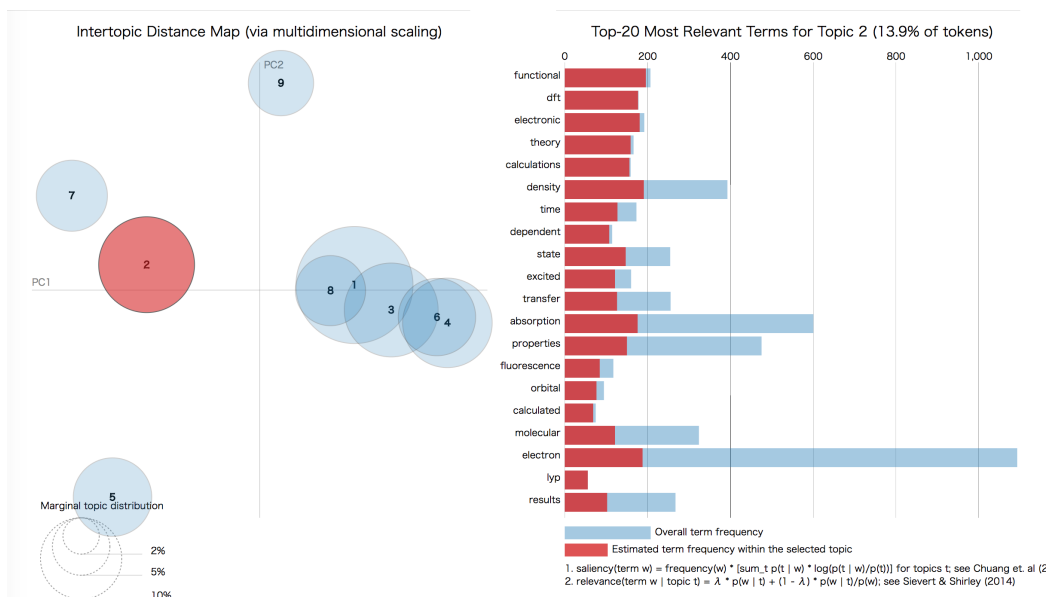


図 28. 太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第二トピック

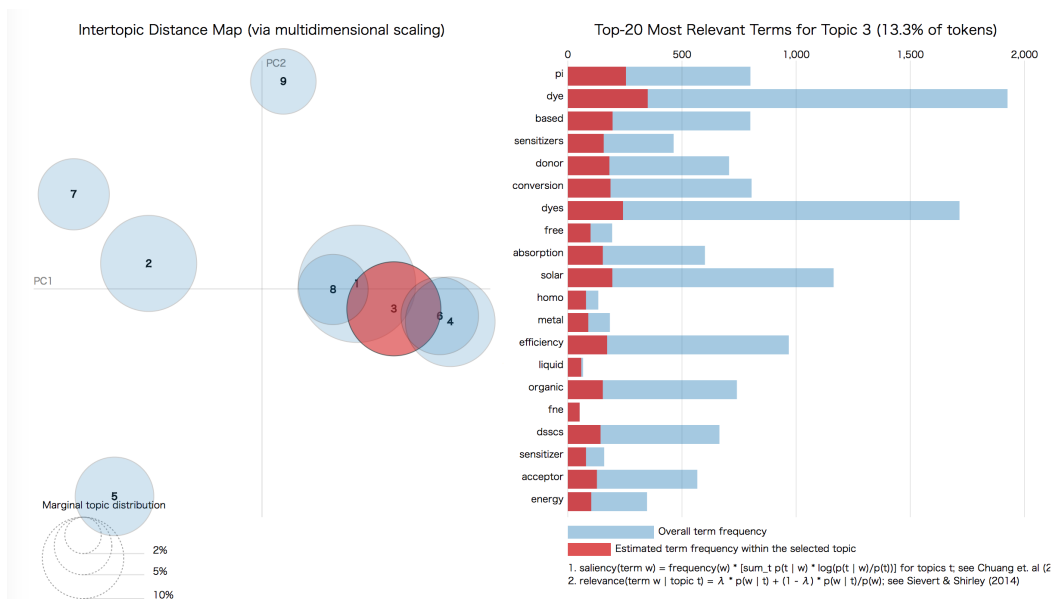


図 29. 太陽光発電分野 2015 年出版論文クラスタ 2-3-1 における第三トピック

3.4 考察

本研究では、ナノカーボン分野および、太陽光発電分野に対して、出版後3年後に萌芽論文となるかどうかを予測するモデルを適用した。ナノカーボン分野については、2011年に出版された論文を対象に予測された結果の上位10論文中9本が定義上の萌芽論文であることが確認できた。また太陽光発電分野については、2011年に出版された論文のうち萌芽論文になると予測された論文のうち上位10論文中7論文が実際に萌芽論文となった。構築したモデルは十分に評価できると考えられる。通年においても安定的に Fmeasure は 0.8 程度を推移しており、適合率および再現率ともにバランスのとれたモデルが構築できたと考える。以降に、それぞれのデータ・セットにおける個別の考察を行う。表5および、表8から、いずれにおいても最も寄与の高い特徴量はページランク (CNT_PAGER) であった。ページランクはウェブページの重要性を引用関係に基づいて評価しするために提案された手法であるが、本研究では学術論文を評価する目的で用いたものである。この特徴量は、被引用数が多い論文を引用することで上昇する指標のひとつと解釈できる。同時に、この指標は相互引用などのローカルなコミュニティ間での引用を有する論文の相対的重要性を減少させる。なお、本研究では、特徴量の計算コストの観点からも引用関係ネットワークを無向ネットワークとして中心性などを算出している。厳密には引用と被引用の合算ではあるが、出版年における被引用は極めて少なく多くは引用に基づくものと考えて良いと判断した。

次に重要な特徴は次数中心性 (CNT_DEGRE) である。多くの論文を引用している論文ほど、萌芽上位にランクしやすい事を意味する。特徴量であると言える。これら2つの特徴量が上位にきていることから言えることは、将来的に被引用数を稼ぐとされるような論文(すなわち本研究における萌芽論文)は、対象分野を適切に調査したうえでの論文であるということができる。

3.4.1 ナノカーボン分野における考察

Zhang et al.(2011) は CNT の大量生産に注目した論文である [117]。当初 CNT の製造方法には、アーク放電法とレーザ蒸発法が主に用いられてきた。アーク放電法は欠陥が少なく品質の良い CNT を製造することのできる手法であるが、まとまった量を製造することができない。またレーザ蒸発法は、比較的純度の高い CNT を製造することのできる方法であるがこれもまた、工業的

製造技術としては向かない手法と考えられている [117]。そのような背景の中、大量合成にむいているとされる CVD 法 (Chemical Vapor Deposition) に注目が集まってきた。信州大学の遠藤教授による”Carbon Multiwall Nanotubes”や、オクラホマ大学による”CoMoCATProcess at SWeNT”などの提案を元に、これまで複数の製造技術が開拓され、既に実用化されつつある。当該論文は、CVD 法にかぎらず CNT の大量製造方法に関する研究について網羅的に紹介し議論をしたものである。この論文は 3 年間で 84 件の被引用数を獲得したことがわかり、注目を浴びていることが分かる。表 6 において 2 番目にランクした論文は、配向 CNT における現在の状況と物理的特性およびその適用領域に関するレビュー論文である。CNT の製造技術のうち CVD 法については、その高い配向性も期待されている。そのような Aligned CNTs においては広い適用領域 (例えば、発光、光学アンテナ、サブ波長光伝送、ナノ同軸構造構造の太陽光発電など [118]) が期待されている。この論文は、本研究の定義における萌芽論文の条件は満たさなかったものの、一定の被引用数を獲得していることが見て取れる。3 番目にランクした論文は、カーボンナノチューブおよびグラフェンを 3 次元アセンブリする際の大量生産手法に着目した論文である [119]。カーボンナノチューブおよびグラフェンの 3 次元ネットワークに関連する研究としては Dasgupta et al.(2016) が詳しい [147]。3 次元構造を有する多孔質膜の研究は依然として初期段階であるとしており今後より実用化に寄与する材料が必要であるとしている [147]。グラフェンはグラファイト結晶の 1 原子面を取り出したものである。2004 年に Geim らが高配向性の無水グラファイト (Highly Oriented Pyrolytic Graphite, HOPG) の表面を粘着テープで剥離し、剥離した表面をさらに剥離するという方法でグラフェンの薄片を取り出すことに成功した。この報告以降、グラフェンについては、電気・電子・機械・科学的な特性について明らかになってきた [148]。特にグラフェン内は高い電子移動度を有することが明らかになっている。電子移動度は固体中の電子の速さの目安である。理論的には $2,000,000 \text{ cm}^2/\text{Vs}$ が予測されており [149]、実験的にも $200,000 \text{ cm}^2/\text{Vs}$ という結果が得られている [150]。シリコン中電子の移動度が、 $1000 \text{ cm}^2/\text{Vs}$ であることを考えると、グラフェンの電子移動度はその 100 倍以上となる。高い電子移動度を有するという事は、例えば、高速トランジスタの実現において重要な要素である。当該論文は、グラフェンが有する高い電子移動度に関して注目し、その電気的特性と適用について論じたレビュー論文である。4 番目にランクした論文は、2014 年時点で、664 件の引用数を獲得

していることが確認できる。5番目にランクした論文はグラフェンの物理的特性を網羅的に論じた論文である。グラフェンは高い電子移動度だけでなく、高い熱安定性や優れた強度も有している。当該論文はグラフェンをベースにした、FETs, メモリ, 太陽光デバイス、センシングプラットフォームに関する応用についても網羅的に記述している。2014年時点で587件の被引用数を獲得している。表6における6番目の論文は、ナノ材料の構造分析方法に焦点を当てた論文である。ナノ材料の構造分析において、威力を発揮する手法の一つにラマン分光法が挙げられる。特に炭素材料のラマンスペクトルには、グラファイト構造に由来したG-bandピークと欠陥由来のD-bandピークが現れることが分かっている。これらのピーク比を用いて、ナノカーボン材料の結晶性の純度や欠陥濃度を評価することができるとされている。当該論文は、カーボンナノチューブとグラフェンにおけるラマン分光スペクトルについて注目し、関連研究をまとめたレビュー論文である。表6における7番目にランクした論文はメタン触媒分解の反応原理から、生成されるナノ炭素材料の形状とその形成原理について関連する先行文献を包括的にまとめたレビュー論文である。メタンを高温環境下において触媒に通し水蒸気改質すると、水素と炭素を生成することが可能である。生成された水素は燃料電池の燃料として用いることが可能であり、おもに水素の生成手段として注目を集めていた。一方で、生成された炭素もダイレクトカーボン型燃料電池の材料としても用いることができることなどから、ナノ炭素材料の生成手段という点でも注目を浴びている手法の一つである。カーボンナノチューブは、その構造がアスベストに類似しているため人体に対する毒性を指摘されている。その他のナノ炭素においても、毒性を低減させるための試みは、ナノ炭素材料の活用において注目されている研究領域であるといえる。当該論文は、Nanotoxicology と称されるこの分野において、体系的な知識提供を試みた論文であるといえる。また、著者らは低毒性を実現するための具体的な課題について明示している。当該論文は化学変化によって、カーボンナノ材料の生物学的、毒性学的な変質をもたらす手法について議論するものである。Singh., et al.(2011)による論文は、グラフェンの歴史から、その特性、生成手段、グラフェンが、電気デバイス、オプトロニクスデバイス、科学センサーナノコンポジット、エネルギーストレージといった多様な分野への適用におけるインパクトについて網羅的にまとめている論文である。2014年時点で506件の被引用数を獲得していることが確認できる [125]。10位にランクされた論文のタイトルは、Carbonaceous nanomaterials for enhancement of

TiO₂ photocatalysis である [126]。一般的に光触媒の材料として用いられるものには、酸化チタン (TiO₂) があげられる。しかしながら、その効率や応答範囲の狭さから課題が指摘されていた。酸化チタンにナノカーボン材料を組み合わせることでその特性を大きく変えることができるとされている。当該論文は、nanocarbon-TiO₂ による光触媒の論文として、その生成手段、特徴、今後の方向性についてガイドラインを示している。当該論文は 2014 年時点で、232 件の被引用数を獲得したことが確認できる。

さて、図 17 において、特に推定頻度の割合が高い単語に注目すると、“electronic”、“band”、“zigzag”、“gap”、“armchair”などが目立つことをしめした。単相カーボンナノチューブ (SWCNT) の構造は螺旋度 (chirality) によって導電性などが異なることが知られている。たとえば、ジグザグ型では金属 1/3 半導体 2/3 の割合の特性を、カイラル型では半導体の特性を、アームチェア型は金属の特性を持つ。2010 年時点においてそれぞれにおいて化学合成に課題があったが、2011 年 10 月にはじめてカイラル型とアームチェア型の合成に成功している。残るジグザグ型カーボンナノチューブであるが、これも 2012 年 7 月米国科学会誌 (Journal of The American Chemical Society) において Hitosugi et al.(2012) により、Bottom-up synthesis and thread-in-bead structures of finit (n,0)-zigzag single-wall carbon nanotubes ((n,0)-ジグザグ型単層カーボンナノチューブのボトムアップ化学合成とその折り込み構造) というタイトルで論文が発表された [151]。このように、2011 年前後において、カイラル型、アームチェア型、ジグザグ型の単相カーボンナノチューブに関する化学合成が盛り上がっており、図 17 における“band”、“zigzag”、“gap”、“armchair”といった単語はまさにそれを予期している。一方で、Hitotugi et al.(2012) の上記論文の基となる先行関連論文が 2011 年に出版されている [152]。実際には被引用数を 44 獲得しており、当該モデルにおいて萌芽論文の定義を満たす論文である。しかしながら、萌芽スコア自体は 11932 位にランクされた論文であり、すなわち論文単位だけに注目した萌芽予測モデルだけでは特定できなかった論文である。このことから、論文単位だけでなく、用語単位において萌芽研究という粒度で今後盛り上がる研究分野を特定することは一定程度有効であるとみなすことができる。

3.4.2 太陽光発電分野における考察

Calogero et al.(2015) による論文 [137] は植物などの植物色素を用いた色素増感型太陽光発電に関するレビュー論文である。Vegetable dyes とは、藻類や草花、果物から抽出した sensitizers であり、これを元に色素増感型太陽光発電に活用するアイデアが Vegetable based Dye-sensitized solar cells である。このアイデアは、1993 年に [153] によって提案されたものであり、その後多くの研究者が、このアプローチに関する研究を行っている、なお [153] の現在 (2015.Sep.3rd) の被引用数は 733 件である。萌芽論文 [137] として抽出された論文では Vegetable based Dye-sensitized solar cells の課題はもちろん今後の展望について触れており今後のこの分野の動向については注目に値する対象の一つであると考えられる。色素増感型太陽光発電における電極材料としてカーボンナノチューブを用いる提案がされている。カーボンナノチューブが持つその電氣的、光学的、化学的、物理的、触媒的特徴より、既存の電極材料であるやプラチナと比較して、安価に優れた性能が得られることが期待されている。Batmunkh, M. et al.(2015) は、色素増感型太陽光発電の透明導電性ダイオードや、半導体層、対電極にカーボンナノチューブを用いた研究について総括的に議論した [154]。著者らによれば、カーボンナノチューブを用いて適切なパラメータ設定を行えば、極めて高い変換効率を得ることが期待できる [154]。近年ペロブスカイト型太陽電池が広く注目を集めている。実際、Science および Nature によって 2013 年の最大のブレイクスルーであるとされている。ペロブスカイト型太陽電池は安価な材料および溶液法から作製可能であるため高効率かつ安価な太陽電池として期待されている。ペロブスカイトとはチタン酸カルシウム (灰チタン石, CaTiO_3) の結晶構造を発見したロシアの研究者、Lev. Perovski にちなんで命名された結晶構造のことであり、このペロブスカイトを光電変換につかった最初の研究は 2009 年に発表されたものであるが、2014 年に最高効率は 20.1% を達成した報告例が一躍注目を浴びた。Chueh, CC et al(2015) らは、近年の高分子系太陽電池及びペロブスカイト太陽電池の性能改善に貢献した溶液処理界面層の近年の開発動向を記述した。また、Shi, SW et al(2015) は有機金属ハライドペロブスカイトベースの全固体 (all-solid-state) 太陽電池の発展について論じている。

さて図 21 において、多くが色素増感型太陽電池に関連する単語であることがみておられるが、その中でも特に特徴的な単語として”pi”が挙げられる。色

素増感型太陽電池は、多孔質の酸化チタンに吸着した色素に光が当たると発電する仕組みである。この多孔質酸化チタンに吸着させる色素としてはルテニウム金属錯体色素が一般的で有名である。ルテニウムは、無機色素の代表的な色素であり変換効率や耐久性が高い。近年では東京大学瀬川教授により、長波長領域（近赤外光）を効率よく光電変換できるルテニウム錯体色素が開発されている。その一方で、ルテニウムは高価であることが課題となっており、代替する金属による電極材料の研究が進められている。そのようななか、2009年にドナー- π -アクセプター色素による高効率化が実現された。2011年予測モデルにおいて特徴語としてあげられた”pi”はまさにドナー- π -アクセプター色素に関する研究を示すと考えられる。実際、関連する論文として、Yella et al.(2011)による論文、”Porphyrin-Sensitized Solar Cells with Cobalt (II/III)-Based Redox Electrolyte Exceed 12 Percent Efficiency” が2011年に出版された [155] が、この論文は2011年モデルにおいて萌芽スコアで1位として予測されていただけでなく、事実、出版から3年後の2014年時点の被引用数は、3,732件を獲得しており、紛れもなく予測可能である萌芽論文であることが確認できた。高価とされるRu系色素を使わずに高い変換効率を実現できることからこの研究(2011発表)を皮切りにドナー- π -アクセプター系色素の活用が2012時点で注目を浴びることになったことを示唆する。それだけでなく、2015年モデルにおけるクラスタ2-3-1における第一トピックを表す図21では、”acceptor”、”pi”、”donor”と上位に特徴語が並んでおり、よりこの分野が顕著に注目されたことと評価することができる。

3.5 まとめ

本研究では、ナノカーボン分野411,084本、太陽光発電分野12,999本における膨大な学術論文情報より今後萌芽となりうる論文を予測するモデルを適用した。早期に萌芽的な領域を特定するため、論文が出版された直後少なくとも1年未満の情報のみから、3年後の当該論文の引用数の伸びを予測することを目的としている。既存研究と比較して、多様な特徴量、ネットワーク指標やクラスタリング結果を用いている点が特徴的であり、論文が発表された直後の時点での特徴量から、数年先の引用数の伸びを予測している。予測モデルで利用する特徴量は大きくわけて4つのカテゴリ（ネットワーク、クラスタ、中心性、引用関係）を用いたがいずれもネットワークに着目して構築され得た特徴量である。本研究はこれらのネットワーク指標によって萌芽論文を特定した後、萌芽

論文割合が高いクラスタにおいて論文中が用いられている単語に着目しトピックモデルによって萌芽的な研究分野を点（論文）のみならず面（研究分野）で特定することを試みたものである。

萌芽論文予測モデル自体、ナノカーボン分野においても太陽光発電分野においてもいずれも一定の精度を獲得しており、有用性のたかいモデルが実現できた。寄与度 (Weights) の高い特徴量 (Feature) が Pagerank であることは、被引用数を多く集めている論文に引用されると、その論文の被引用数も増大する可能性が高いことを意味する。また、近接中心性が寄与しているということは、多くの論文との距離が近い論文であることを意味し、すなわち分野の中心的な点であることを意味する。これらから、萌芽論文の特徴は、当該分野の知見を十分に調査し、コミュニティにおいて評価される課題設定をしている論文であることがわかる。いわば著者の能力を定量化した指標の一つと考えることができる。萌芽と予測された論文割合が高いサブクラスタに対する特徴語に着目するとナノカーボン分野においてはジグザグ型単層カーボンナノチューブに関する化学合成に関する研究や、太陽光発電分野においては、ドナー π アクセプター型色素に関する研究にフォーカスを当てることが出来、論文とといった「点」だけではなく、研究領域といった「面」として萌芽分野を考察することに成功している。

ところで、太陽光発電分野においては、構築したモデルに対し、2015年に出版された論文を入力することで今後の太陽光発電分野における洞察を得た。2015年に出版された論文より萌芽と予測された論文上位10件のうち、3本がペロブスカイト太陽電池に関わるレビュー論文であった。ペロブスカイト太陽電池は提案がされてからわずか5年でエネルギー変換効率が5倍に増え、現在主流のシリコン系太陽電池に近づきつつある。また、製造コストが圧倒的に安い点についても注目がされている。我々は、今後ペロブスカイト太陽電池が極めて注目に値すると考える。また、その他のテーマにおいても、予測上位論文では共通しているテーマが少なくない。萌芽領域という点で注目に値する。今後、情報が膨大に増加し知識の構造が複雑化していくほど、企業が行う研究開発投資における意思決定や、国が行う科学技術政策としての予算配分意思決定は困難を極める。一方で、科学や技術の動向見通しは、主体的に作り上げていくものであるという点も忘れてはならない。本研究のような予測モデルが担う役割は、人間が意思決定を行う際にそれを支援するためのものである。膨大な情報から将来的に有用となりうる論文を抽出する支援するための手法は今後より

増大していくと考えられる。

4 第4章：国際特許分類の階層情報を用いた技術融合予測

4.1 背景と目的

4.1.1 技術融合

技術や社会の複雑性に伴い特定の技術領域だけで製品が構成されるような産業構造はかつてのものとなりつつあり、多くの産業における製品は複数の要素技術同士の組み合わせによって成り立つ。また、科学技術の多様性が高くなりつつある中でいまやどの要素技術同士が突然関係性を有しても不思議ではない。産業と技術が 1:1 に対応するような場合、すなわち一つの産業において中核となる技術が一つ程度に代表されるような場合においては、企業における研究開発では既存の技術を代替するような技術を開発する方向性が取られてきた。そのような方向性を実現するためには、特定の技術的性能を向上させていくようなブレークスルーアプローチが有効であった [156]。一方で産業と技術の対応関係が 1:N、すなわち一つの産業において必要な技術は複数かつ多様である場合、においては特定の性能や特定の技術のみに注力し研究開発を進めても最終的な製品に直結するとは限らない。このような産業の研究開発や製品開発において必要な視点は既存の技術同士の融合性に着目し、あらたな組み合わせを提示することで価値を創出することにある。Kodama(1992) は、後者の現象を技術融合 (Technology Fusion) と呼んでいる [156]。かつて光学分野と電子工学分野が融合することで、新たな分野としてオプトロニクスが生み出されたことや、機械工学分野と電気工学分野の融合が産んだメカトロニクス革命が工作機械業界に変革をもたらしたこともこの技術融合によるものと考えられる [156][157]。近年ではバイオテクノロジーと情報技術の融合としてバイオインフォマティクスなども技術融合の一例である。イノベーションは新結合と訳されることから、新たなものをゼロから作り上げるのではなく既存のものを融合することで価値を生み出すことであるという認識に基づけば、技術融合の本質は、イノベーションそのものであるといえる [158][159]。

技術融合は、より広義には Technology convergence, fusion, merging, cross-fertilization, hybridization などと呼称される [160]。Curran(2013) によれば、この融合という現象は 2 つの技術領域が有する境界線がぼやけていく (blurring) 現象であると定義している [161]。この定義は、技術に限らず、マーケッ

トや産業においても同様に適用することが可能であり、実際にこれらを対象とした研究も少なくない。例えばマーケットにおける融合現象においては、企業が出すプレスリリースから製品仕様情報を分析することでその現象を説明する試みが行われている [162]。また、産業における融合現象に関する研究では、企業や産業の属性情報と特許情報をともに用いる手法が適用されることが多い [163][164]。

一方で、科学技術における融合を議論する研究では多くが論文書誌情報を対象にするがその分析手法は多様である。共起語分析によるもの [165][166] や、共引用によるもの [167][168]、共著分析によるもの、ジャーナルのカテゴリ共起によるものなどにわけられる。技術を対象とした融合については特許情報を対象にした IPC の共起によるもの [169] や、共引用によるものが存在する [170]。このように、科学技術を対象とする融合をテーマとした研究では科学計量書誌情報を用いることが多い。

4.1.2 書誌情報に基づく知識発見とリンク予測

書誌情報を用いた代表的な知識発見手法としては Swanson(1986) によるモデルが挙げられる [171]。Don R.Swanson は書誌情報をもとに、レイノー病と魚油の間に当時まだ知られていなかった関係性を仮説付け、実際に検証することに成功している。Swanson はまず、魚油を摂取することによって血液粘性、血小板凝集、血管反応の抑制の効果があることを発見した。一方で、血液粘性や血小板凝集の抑制がレイノー病に効果があるということも発見した。これらの発見は、書誌情報に基づきそれぞれの分野の情報から独立に抽出された知見である。この知見をもとに、Swanson は「魚油を摂取することで、レイノー病を予防することができるはずである」という仮説を導き出した。シンプルな三段論法ではあるが、魚油の効果に関する知識とレイノー病の病理的知識の双方のドメインの知識がなければつなげることの出来ない仮説発見である。Swanson は情報学の学者であるがゆえ、魚油についてもレイノー病についても専門的知識を持つ立場にない。あくまで魚油の書誌情報とレイノー病の書誌情報それぞれから偶然的に抽出された情報をつなぎ合わせることで新たな発見を導いた。ABC モデルと呼ばれることのあるこの手法は、2つのコンセプト (A と C) の間における隠れた関係性がその間を仲介する B というコンセプトによって明らかになることを意味している。Swanson は同様の手法で、マグネシウムが偏頭痛に効果的であるということや、アルギニンがソマトメジン C の分泌に効果的

であるということを見つけた [172][173]。このように書誌情報を用いてこのように新たな知見を見出す手法は、Literature-Based Discovery(LBD) と呼ばれ Swanson のこの事例を皮切りに多くの研究がなされてきている。

この Literature-Based Discovery の手法は、ネットワーク分析の文脈ではリンク予測の 1 手法と等価であると考えられることができる。ネットワークはノードとノードのあいだの関係性をリンクとよばれる紐帯で表現する系である。ネットワークにおいて繋がりが薄い部分すなわちリンクが疎な部分はネットワークの中に存在する隙間のことを構造的空隙とよぶ [174]。構造的空隙を埋めるネットワークは、企業がイノベーションを実現するために重要である。ネットワークがどの程度つながりが強いかは、クラスタリング係数で算出できる [175]。例えば 3 つのノードの間に存在するリンクの最大数は 3 であり、4 つのノードがある場合その間に存在するリンクの数、すなわち最大次数は 6 である。このように全てのノードが全てのノードと互いにつながるようなネットワークを完全グラフと呼ぶ。しかしながら現実社会に存在する実際のネットワークはリンクの張られ方に局所的な濃淡があり、完全グラフのように均質ではない。対象のネットワークの粗密の程度を完全グラフと相対的に比較する指標としてクラスタ係数があげられる。現実社会のネットワークはその多くにおいて次数分布がべき分布を表すことが知られている [176][177][178][179]。この性質はスケールフリー性とよばれる特徴であり実社会のネットワークの多くがこの性質を有する。すなわち、実際のネットワークはその性質上、繋がりの密度が濃い部分と繋がりの薄い部分に二分される。このように現実のネットワークには密度の疎な箇所がかならず存在する。従って対象をネットワークと見た場合、将来的にあるノード同士がリンクを有するか否かといった問題は関心のある課題である。この課題を解決する予測手法の一つがリンク予測である [180][181][160]。論文や特許といった書誌情報を、ネットワークにおけるノードとした場合でも同様である。例えば、Shibata et al.,(2012) は、論文間の引用ネットワーク予測の手段としてリンク予測手法を用いている [182]。

技術融合の文脈においても IPC の共起関係をもとにしたリンク予測の視点が提案されている [160][183]。Lee et al.(2015) は、一定の時間窓において生じた IPC の共起情報から Adamic-Adar 指標を算出して将来にリンクが生じた割合を示しているが、Adamic-Adar と予測結果との関係については十分に議論しておらず予測という観点では検証の予知があると考えられる [160]。Caviggioli (2016) も技術融合を特定するために IPC 情報を用いているが、技術の組み合わせ

わせという観点における視点について十分な議論に至っていない [183]。

4.1.3 技術的スピノフ

技術は本来の目的とは異なる目的で用いられる事がある。例えばしばしば、巨額の予算や長期的国家プロジェクトによって生まれた技術が、当初思いもよらなかった市場において適用される現象が観察できる。このような現象を技術的スピノフと呼ぶことがある。Chiang(1992) は、政府主導などのミッションによって生まれた技術が、民間セクターにスピルオーバーする現象のことを、技術的スピノフと定義している [184]。これは技術融合の一形態と考えることができる。

我々の身の回りでは多くの技術的スピノフ事例を目にすることができる。例えば車輻に搭載されているエアバッグの技術は、ひとつのスピノフ事例といえることができる。エアバッグに用いられる技術は、当初米国海軍における産業エンジニア John W Hetrick が魚雷で用いられている空気圧縮技術を応用して、自動車事故の安全性を高められることを思いついたことが発端となっている。宇宙構造物における薄膜展開構造物の折りたたみ技術にミウラ折りと呼ばれる折り方がある。宇宙太陽光パネルを代表とする薄膜展開構造物は、打ち上げ時は小型で軽量である必要がある一方で、軌道投入後は大きな面積を確保する必要がある。確実な展開のためには可能なかぎり単純な機構である必要があり、さらには同じ機構によって衛星軌道上で折りたたみが自由に行われることが望ましい。これは、1970年に宇宙科学研究所にて宇宙構造工学を研究していた三浦公亮によって、構造の新しい形 PCCP シェルが発見されたことがきっかけとなる。もともとは、航空機のような円筒形構造物は長さ方向に圧縮すると概不伸張有限変形で座屈するという構造破壊の研究として吉村慶丸が発表したものが元になっているが、三浦公亮はこれを構造の形として発表した。偶然にも、飲料缶製造会社の技術者が薄肉かつ高強度の缶構造の開発に取り組んでいた際に三浦の論文をみつけたことから、飲料缶への適用に至ったとされており、現在も飲料缶のパッケージとしてみる事ができる。この折り方は英国折紙協会によって「ミウラ折り」と名付けられ、その後インテリア、携帯用地図、スタッドレスタイヤの溝などにも用いられることとなった。

ハードウェアのみならず、ソフトウェアにおいてもこのようなスピノフ事例は少なくない。家庭用ロボット掃除機を開発、販売する iRobot 社の売上を中心事業は軍用ロボットであり、家庭用ロボット掃除機地震も要素技術は地雷

除去プログラムがもととなっている。

また、国境を超えることで、オリジナルと異なる用いられ方をする技術も少なくない。特に米国におけるいくつかの軍事技術は、国を変え日本においては市民の身の回りを支える技術となっている。形状記憶合金は当初、戦闘機の内燃機関用のパーツとして米国において用いられたが、日本においては肩パッドや下着などのワイヤーとしての用途が有名である。チタン合金もまた、戦闘機の胴体において用いられる合金であったが、日本においてはフライパンを始めとする調理用器具として用いられることになる。新幹線のブレーキとして用いられている技術も当初は米国における航空機のエアブレーキが元になっている。このような技術が当初の目的と異なる適用がなされる事例は、官需から民需にかぎらず多く見ることができる。

NASA では宇宙開発技術をシーズとするような技術的スピノフを 2000 程度アーカイブしている [185, 186]。技術的スピノフ自体はその定義上、特定の産業において、多く見出すことができる傾向にあり、航空宇宙産業、原子力産業、防衛産業において生まれた技術がもととなることが多い。このような産業は、研究開発として先進的な技術が必要であり、莫大な財源はもとより長期にわたる継続的投資を必要とする産業である。こういった産業においては、公共資源が費やされることもあり、生まれた技術が本来の目的として有用であることを示す説明責任が求められる。従って、この産業で生まれたものの当初の目的としてはうまく活用することのできなかつた技術の有用性を示すという観点でも、既存の技術の融合ポテンシャルを特定することによって、あらたな技術的スピノフとして提案することは有用であると考えられる。

技術的スピノフによるとされる多くの事例は、リニアな研究開発によって生み出すことは困難であろうことが容易に想像できる。技術的スピノフの誕生は偶発的に発見されたようなものが少なくない。技術的スピノフは、あらかじめ計画にない外生的な要因によって、研究開発の主流を飛び越えて生まれたものである [184]。例えば、レーダー技師であったパーシー・スペンサーは、作動中のマグネトロンの前に立っているときに、偶然にもポケットの中に溶けたキャンディを見つけたことで、今日の電子レンジの原理を思い立ったとされている。このキャンディの偶然的発見こそが、全人類にとって運命的なものとなった。

技術的スピノフとなるポテンシャルをあらかじめ特定し、その市場や適用先を見つけ出す事ができれば、新たな市場の創出や雇用の創出にもポジティブ

な影響をあたえることができる。死蔵していた技術に新たな息を吹き込むことができる。NASA は、積極的に技術的スピノフを模索する試みをしているが、その目的は航空宇宙への研究開発投資に起因する実用的可能性を促進することであるとしている。実際に、技術的スピノフの情報が政策立案者、産業、公共セクタに共有する手法を提案されるなど、技術的スピノフの活用の関心は高い [186]。これまで、技術的スピノフに基づく研究としては特許情報の前方引用にもとづいた手法の提案 [187] がある。しかしながら、特許情報における直接引用関係は、学術論文情報におけるそれと比較して疎であることが知られている。また、公開公報になって初めて引用対象となるため、論文以上にサイテーションラグが存在してしまう点に課題がある。本研究では、技術融合を予測する手法のひとつとして、これまで共起がなかったにも関わらず特定のタイミングで共起する IPC に注目をする。IPC の共起という観点であれば上述のようなサイテーションラグの問題や直接引用のようなスパースティの問題から比較的解放される。

本研究の着想として、技術融合のなかでも特に距離が遠い技術同士の融合のひとつを技術的スピノフとした。具体的にはセクションが異なる IPC が付与された特許は、異なる分野に融合を図ろうとしている特許であるとし、技術的スピノフが発生した一形態を表すものと定義する。本研究の目的は、これまで関係性を持つことがなかった要素技術同士が突然関係を持つ技術融合の事象のうち特に距離が遠い技術同士のそれを技術的スピノフであると定義し、それらを事前に予測する手法を提案、検証することである。ネットワークにおけるリンク予測問題で用いられる Common Neighbor などの指標は、スワンソンの ABC モデルと同義であると再解釈することができると考え、技術融合の予測は要素技術間のリンク予測の問題として解くことができると考えるものである。

4.1.4 国際特許分類 (IPC) の活用

特許分析では前述の通り引用関係が論文のそれと比較して極めて疎であることが指摘される [188][189]。このことは、学術論文コミュニティとくらべて知識の繋がりを意味する関係性を抽出しにくい事を意味する。また、特許公報に用いられる単語の表記をより拡大して記述する傾向にあるといった特性がある。たとえば、ハードディスクはその技術的特性から特許公報においては磁気記録媒体と表記することが権利範囲上望ましいと判断される。すなわち同じ技術であっても分野によって異なる表現がされることもあり、また単語の

表記ゆれの問題も議論される。この特性上、一般的に用いられる単語は愚か
学術論文で用いられる単語でさえ整合性がとりづらい。また、特許情報同士
の引用関係や類似度関係が、発明者同士の人的関係と無関係である場合が多
い。学術論文であれば、同分野の研究をしている以上少なくとも一方が見知っ
ている事が多いが、特許についてそのような現象は起こりにくい。特許分析で
テキストマイニングを用いた書誌情報分析をする際には、このような課題が
情報の一意性を毀損することが少なくない。このような課題があるなかで特
許分析を行う方法の一つでは特許分類とよばれる特定の分類体系の情報を用
いることがある。特許分類は技術の特徴を類型化したものであり、検索やイ
ンデキシングを目的にすべての特許に付与されるタグである。特許分類は複
数存在し、とくに各国によってことなる特許分類の体系が存在する。欧州で
は、European Classification System (ECLA) 及び In Computer Only(ICO)、
米国では、United States patent classification(USPC) が用いられている。ま
た、日本特許庁では日本特許分類 Japan Patent Classification(JPC) や File
Index(FI) 分類が用いられてきており、さらに詳細な特許分類として F ターム
という独自の特許分類を定めている。各国において特許の分類方式が変わるの
はその国の文化や技術的背景に基づくものであり、必然的なものであるが技術
や知識のグローバル化によって統合したものを採用することが合理的であるとい
う考え方がすすんでいる。たとえば、2013 年においては欧州と米国の共同特許
分類 Cooperative Patent Classification(CPC) が用いられ始めている。国際特
許分類 (International Patent Classification:IPC) は、特許分類のうち国際的
に共通して用いられる特許分類であり、国際特許分類に関するストラスブール
協定によって 1971 年に基づいて作成されたものである。IPC もほかの技術分
類同様全ての特許公報に対して付与される技術タグの分類である。世界的所
有権機関によれば、IPC はおおよそ 7 万エントリーの技術要素が階層的構造に
よって構成されている。図 30 は IPC の一例を表したものである。最も上位レ
ベルはセクションと呼ばれる階層で 8 分類される。セクションは、A (生活必
需品)、B (処理操作：運輸)、C (化学；冶金)、D (繊維；紙)、E (固定構造
物)、F (機械工学；照明；加熱；武器；爆破)、G (物理学)、H (電気) にわ
けられる。これらのセクションは、それぞれ、クラスと呼ばれる第二階層から
構成される。クラスは、さらにサブクラスの階層によって構成され、サブクラ
スの階層はグループの階層によって構成される。グループはメイングループに
よって構成され、メイングループはさらにサブグループによって構成される。

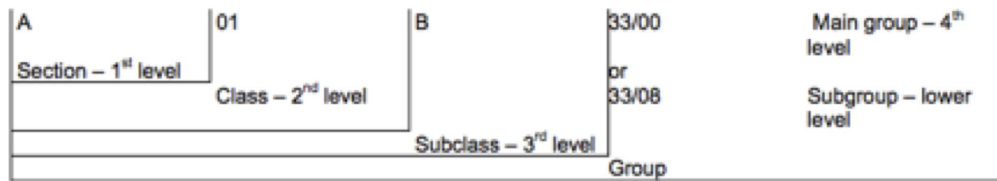


図 30. 国際特許分類 IPC の階層別名前 (Source: (WIPO,2014))

IPC を始めとした特許分類は、特定の技術が階層構造によって示されるため、下階層であるほどより詳細な技術分野を表す。これら IPC は一つの特許公報に対して複数付与される。複数の IPC が付与される特許は、複数の異なる技術によって構成されると解釈できる。例えば、上位層は共通であり、具体的な技術要素をあらあわす下階層部分だけが異なるタグが複数付与されるような特許技術は、ある特定の分野に構成される。一方で、セクションが異なる IPC が複数付与されるような特許は、分野横断的な技術によって構成されている特許であると考えられることができる。例えば電気分野に関わる特許はその多くがセクション H(電気) に分類される IPC で構成されると考えられる。しかしながら、中には他のセクションである例えば C (化学 ; 冶金) セクションを有する IPC と共起するような特許も存在する。

これまで、技術融合のありかたのひとつとして IPC に着目する発想自体は [160][183] などが挙げられるが、彼ら自身が Future work として掲げている点の一つに、特許分類情報が持つ階層構造情報を役立てることが出来ていないという点が存在する。本研究では、これまで十分に扱われることのなかった階層構造情報に着目し、階層情報が技術融合予測に役立てることを示す事を目的とするものである。

4.2 手法

本研究は、特許公報が持つ技術タグである国際特許分類 (IPC) の共起ネットワークに着目し、それぞれの階層におけるネットワークを構築することで、リンク予測の手続きを適用することによって技術的スピノフに代表される技術融合のポテンシャルを特定できると考えた。31 に、予測対象となる共起関係と階層関係の図を示す。

前述の通り IPC は複数の階層構造によって技術階層を類型化するツリー構造のデータとなっている。本研究ではこれらの階層について、サブクラス

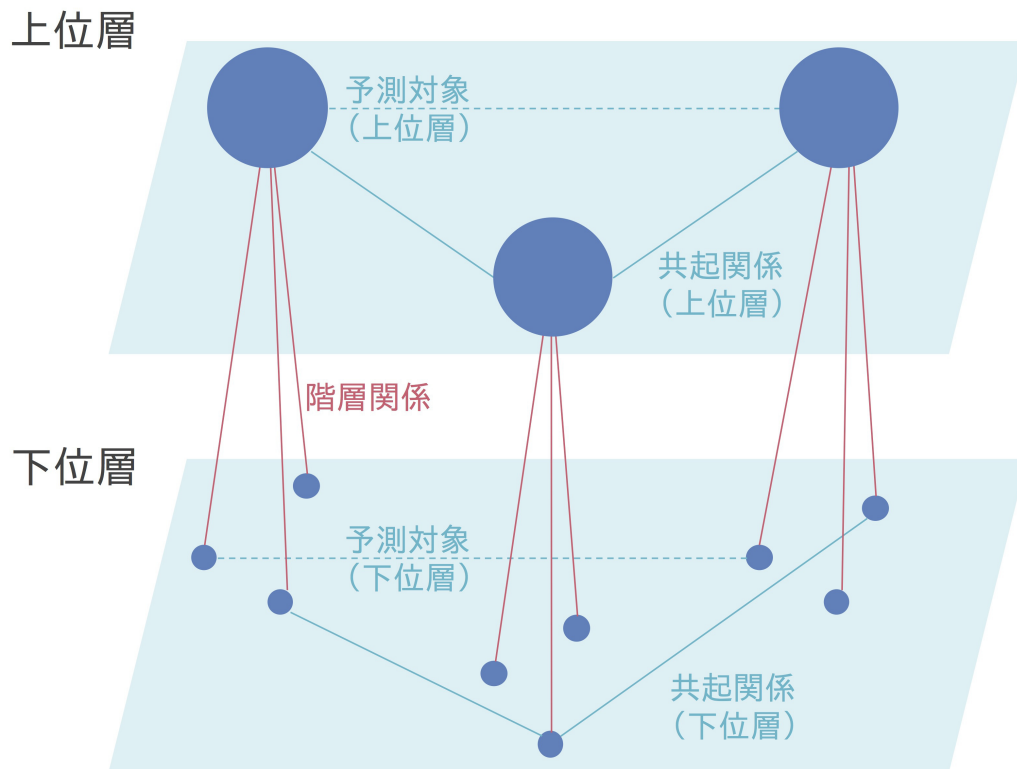


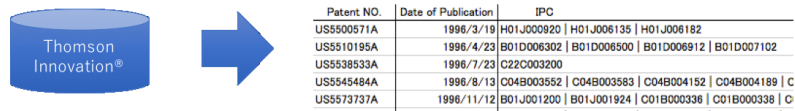
図 31. 階層概念

(4-digit)、グループ (8-digit)、サブグループ (10-digit) の3階層に分けて考えることとする。すなわち、通常のリンク予測問題ではそれぞれの階層の中でリンク予測指標を算出し予測を行ってきた。本研究では、異なる階層の情報を用いることでリンク予測パフォーマンスに与える影響を考察する。

また本研究では、この最上位階層であるセクションが異なる IPC のうち、これまで共起が起きなかつともかかわらずある特定の時点に於いて共起するような IPC のペアを技術的スピンオフポテンシャルを有するペアであると定義した。

図 32 に予測分析のプロセスを示す。まず、対象となるデータセットから必要なフィールドとして特許公報番号、公開公報年月日、IPC を抽出する (ステップ 1)。本研究では、データソースとしてトムソン・ロイター社が提供する特許公報提供サービスである”Thomson Innovation”を用いる。本サービスは、網羅的に世界中の特許情報を取り扱っている包括的データベースである。このサービスのうち、トムソン・ロイター社によって付加価値情報が付与された特許情報データベース DWPI(Darwent Patent Index) を用いる。DWPI では特許

ステップ1. 特許公報番号、公開公報年月日、IPCを抽出



ステップ2. 各時間におけるIPC共起ネットワーク(時間拡張ネットワーク)の構築



ステップ3. 予測精度の算出

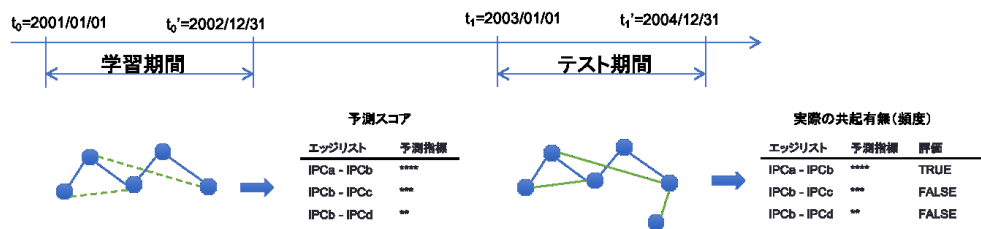


図 32. 分析プロセス概要

ファミリーがひとつのレコードとして集約されるため、重複した技術を異なるレコードとして取ることを避けることができる。対象となる特許群から、IPCの共起ネットワークを構築する(ステップ2)。各年月においてそれまで、共起していたIPCペアすべてを対象とする。IPCをノード、共起関係をリンクとした累積情報のネットワークであることから、ノードおよびリンクは時間の経過とともに増加する時間拡張ネットワークである。

本研究では、分析対象となる時間軸の全体を、学習期間 $t_0 \sim t_0'$ とテスト期間 $t_1 \sim t_1'$ に分けて、時間窓を定義する。複数の条件を検証するため、学習期間については2年間、3年間、6年間の3種類の長さの窓を定義し、またテスト期間においては1年間、2年間、3年間の3種類の長さの窓を定義した(表11)。学習期間やテスト期間によって予測精度を算出しモデルの安定性を検証する(ステップ3)。

ネットワークの類似指標にもとづくリンク予測は与えられたネットワークの局所的な構造特性に依存する。リンク予測問題における基本的な考え方は、ネットワークの任意のノードが共通した隣接ノードを有する場合それら任意のノードの間にもリンクが生じうるであろうという考え方である。一般的にネットワークにおけるリンク予測は、任意の2つのノードの間にリンクが生じうる

表 11. 学習期間窓およびテスト期間窓の組み合わせ

学習開始年:t0	学 習 終 了 年:t0'	テ ス ト 開 始 年:t1	テ ス ト 終 了 年:t1'	学習期間 [年]	テ ス ト 期 間 [年]
2011	2013	2014	2015	2	1
2009	2013	2014	2015	4	1
2007	2013	2014	2015	6	1
2010	2012	2013	2015	2	2
2008	2012	2013	2015	4	2
2006	2012	2013	2015	6	2
2009	2011	2012	2015	2	3
2007	2011	2012	2015	4	3
2005	2011	2012	2015	6	3

か否かをそれぞれ正例、負例とする 2 クラスの分類問題として解くことが多い。本研究では 2 クラス分類だけでなく、徐々に共起情報が増加していくさまを予測することも対象にするため、回帰問題としての予測も行う。

また、これまで IPC を用いて技術融合における現象を説明する研究ではいずれもリンクの有無のみを特定する重みなしのネットワークとして扱ってきた。しかしながら、漸近的な技術の融合過程を特定するには共起情報は重み付きとして考えることが望ましいと考える。本研究では、IPC 同士の共起を特定するための指標に対して次の重み付き指標を用いる。本研究では、IPC の共起の予測を行う予測指標として 4 つの重み付きリンク予測指標を用いる。それぞれ、重みつき共通隣接ノード指標 S_{WCN} 、重み付き資源配分指標 (WRA) S_{WRA} 、重み付き Adamic-Adar 指標 S_{WAA} 、また、それらの合計 S_{sum} である。

重みつき共通隣接ノード指標 S_{WCN} は式 3 のように定義できる。

$$S_{wcn}(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} w(x, z) + w(z, y) \quad (3)$$

任意の IPC をノードとし、2 つのノードの間に生じるリンクは特許において生じた共起関係を意味する。リンクの横に付したそれぞれの重みを当該 2 つの IPC 間において生じた共起回数を意味する。共通隣接ノード指標は、共通の友人が多い 2 人は友人になる可能性が高い事を意味する指標である。本研究で取り扱う重み付きの共通隣接ノード指標は、IPC 同士の仲の良さのようなものを考慮することとなる。実際には過去の学習期間に共起した回数が重みとして考慮される。

重み付き資源配分指標 S_{WRA} は式 4 の通り定義される。

$$S_{wra}(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(z, y)}{s(z)} \quad (4)$$

重み付き Adamic-Adar 指標 S_{WAA} は式 5 の通り定義される。

$$S_{waa}(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x, z) + w(z, y)}{\log(1 + s(z))} \quad (5)$$

重み付き資源配分指標 S_{WRA} および、重み付き Adamic-Adar 指標 S_{WAA} は、いずれも重み付き共通隣接ノード指標を元にした指標であり、人づき合いの悪い友人を共通の友人として持つ 2 人は友人になる可能性が高いと解釈することができる。

さて、元田 (2011) は、リンク予測において多種類の類似度指標を足し合わせることで新たな指標を定義し、それによってより高い精度が得られることを示している [190]。本研究で同様に、式 6 のように、WCN、WRA、WAA の和も特徴量として用いる。

$$S_{sum}(x, y) = S_{wcn}(x, y) + S_{wra}(x, y) + S_{waa}(x, y) \quad (6)$$

以上の予測指標をサブクラス (4digit)、グループ (8digit)、サブグループ (10digit) 全ての IPC 粒度において算出する。

4.2.1 予測モデル

予測対象はそれぞれの IPC 層を対象にし、予測指標の組み合わせによって複数のモデルを構築する。大きくわけて、下位情報を用いて上位層の予測を行う場合と、上位情報を用いて下位層を予測する場合の 2 つに分けることができる。まず、下位情報を用いて上位層の予測を行う場合については次の組み合わせが存在する。

- サブクラス (4digit) における予測に、グループ (8digit) の情報を用いた場合、サブクラス (4digit) 自身の情報を用いた場合、グループ (8digit) とサブクラス (4digit) の双方を用いた場合の 3 種類のモデルを構築する。
- グループ (8digit) における予測に、サブグループ (10digit) の情報を用いた場合、グループ (8digit) 自身の情報を用いた場合、サブグループ (10digit) とグループ (8digit) の双方を用いた場合の 3 種類のモデルを構築する。
- サブクラス (4digit) における予測に、サブグループ (10digit) の情報を用いた場合、サブクラス (4digit) 自身の情報を用いた場合、サブグループ (10digit) とサブクラス (4digit) の双方を用いた場合の 3 種類のモデルを構築する。

一方、上位情報を用いて下位層の予測を行う場合については次の組み合わせが存在する。

- グループ (8digit) における予測に、サブクラス (4digit) の情報を用いた場合、グループ (8digit) 自身の情報を用いた場合、グループ (8digit) とサブクラス (4digit) の双方を用いた場合の 3 種類のモデルを構築する。
- サブグループ (10digit) における予測に、グループ (8digit) の情報を用いた場合、サブグループ (10digit) 自身の情報を用いた場合、サブグループ (10digit) とグループ (8digit) の双方を用いた場合の 3 種類のモデルを構築する。
- サブグループ (10digit) における予測に、サブクラス (4digit) の情報を用いた場合、サブグループ (10digit) 自身の情報を用いた場合、サブグループ (10digit) とサブクラス (4digit) の双方を用いた場合の 3 種類のモデルを構築する。

本研究において対象となる予測問題は分類モデルと回帰モデルそれぞれについてモデルを構築して予測を行う。分類モデルでは、表 11 における学習期間において構築された共起 IPC のみによって生成されるネットワークを元に予測指標を計算する。ここで計算された指標を予測モデル生成のための特徴量とする。また、学習期間でリンクが生成されなかったが、検証期間でリンクが生成されたペアを正例として定義し、このフラグをラベルとする。すなわち 2 クラス分類と定義し、特定の IPC 間の共起有無を予測するモデルとする。

本研究では、回帰モデルとして定義した予測モデルも構築する。回帰モデルでも分類モデルと同様に表 11 の学習期間において構築された共起 IPC のみによって生成されるネットワークを元に予測指標を計算する。検証期間で構築されたリンクの重みを予測対象とする。

分類モデルと回帰モデル双方に適用可能な手法としてランダムフォレストによる予測モデル構築を行う [191]。ランダムフォレストは、決定木を用いたアンサンブル学習（集団学習）の一つであり、分析の性質上特に非線形な問題についても高いパフォーマンスを実現することができる手法の一つである。また目的変数に対する特徴量の重要度を推定することができる上、欠損値を持っていてもある程度有効に動作することが知られている。ランダムフォレストでは訓練データに対してサンプルのブートストラップと説明変数のサンプリングを行って複数の決定木を構築し、テストデータに対して各決定木の予測結果の多

数決（分類問題）もしくは平均値（回帰問題）によって予測を実行するものである。決定木の代表的なアルゴリズムである、CART アルゴリズム [192] では例えば条件ノード A を条件ノード A_L と条件ノード A_R に分けるときに次に示す ΔI を最大化する分割を行うそれぞれ分類木と回帰木に応じて次の評価基準で木を構築していく。分類木では

$$\Delta I = P(A)I(A) - P(A_L)I(A_L) - P(A_R)I(A) \quad (7)$$

この場合、Entropy を基準とする場合 $I(A) = -\sum p(k|A) \log p(k|A)$ であり、GINI 係数を基準とする場合 $I(A) = 1 - \sum_k p(k|A)^2$ で定義される。また $p(k|A)$ は条件ノード A でクラス k を取る条件付き確率を意味する。一方で回帰木の場合、 ΔI は次で与えられる。

$$\Delta I = \frac{S(A) - S(A_L) - S(A_R)}{N} \quad (8)$$

なお、この時の $S(A)$ は尤離度 (deviance) とよばれ、 $S(A) = \sum_{k \in A} (t_k - t_A)^2$ で与えられ、 t_A は条件ノード A での目的変数 t の平均値を意味する。ランダムフォレストでは、まず学習データからランダムに特定のサンプル集合を抽出する。つぎに、それぞれのサンプル集合において特徴量をランダムに抽出する。全てのサンプリングセットにおいて同じようにランダムに抽出した特徴量で決定木モデルを構築し、複数の決定木から総合的な予測モデルを構築する。本研究では計算には、上記のランダムフォレストアルゴリズムを高速処理できるように実装されたパッケージである ranger を用いる [193]。また、分析には R を用いる [194][195]。

正例データと負例データのバランスが不均衡である事が少なくない。通常、実データにおける不均衡データは正例のサンプルサイズが少ない場合が一般的であるが、本研究におけるデータもその定義上正例のサイズが負例よりも少ないことが十分に想定される。不均衡データに対する対処には、正例を誤答したときのペナリティを重くする方法と、サンプリングサイズを調整する方法の二種類が代表的である。また、後者には正例サンプルサイズを負例にあわせるオーバーサンプリングと、正例のサイズに合わせて負例を減らすアンダーサンプリングが挙げられるが、本研究ではアンダーサンプリングを採用する。

4.2.2 評価指標

分類モデルにおける評価基準には Precision、Recall、Fmeasure を採用する。それぞれ次式で定義される評価指標であり、2 クラス分類において極めて一般

的に用いられる指標である。適合率 (precision) はモデルが正と予測したデータのうち実際に正であるものの割合である。

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

一方、再現率 (recall) は実際に正であるデータのうちモデルが正と予測したものの割合である。

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

F 値は適合率と再現率の調和平均であり、次に定義される。

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision} \quad (11)$$

TP は、真陽性 (True Positive) を、 FP は偽陽性 (False Positive) を、 FN は偽陰性 (False Negative) を表す。Precision は正例であると予測されたサンプルのうち、実際に正例であるサンプルの割合である。Recall は実際に正例であるサンプルのうち、正例であると判定されたサンプルの割合である。一般的に Precision と Recall はトレード・オフの関係にある。本研究では、Precision と Recall の調和平均である Fmeasure もあわせて評価を行う。

また、回帰問題における評価基準には、予測器が出力した結果と、実際のテスト期間に生じたリンクの重みの間に生じる回帰直線がもつ調整済み決定係数 R^2 と平均二乗誤差 $RMSE$ を用いる。

以上より、本研究における予測分析のパラメータを整理すると表 12 のように整理できる。

まず、学習窓予測窓の組み合わせは表 11 の通り 9 種類である。次に、予測対象とする IPC のペアは、全ての IPC を対象としたものと、上位の IPC が異なる組み合わせのみを予測対象としたものの 2 種類である。さらに、予測対象とする IPC 層はサブクラス (4digit)、グループ (8digit)、サブグループ (10digit) の 3 種類である。次に、学習層と予測層の組み合わせは、下層情報を用いて一つ上層のリンクを予測する場合と、上層情報を用いてひとつ下層のリンクを予測する場合の 2 種類である。問題は全てにおいて 2 クラス分類と回帰問題の 2 種類の分析を行う。2 クラス分類における評価基準は、Precision、Recall、Fmeasure を用いる。回帰問題における評価基準は、予測値と実測値の回帰直線による調整済み決定係数 R^2 と最小二乗誤差 $RMSE$ を用いる。

表 12. 本研究における分析条件の整理

学習予測窓 (9種類)	学習期間 2,4,6 年 × 予測期間 1,2,3 年 表 11 参照		
予測対象 IPC (2種類)	上位 IPC が異なる組み合わせのみを対象	全組み合わせを対象	
予測対象層 (3種類)	4digit 層	8digit 層	10digit 層
学習層と予測層の組み合わせ (2種類)	下層情報を用いて上層を予測	上層情報を用いて下層を予測	
問題 (2種類)	2 クラス分類	回帰	
2 クラス分類問題における評価基準 (3種類)	Precision	Recall	Fmeasure
回帰問題における評価基準 (2種類)	調整済み決定係数 R^2	最小二乗誤差 RMSE	

4.3 結果

4.3.1 CFRP データセットにおける結果

4.3.1.1 データセット概要

用いたデータセットのうち CFRP に関するデータセットの概要を示す。CFRP データセットは抽出したデータのうち 1 月から 12 月のすべてのデータを含む 1978 年から 2015 年を分析の対象データとした。各年月において得られる IPC の共起関係は時間拡張ネットワークとして徐々に成長していく。ここでは年月ごとのネットワークにおけるグローバル推移を計算している。まずはじめに FGM に関する論文出版推移と、それに応じたユニーク IPC 数推移、IPC の共起数推移、ネットワークにおけるクラスタリング係数推移、を示す。図 33、図 34、図 35 は、それぞれ IPC におけるサブクラス (4digit)、グループ (8digit)、サブグループ (10digit) を対象とした結果である。

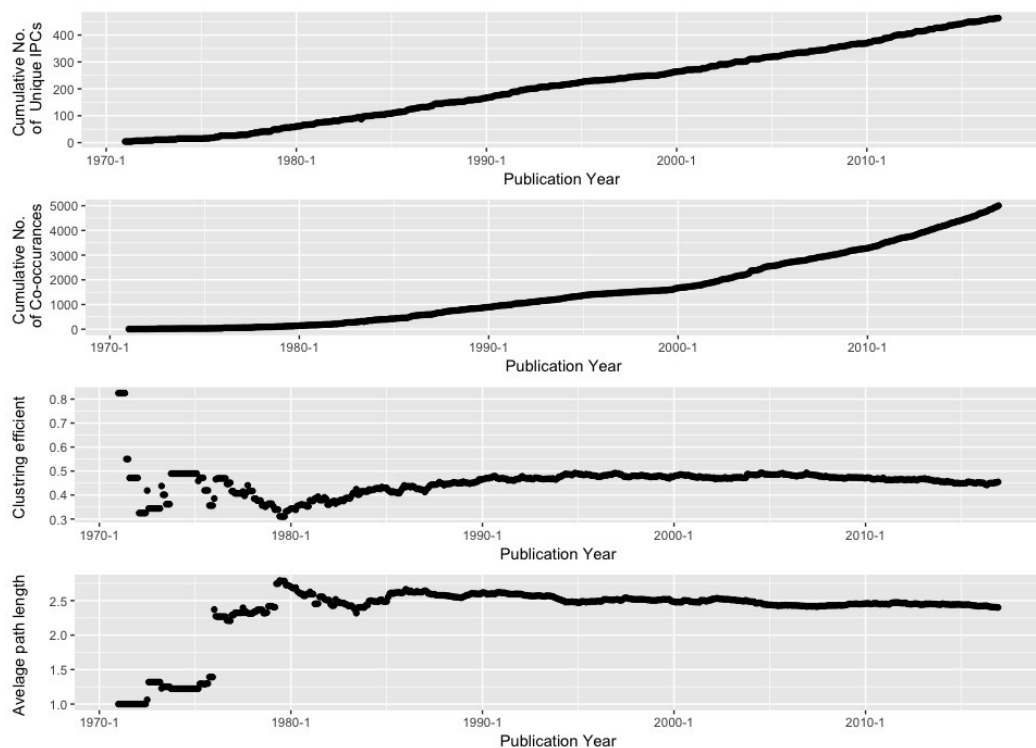


図 33. ネットワークにおけるグローバル指標の推移 (CFRP データセット：サブクラス)

CFRP データ・セットにおける IPC の共起のうち、データセットの最終年月である 2015 年 12 月における最終的に最も頻度の高い組み合わせは、サブク

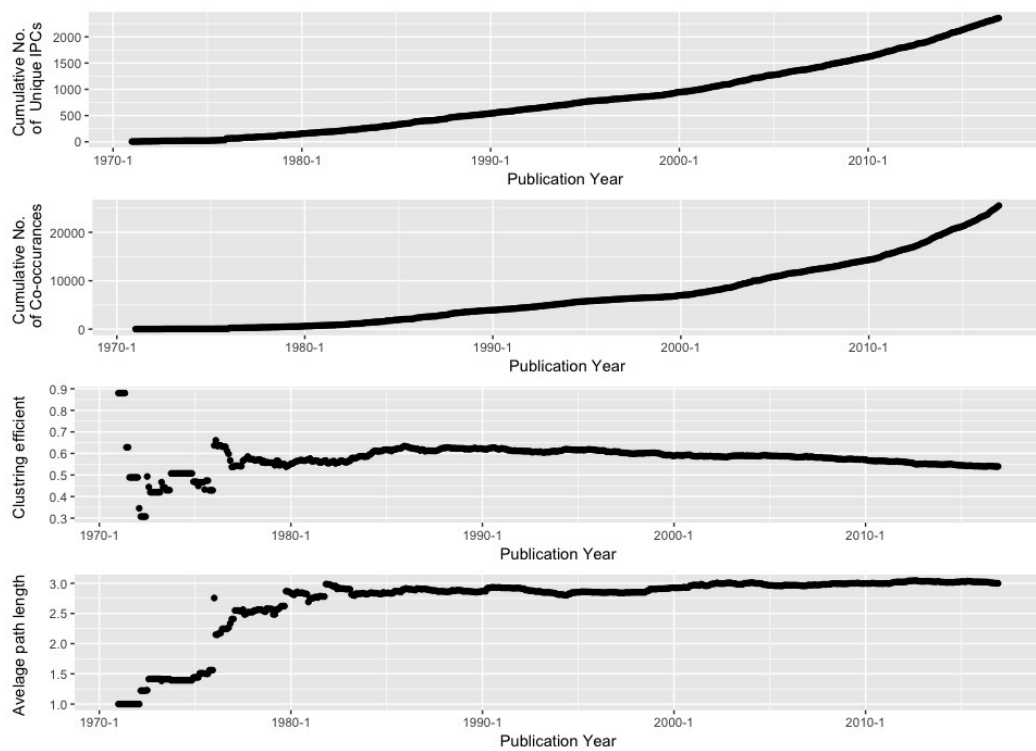


図 34. ネットワークにおけるグローバル指標の推移 (CFRP データセット：グループ)

ラス、グループ、サブグループにおいてそれぞれ表 13、表 14、表 15 にしめすとおりである。

表 13. 上位 5 ペア (CFRP データセット：サブクラス)

				共起数
C08K	無機 非高分子有機物質の添加剤の使用	C08L	高分子化合物の組成物	35,500
C08K	無機 非高分子有機物質の添加剤の使用	C08K	無機 非高分子有機物質の添加剤の使用	25003
C08L	高分子化合物の組成物	C08L	高分子化合物の組成物	16,548
B32B	積層体, すなわち平らな平らでない形状, 例. 細胞状 ハニカム状の層から組立てられた製品	B32B	積層体, すなわち平らな平らでない形状, 例. 細胞状 ハニカム状の層から組立てられた製品	10,449
D06M	繊維, より糸, 糸, 織物, 羽毛 このような材料から製造された繊維製品のクラス D 0 6 の他に分類されない処理	D06M	繊維, より糸, 糸, 織物, 羽毛 このような材料から製造された繊維製品のクラス D 0 6 の他に分類されない処理	6,595

表 14. 上位 5 ペア (CFRP データセット：グループ)

				共起数
C08K0003	無機 非高分子有機物質の添加剤の使用・・・無機配合成分の使用	C08K0007	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用	4,989
C08K0003	無機 非高分子有機物質の添加剤の使用・・・無機配合成分の使用	C08K0003	無機 非高分子有機物質の添加剤の使用・・・無機配合成分の使用	2,828
C08K0003	無機 非高分子有機物質の添加剤の使用・・・無機配合成分の使用	C08K0005	無機 非高分子有機物質の添加剤の使用・・・有機配合成分の使用	2,816
H01M0004	化学的エネルギーを電気的エネルギーに直接変換するための方法 手段, 例. 電池・・・電極	H01M0004	化学的エネルギーを電気的エネルギーに直接変換するための方法 手段, 例. 電池・・・電極	2572
C08K0005	無機 非高分子有機物質の添加剤の使用・・・有機配合成分の使用	C08K0007	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用	2,045

表 15. 上位 5 ペア (CFRP データセット：サブグループ)

				共起数
C08K000304	無機 非高分子有機物質の添加剤の使用・・・無機配合成分の使用・・・炭素	C08K000706	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用・・・元素	760
C08K000706	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用・・・元素	C08K001304	無機 非高分子有機物質の添加剤の使用・・・メイングループ 3 / 0 0 から 1 1 / 0 0 のいずれのメイングループにも包含されない配合成分であって, そのいずれの化合物も本質的な配合成分からなる混合物の使用・・・形状によって特徴づけられた配合成分および有機 無機の配合成分	556
C08K000706	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用・・・元素	C08K000714	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用・・・ガラス	515
C08K000706	無機 非高分子有機物質の添加剤の使用・・・形状に特徴を有する配合成分の使用・・・元素	C08L006300	高分子化合物の組成物・・・エポキシ樹脂の組成物 エポキシ樹脂の誘導体の組成物	403
H01M000413	化学的エネルギーを電気的エネルギーに直接変換するための方法 手段, 例. 電池・・・電極・・・非水電解質二次電池用の電極	H01M001005	化学的エネルギーを電気的エネルギーに直接変換するための方法 手段, 例. 電池・・・二次電池 その製造・・・非水電解質二次電池	367

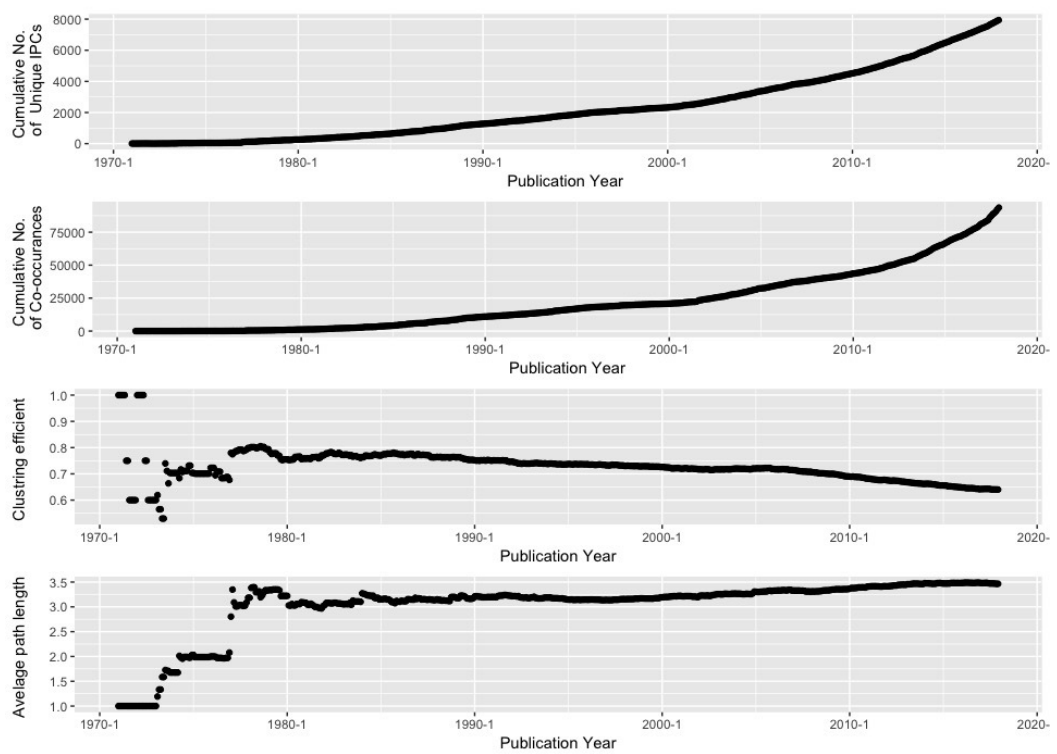


図 35. ネットワークにおけるグローバル指標の推移 (CFRP データセット：サブグループ)

4.3.1.2 モデル評価（分類）（上位層の予測に下位層の情報を用いる場合）

まず、CFRP データセットにおいて上位層の予測に下位層の情報を用いた場合の結果を示す。予測対象となる上位層の IPC に対応する下位層の IPC は複数存在するため、一つの統計要約量に集約することによって特徴量を代表させた。本研究では中央値を用いた。CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 36 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 37 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 38 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。



図 36. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]



図 37. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]



図 38. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

CFRP データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 39 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 40 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 41 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。



図 39. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [CFRP データセット全 IPC を対象]



図 40. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [CFRP データセット全 IPC を対象]



図 41. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [CFRP データセット全 IPC を対象]

4.3.1.3 モデル評価（分類）（下位層の予測に上位層の情報を用いる場合）

つづいて、CFRP データセットにおいて、下位層の予測に上位層の情報を用いた場合の結果を示す。予測対象となる下位層の IPC に対応する上位層の IPC は 1 つしかないため、上位層の予測とことなり、統計代表値によって集約する必要はない。CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 42 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 43 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 44 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

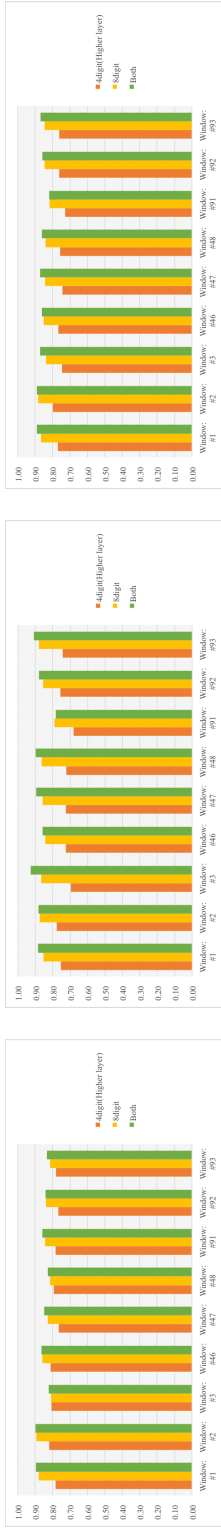


図 42. グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision (左図), Recall (中図), Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

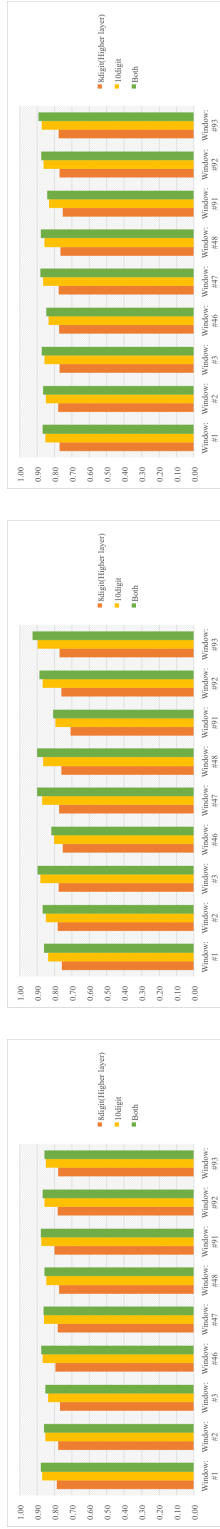


図 43. サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の Precision (左図), Recall (中図), Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]



図 44. サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision (左図), Recall (中図), Fmeasure (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

CFRP データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 45 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 46 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 47 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

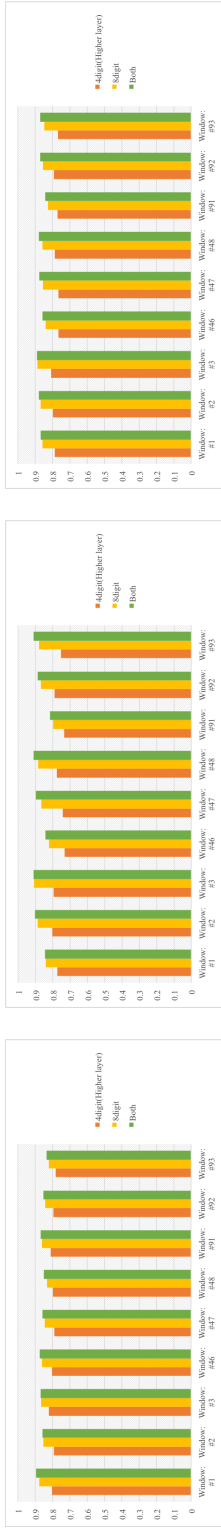


図 45. グループ (8digit) の予測に上層情報であるサブクラス (4digit) を用いた場合の Precision (左図), Recall (中図), Fmeasure (右図) [CFRP データセット全 IPC の組み合わせ対象]

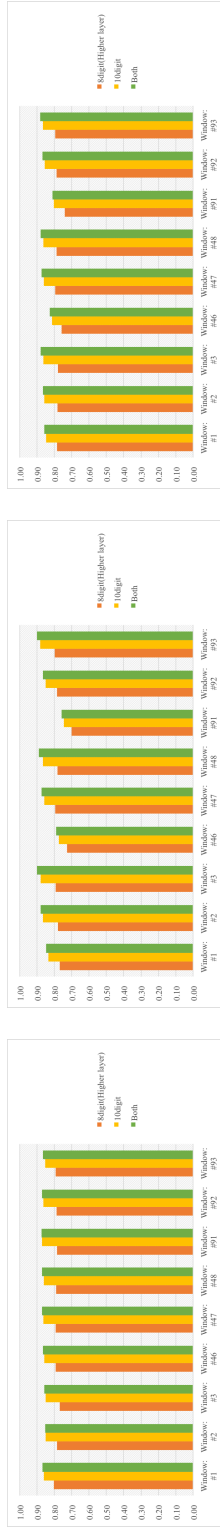


図 46. サブグループ (10digit) の予測に上層情報であるグループ (8digit) を用いた場合の Precision (左図), Recall (中図), Fmeasure (右図) [CFRP データセット全 IPC の組み合わせ対象]



図 47. サブグループ (10digit) の予測に下層情報であるサブクラス (4digit) を用いた場合の Precision (左図), Recall (中図), Fmeasure (右図) [CFRP データセット全 IPC の組み合わせ対象]

4.3.1.4 モデル評価（回帰）（上位層の予測に下位層の情報を用いる場合）

次に、CFRP データ・セットにおける回帰分析結果を示す。回帰モデルの評価は前述の通り、RMSE と調整済み決定係数の2つである。まず、CFRP データセットにおいて上位層の予測に下位層の情報を用いた場合の結果を示す。CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした調整済み決定係数（左図）, RMSE（右図）を図 48 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数（左図）, RMSE（右図）を図 49 に示す。CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数（左図）, RMSE（右図）を図 50 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

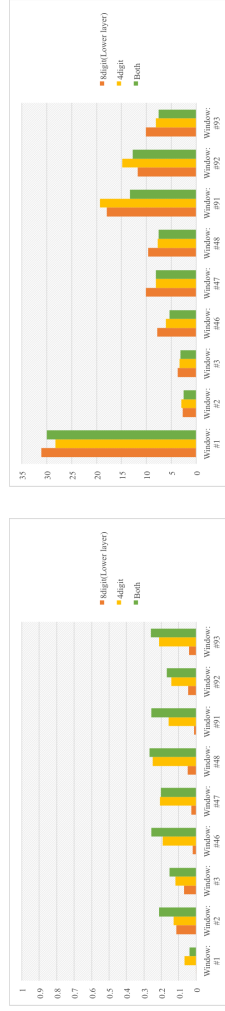


図 48. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

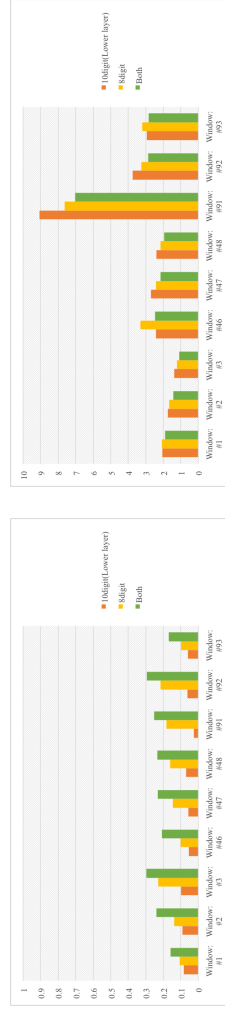


図 49. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

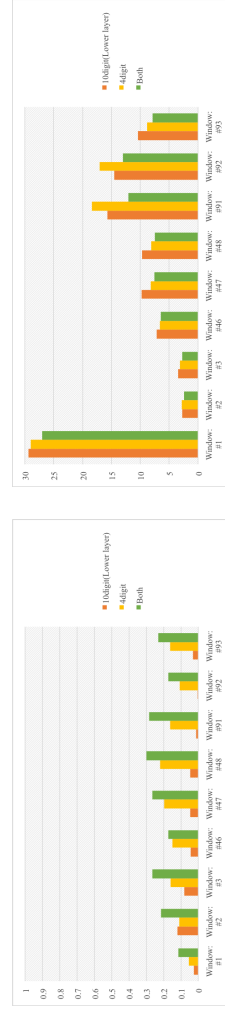


図 50. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

同様に、CFRP データ・セットに対して、全ての IPC を対象にした結果を示す。CFRP データセットにおいて上位層の予測に下位層の情報を用いた場合の結果を示す。CFRP データセットにおいて、全ての IPC の組み合わせを対象とした調整済み決定係数 (左図) ,RMSE (右図) を図 48 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 49 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 50 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

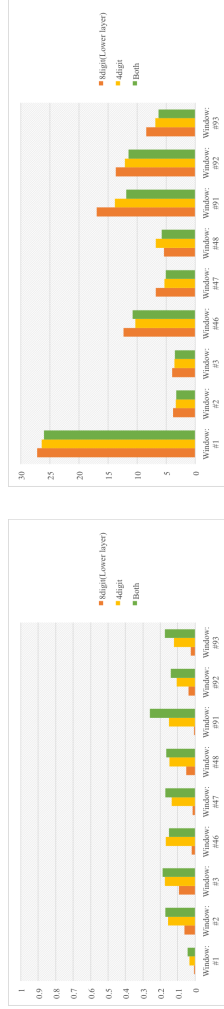


図 51. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]

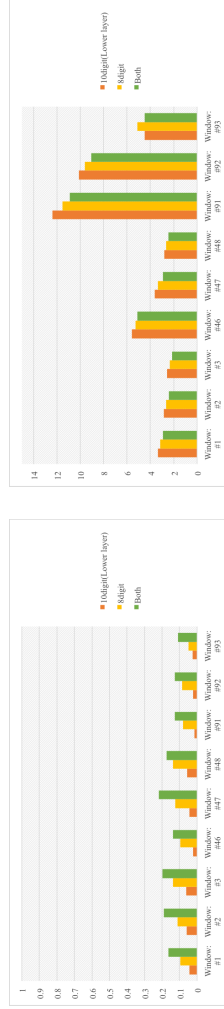


図 52. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]

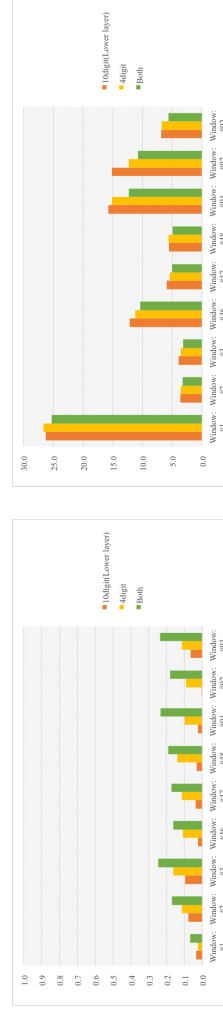


図 53. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]

4.3.1.5 モデル評価（回帰）（下位層の予測に上位層の情報を用いる場合）

次に、CFRP データセットにおいて下位層の予測に上位層の情報を用いた場合の結果を示す。CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 54 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 55 に示す。

CFRP データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 56 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

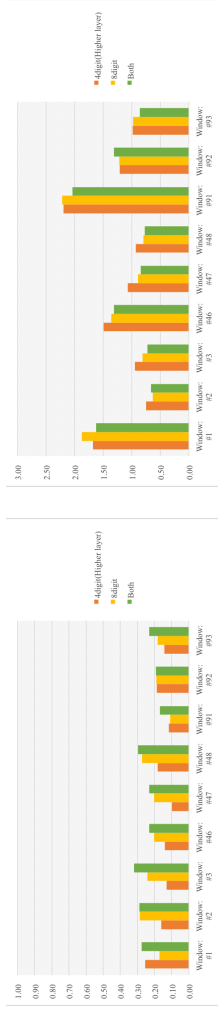


図 54. グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

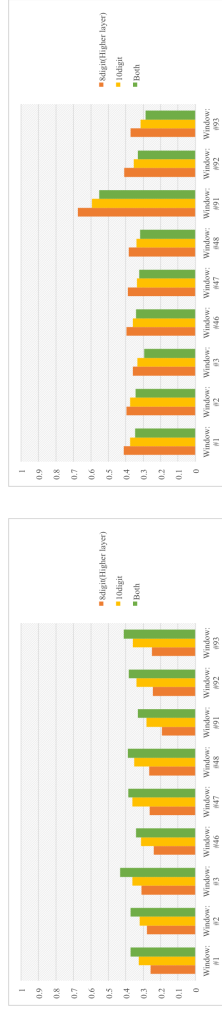


図 55. サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

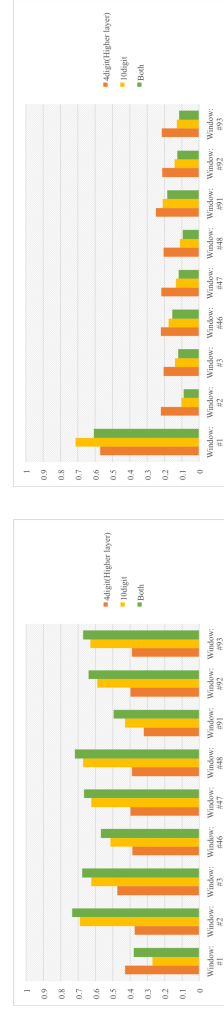


図 56. サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図), RMSE (右図) [CFRP データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

同様に、CFRP データ・セットに対して、全ての IPC を対象にした結果を示す。CFRP データセットにおいて、全ての IPC の組み合わせを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 57 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせを対象とした予測において、グループ (8digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 58 に示す。

CFRP データセットにおいて、全ての IPC の組み合わせを対象とした予測において、サブクラス (4digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 59 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

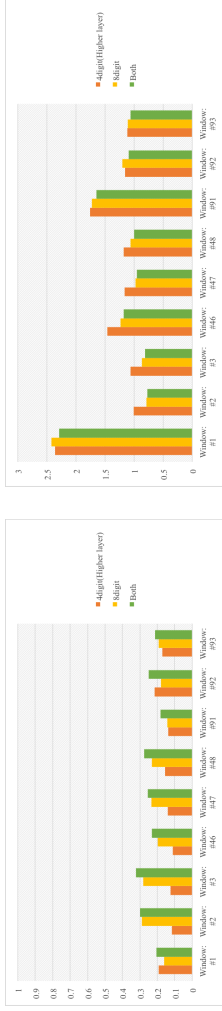


図 57. グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]

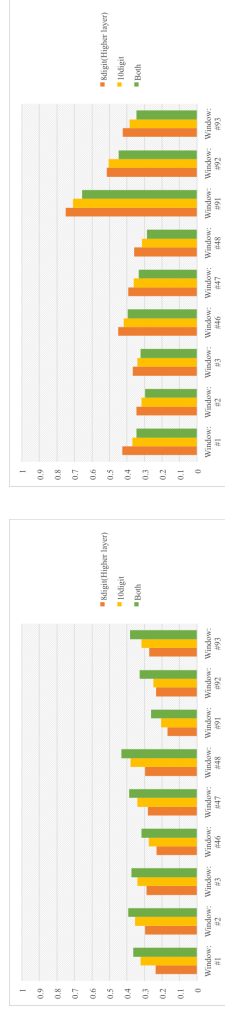


図 58. サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]

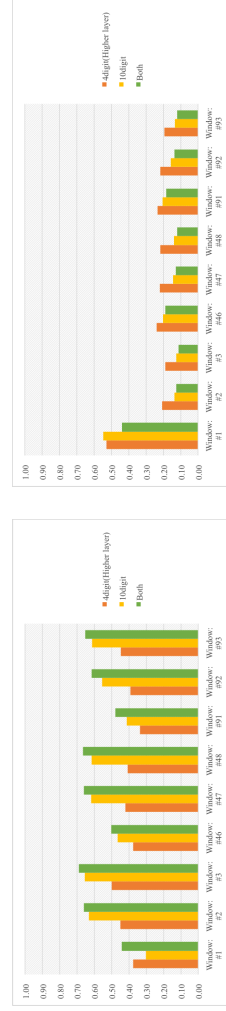


図 59. サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) 下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [CFRP データセットにおいて全ての IPC の組み合わせを対象]

4.3.2 FGM データセットにおける結果

4.3.2.1 データセット概要

用いたデータセットのうち FGM に関するデータセットの概要を示す。FGM データセットは抽出したデータのうち 1971 年 1 月から 2016 年 12 月の 10,953 特許を分析の対象データとした。各年月において得られる IPC の共起関係は時間拡張ネットワークとして徐々に成長していく。ここでは年月ごとのネットワークにおけるグローバル推移を計算している。まずはじめに FGM に関する論文出版推移と、それに応じたユニーク IPC 数推移、IPC の共起数推移、ネットワークにおけるクラスタリング係数推移、を示す。図 60、図 61、図 62 は、それぞれ IPC におけるサブクラス (4digit)、グループ (8digit)、サブグループ (10digit) を対象とした結果である。

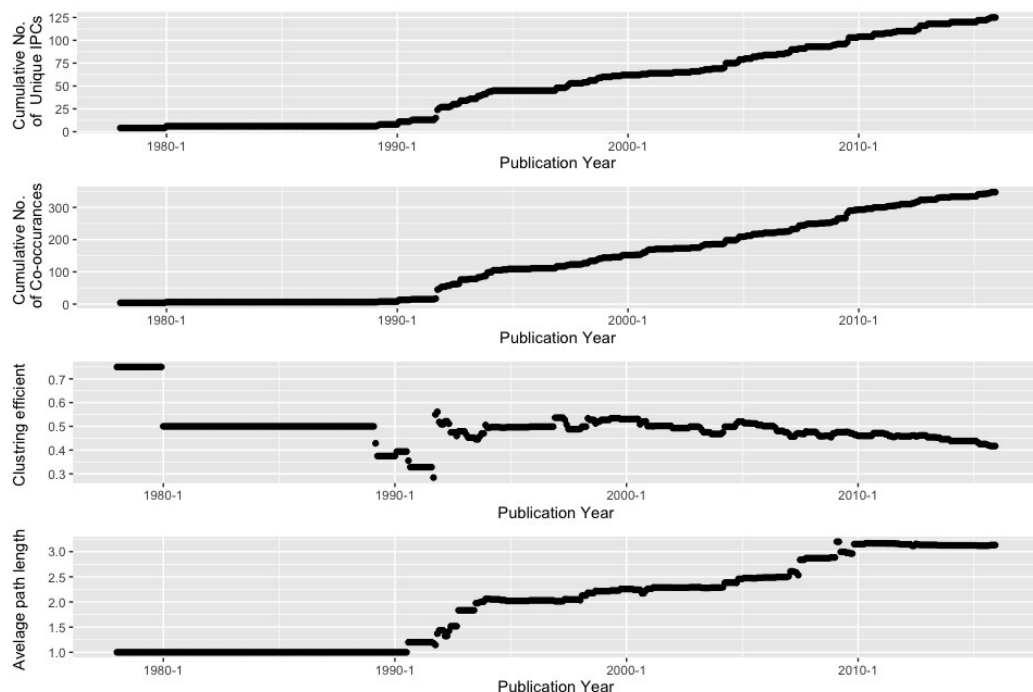


図 60. ネットワークにおけるグローバル指標の推移 (FGM データセット：サブクラス)

FGM データ・セットにおける IPC の共起のうち、データセットの最終年月である 2015 年 12 月における最終的に最も頻度の高い組み合わせは、サブクラス、グループ、サブグループにおいてそれぞれ 161718 にしめすとおりである。

4.3.2.2 モデル

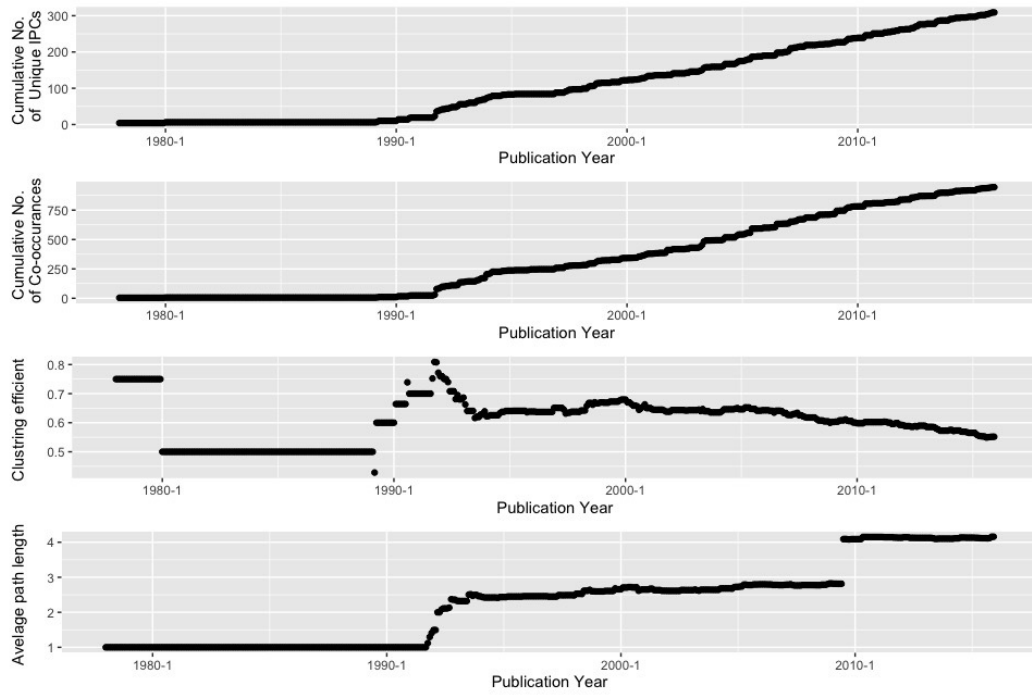


図 61. ネットワークにおけるグローバル指標の推移 (FGM データセット：グループ)

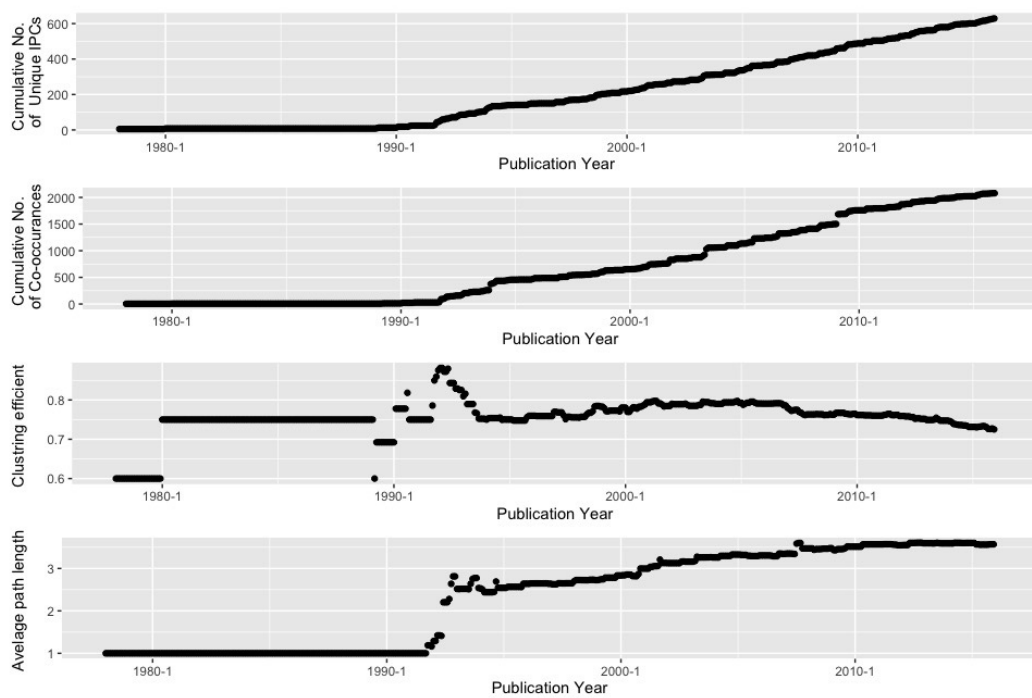


図 62. ネットワークにおけるグローバル指標の推移 (FGM データセット：サブグループ)

表 16. 上位 5 ペア (FGM データセット : サブクラス)

				共起数
A61K	医薬用, 歯科用又は化粧品用製剤	A61K	医薬用, 歯科用又は化粧品用製剤	154
A61K	医薬用, 歯科用又は化粧品用製剤	A61P	化合物 医薬組成物の治療活性	133
C04B	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス	C04B	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス	107
C22C	非鉄合金, すなわち鉄以外の金属を基とする合金	C22C	非鉄合金, すなわち鉄以外の金属を基とする合金	93
B22F	金属質粉の加工 金属質粉からの物品の製造 金属質粉の製造	C22C	非鉄合金, すなわち鉄以外の金属を基とする合金	83

表 17. 上位 5 ペア (FGM データセット : グループ)

				共起数
C04B0035	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス・・・組成に特徴を持つ成形セラミック製品	C04B0035	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス・・・組成に特徴を持つ成形セラミック製品	77
A61K0036	医薬用, 歯科用又は化粧品用製剤・・・藻類, 地衣類, 菌類もしくは植物又はそれらの派生物からの物質を含有する構造未知の医薬品製剤	A61K0036	医薬用, 歯科用又は化粧品用製剤・・・藻類, 地衣類, 菌類もしくは植物又はそれらの派生物からの物質を含有する構造未知の医薬品製剤	46
A61K0036	医薬用, 歯科用又は化粧品用製剤・・・藻類, 地衣類, 菌類もしくは植物又はそれらの派生物からの物質を含有する構造未知の医薬品製剤	A61P001	化合物 医薬組成物の治療活性・・・消化器官, 消化系統の疾患治療薬	32
B22F0003	金属質粉の加工 金属質粉からの物品の製造 金属質粉の製造・・・成形 焼結方法に特徴がある金属質粉からの工作物 物品の製造 特にそのために適した装置	C04B0035	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス・・・組成に特徴を持つ成形セラミック製品	25
B22F0003	金属質粉の加工 金属質粉からの物品の製造 金属質粉の製造・・・成形 焼結方法に特徴がある金属質粉からの工作物 物品の製造 特にそのために適した装置	B22F0007	金属質粉の加工 金属質粉からの物品の製造 金属質粉の製造・・・成形 焼結方法に特徴がある金属質粉からの工作物 物品の製造 特にそのために適した装置	24

表 18. 上位 5 ペア (FGM データセット : サブグループ)

				共起数
B22F000702	金属質粉の加工 金属質粉からの物品の製造 金属質粉の製造・・・成形を行い行わないで粉末を焼結することによって, 金属質粉から成る複合層, 複合作物 複合物品の製造・・・複合層の製造	C04B003564	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス・・・組成に特徴を持つ成形セラミック製品・・・焼成 焼結方法	6
B22D001302	金属の鋳造 同じ方法 装置による他の物質の鋳造・・・遠心鋳造 遠心力の利用による鋳造・・・縦軸のまわりに回転する鋳型中での細長い中空体, 例. パイプ, の製造	C22C002100	非鉄合金, すなわち鉄以外の金属を基とする合金・・・アルミニウム基合金	5
B22F000314	金属質粉の加工 金属質粉からの物品の製造 金属質粉の製造・・・成形 焼結方法に特徴がある金属質粉からの工作物 物品の製造 特にそのために適した装置・・・両者を同時に行なうもの	C22C000105	非鉄合金, すなわち鉄以外の金属を基とする合金・・・合金の製造・・・金属粉末と非金属粉末の混合物	5
C04B003562	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス・・・組成に特徴を持つ成形セラミック製品・・・製造方法 セラミック製品を製造するための無機化合物粉末の処理方法	C04B003564	石灰 マグネシア スラグ セメント その組成物, 例. モルタル, コンクリート 類似の建築材料 人造石 セラミックス・・・組成に特徴を持つ成形セラミック製品・・・焼成 焼結方法	5
H01J006136	電子管 放電ランプ・・・放電灯ガス 蒸気放電ランプ・・・うつわの部品間の封止 導入線の封止 導入線	H01K000138	白熱ランプ・・・細部・・・導入線の封止	5

4.3.2.3 モデル評価（分類）（上位層の予測に下位層の情報を用いる場合）

まず、FGM データセットにおいて上位層の予測に下位層の情報を用いた場合の結果を示す。予測対象となる上位層の IPC に対応する下位層の IPC は複数存在するため、一つの統計要約量に集約することによって特徴量を代表させた。本研究では中央値を用いた。FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 63 に示す。

FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 64 に示す。

fgm データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 65 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

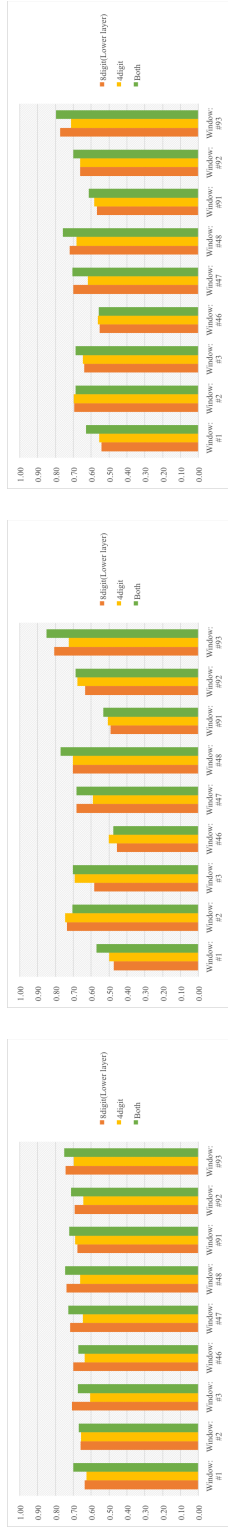


図 63. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [fgm データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

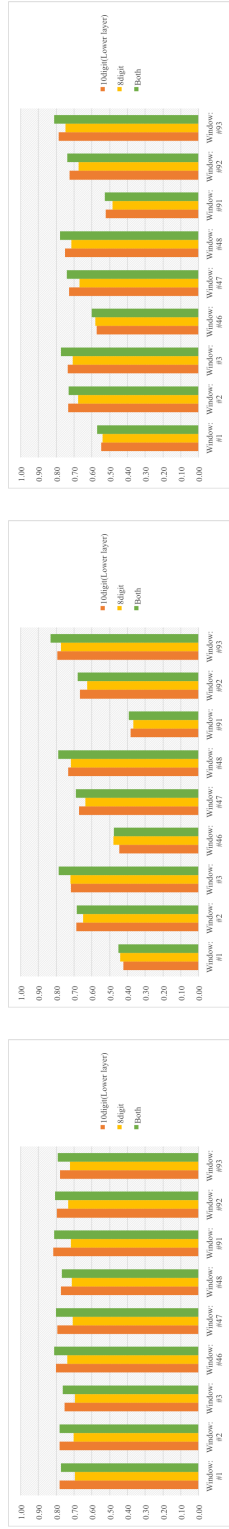


図 64. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

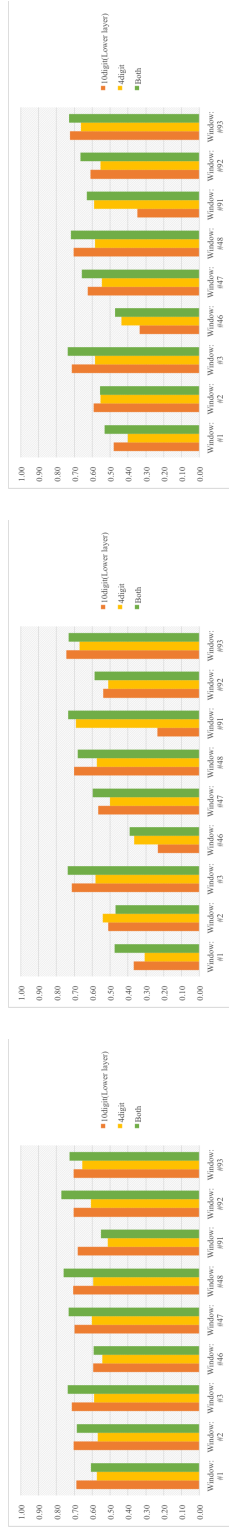


図 65. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

FGM データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 66 に示す。FGM データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 67 に示す。FGM データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 68 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。



図 66. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [fgm データセット全 IPC を対象]



図 67. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [fgm データセット全 IPC を対象]



図 68. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [fgm データセット全 IPC を対象]

4.3.2.4 モデル評価（分類）（下位層の予測に上位層の情報を用いる場合）

つづいて、FGM データセットにおいて、下位層の予測に上位層の情報を用いた場合の結果を示す。予測対象となる下位層の IPC に対応する上位層の IPC は 1 つしかないため、上位層の予測とことなり、統計代表値によって集約する必要はない。FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 69 に示す。

FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 70 に示す。

FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 71 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

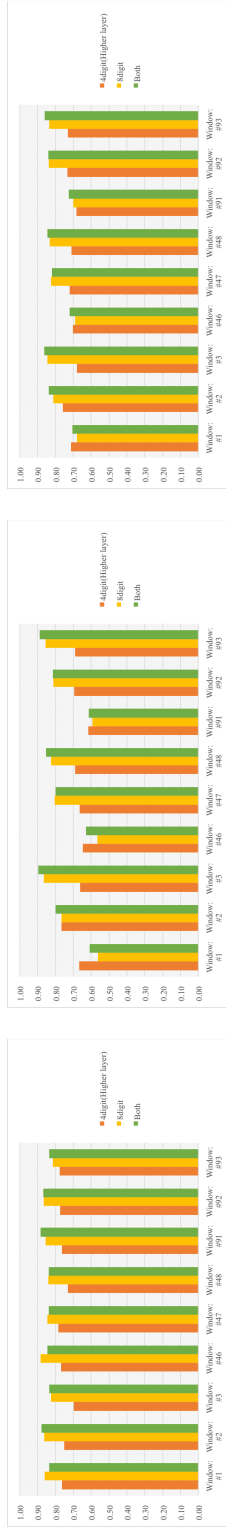


図 69. グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]



図 70. サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]



図 71. サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

FGM データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 72 に示す。

FGM データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 73 に示す。

FGM データセットにおいて、全ての IPC の組み合わせのみを対象とした予測において、サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の Precision(左図) , Recall (中図) , Fmeasure (右図) を図 74 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

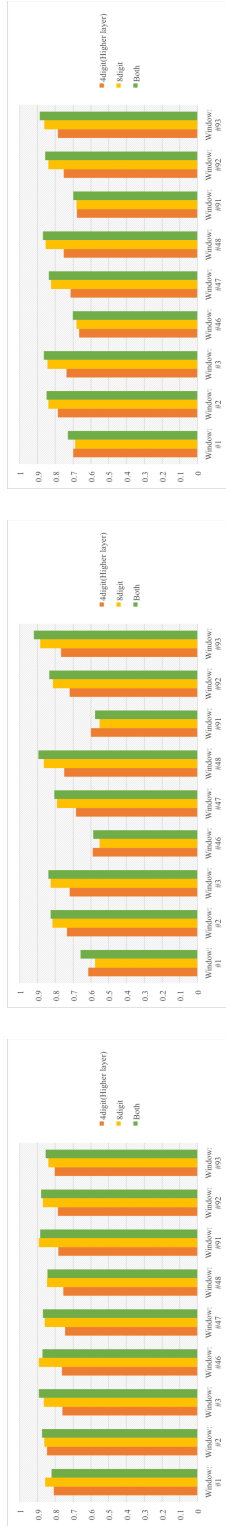


図 72. グループ (8digit) の予測に上層情報であるサブクラス (4digit) を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセット全 IPC の組み合わせ対象]



図 73. サブグループ (10digit) の予測に上層情報であるグループ (8digit) を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセット全 IPC の組み合わせ対象]



図 74. サブグループ (10digit) の予測に下層情報であるサブクラス (4digit) を用いた場合の Precision(左図), Recall (中図), Fmeasure (右図) [FGM データセット全 IPC の組み合わせ対象]

4.3.2.5 モデル評価（回帰）（上位層の予測に下位層の情報を用いる場合）

次に、FGM データ・セットにおける回帰分析結果を示す。回帰モデルの評価は前述の通り、RMSE と調整済み決定係数の2つである。まず、FGM データセットにおいて上位層の予測に下位層の情報を用いた場合の結果を示す。FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした調整済み決定係数 (左図) ,RMSE (右図) を図 75 に示す。

FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 76 に示す。FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 77 に示す。これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

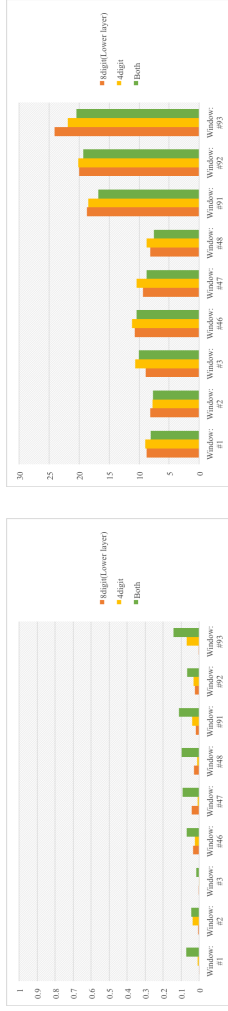


図 75. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

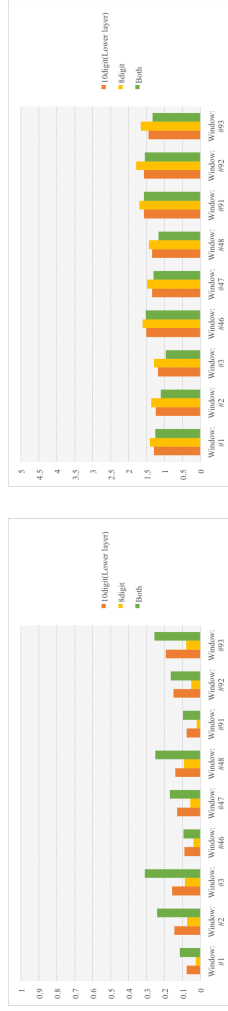


図 76. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [fgm データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

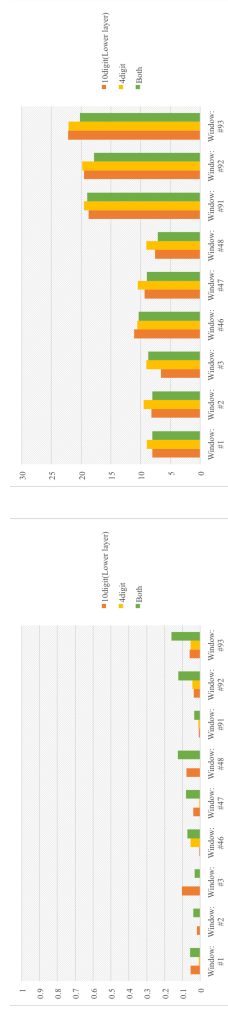


図 77. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

同様に、FGM データ・セットに対して、全ての IPC を対象にした結果を示す。FGM データセットにおいて上位層の予測に下位層の情報を用いた場合の結果を示す。FGM データセットにおいて、全ての IPC の組み合わせを対象とした調整済み決定係数 (左図) ,RMSE (右図) を図 75 に示す。

FGM データセットにおいて、全ての IPC の組み合わせを対象とした予測において、グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 76 に示す。

FGM データセットにおいて、全ての IPC の組み合わせを対象とした予測において、サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 77 に示す。

これらの図における棒グラフはいずれも左から、下層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

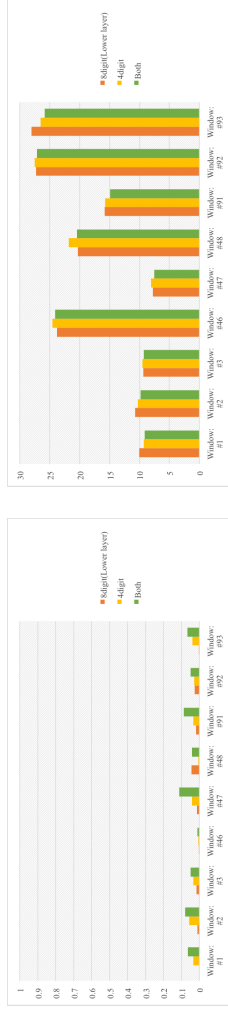


図 78. サブクラス (4digit) の予測にグループ (8digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]

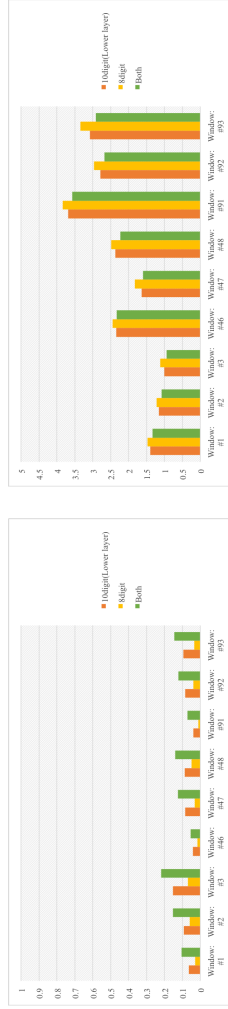


図 79. グループ (8digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]

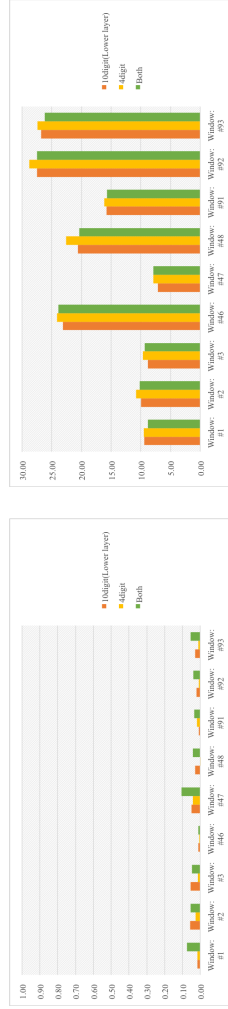


図 80. サブクラス (4digit) の予測にサブグループ (10digit) の下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]

4.3.2.6 モデル評価（回帰）（下位層の予測に上位層の情報を用いる場合）

次に、FGM データセットにおいて下位層の予測に上位層の情報を用いた場合の結果を示す。FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 81 に示す。

FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、グループ (8digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 82 に示す。

FGM データセットにおいて、クラスが異なる IPC の組み合わせのみを対象とした予測において、サブクラス (4digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 83 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

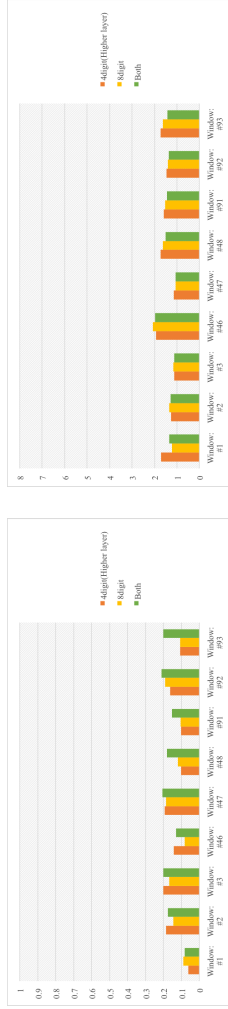


図 81. グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

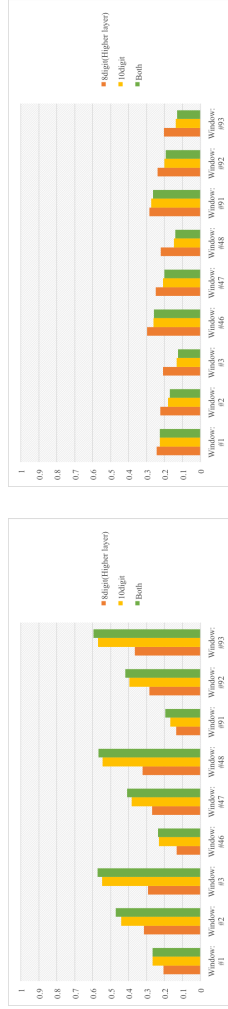


図 82. サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

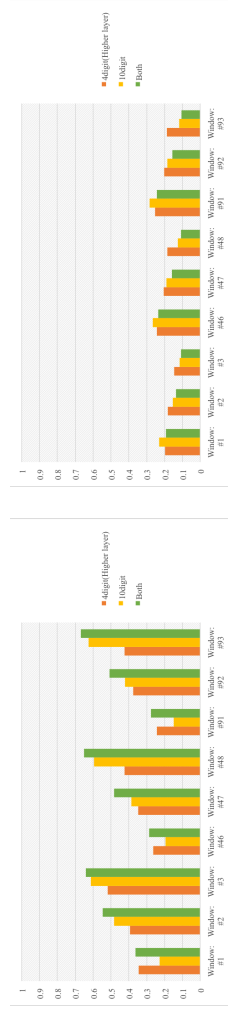


図 83. サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいてクラスが異なる IPC の組み合わせのみ対象]

同様に、FGM データ・セットに対して、全ての IPC を対象にした結果を示す。FGM データセットにおいて、全ての IPC の組み合わせを対象とした予測において、グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 84 に示す。

FGM データセットにおいて、全ての IPC の組み合わせを対象とした予測において、グループ (8digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 85 に示す。

FGM データセットにおいて、全ての IPC の組み合わせを対象とした予測において、サブクラス (4digit) の予測に上位情報であるサブグループ (10digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) を図 86 に示す。

これらの図における棒グラフはいずれも左から、上層のみの特徴量による精度、予測対象層のみの特徴量による精度、両方の層の特徴量を用いた精度としてわけている。色別としては橙、黄、緑で表される。

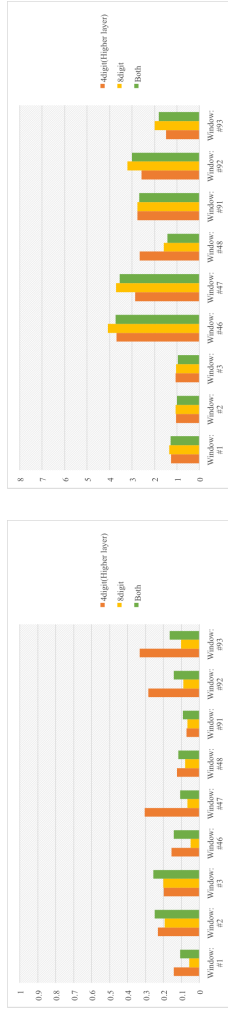


図 84. グループ (8digit) の予測に上位情報であるサブクラス (4digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]

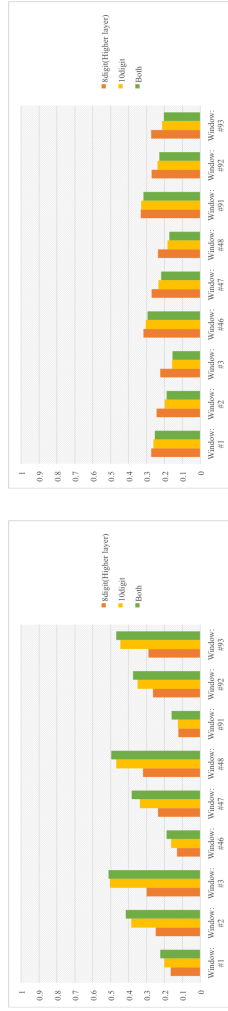


図 85. サブグループ (10digit) の予測に上位情報であるグループ (8digit) を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]

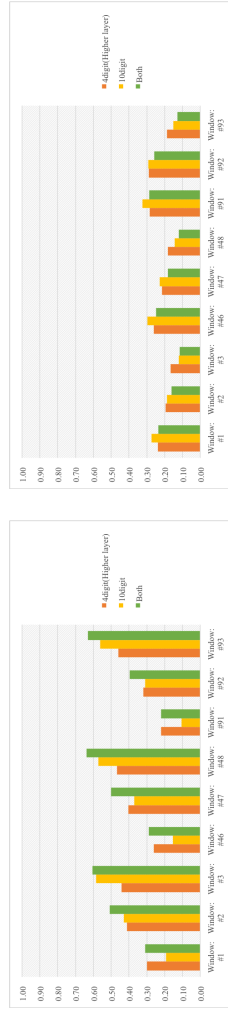


図 86. サブグループ (10digit) の予測に上位情報であるサブクラス (4digit) 下層情報を用いた場合の調整済み決定係数 (左図) ,RMSE (右図) [FGM データセットにおいて全ての IPC の組み合わせを対象]

4.4 考察

まず、CFRP を対象とした結果について次のように議論する。図 36 より、サブクラス (4digit) のレイヤーの予測に粒度の細かいレイヤーであるグループ (8digit) の情報を共に用いることによって、Fmeasure の上昇に寄与していることがわかる (右図棒グラフ緑色)。時間窓によっては、グループ (8digit) の情報のみでの予測だけでもパフォーマンスの向上に寄与している (棒グラフオレンジ色)。Precision だけで見た場合、いずれの時間窓においてもグループ (8digit) のみの情報での予測はサブクラス (4digit) のみでの予測よりも高いことから、Fmeasure の時間窓によって異なる傾向は、Recall による影響であると考えられる。図 37 は、グループ (8digit) の予測にサブグループ (10digit) を用いた場合である。図 36 と比較して、時間窓による傾向は小さく、いずれの時間窓においても両方の特徴量を用いた場合 (緑色)、粒度の細かい情報を用いた場合 (オレンジ色)、予測対象レイヤーのみ (黄色) で予測した場合、の順で精度が良い。サブクラスとグループの関係よりもグループとサブグループの関係のほうが技術粒度の距離は近い。図 38 からその傾向とは見て取れる。図 38 はサブクラス (4digit) の予測にサブグループ (10digit) の情報を用いることの有効性を示しており、いずれにおいても Fmeasure に貢献している事がわかる。これらの傾向については、図 39 から図 41 から全 IPC を対象とした予測においても同様であることが確認できる。一方で、下位層の予測に上位層の情報を用いた場合の結果が図 42 から図 44 に示される。いずれのケースにおいても、隣接レイヤーのみによる予測が予測対象レイヤーのみの予測を上回ることではないことが確認できる。これは、上位の情報だけでは、下位情報以上の情報量を上回ることが出来ないことから理解できる結果といえる。

同様に CFRP データセットに対して回帰予測を行った結果について議論する。上位層の予測に下位の情報を用いたことによる結果を示した図が図 48 から図 50 である。調整済み決定係数の上昇および RMSE (最小二乗誤差) の低下に寄与していることが分かる。図 54 から図 56 は、下位層の予測に上位層の情報を用いた場合の回帰予測結果を表したものである、隣接レイヤーの情報が精度に寄与していることはいずれのケースからも見て取れるが、ここでは図 5457 に注目する。図 54 はサブグループの予測にサブクラスを用いた場合の結果である。多くの時間窓のケースにおいて、調整済み決定係数が 0.7 を超えるなどしており、この組み合わせについては、実利用としても意義のある特徴量

の組み合わせと考えることができる。

4.4.1 CFRP の技術融合に関する具体例

CFRP の技術融合として特定することの出来た具体例について議論する。回帰モデルのなかで最も高い精度 (調整済み決定係数=0.7334) であった条件は、サブグループの予測にサブクラスを用いた場合において、学習期間 4 年 (2009-2013)、テスト期間 1 年 (2014-2015) の学習／予測窓の場合であった。このモデルの中で最も高い予測値であったサブグループの組み合わせは、C08G006942 と D02G000344 の組み合わせであった。このふたつの IPC はそれぞれ C08G006942 「高分子化合物の組成物・・・高分子の主鎖にカルボン酸アミド連結基を形成する反応により得られる高分子化合物・・・ポリエステルアミド」、D02G000344 「繊維, フィラメント, 糸より糸の巻縮処理 糸より糸・・・糸より糸, 例. 飾り糸 他の類に分類されない糸より糸を製造するための方法および装置・・・使用目的に特徴がある糸より糸」である。データ・セット中におけるこのペアが存在する特許公報の一覧を表 19 に挙げる。

表 19. CFRP の技術融合に関する具体例 (サブクラス)

公開公報番号	発明の名称	公開公報日	出願人
CN104611806A	Flame retardant yarn comprising aromatic polymer fiber containing sulfuryl, fabric made of same and preparation method — Flame retardant yarn and fabric and preparation method of prepared aromatic polymer fiber comprises containing sulfone	2015/2/13	Shanghai Tanlon Fiber Co. Ltd.
CN104630957A	Yarn and fabric made of polyphenylene sulfide fibers and blended fibers based on aromatic polyamide and Astrel 360 and production method of yarn and fabric — Yarn and fabric prepared by blending fiber and polyphenylene sulfide fiber based on aromatic polyamide and polyarylene sulfone and preparation method thereof	2015/2/13	Shanghai Tanlon Fiber Co. Ltd.
CN104611804A	Yarn and fabric made by sulfonyl-containing aromatic polymer fibers and high-temperature-resistant flame-retardant fibers and preparation methods — Aromatic polymer fiber comprising yarn fabric and preparation method and high temperature resistance inflaming retarding fiber and containing sulfonyl and preparation method	2015/2/13	Shanghai Tanlon Fiber Co. Ltd.

4.4.2 FGM の技術融合に関する具体例

次に、FGM を対象とした結果について図 63 から図 65 を元に考察する。これらの結果から、CFRP データセットよりも隣接階層レイヤーの情報の貢献が大きいことが見て取れる。また、CFRP データセットと比較して、Recall の時間窓の差異が大きい。Recall が高いということは予測対象の取りこぼし（実際にはリンクが生じたにも関わらず事前に特定できていない対象）が少ないことを意味するが、これはひとつには学習するデータの量によると言える。特許数としては CFRP データセットと比較して半数程度である。CFRP と同様に、サブグループ（10digit）の予測に上位情報であるサブクラス（4digit）を用いた場合の結果が FGM データセットを対象とした回帰予測においても最も優れた結果となった。

FGM の技術融合として特定することの出来た具体例について議論する。サブグループに注目すると、最も多い共起数であり比較的高い予測値として挙げることの出来た IPC のペアとして A61Q001908 と B32B001509 が挙げられる。このふたつの IPC はそれぞれ A61Q001908 「化粧品または類似化粧品製剤の特殊な使用_スキンケア剤_老化防止剤」、B32B001509 「平面以外の一般的形状を本質的に有する積層体本質的に金属からなる積層体_ポリエステルからなるもの」である。データ・セット中におけるこのペアが存在する特許公報の一覧を表 20 に挙げる。

表 20. FGM の技術融合に関する具体例（サブグループ）

公開公報番号	発明の名称	公開公報日	出願人
CN104146873A	Multifunctional eye mask capable of adjusting temperature spontaneously — A multifunctional self-regulate eye mask	2014/11/19	CHEN Lei,CN
CN204033812U	Multifunctional temperature self-regulating eye mask — A multifunctional self-regulate eye mask	2014/12/24	CHEN Lei,CN

本分析の結果から、技術融合予測を分類としてみた場合、相対的下位のレイヤーの情報を用いることは相対的上位のレイヤーの予測に有効であるという点、また、技術融合予測を回帰として見た場合、相対的上位のレイヤーの情報を用いることは相対的下位のレイヤーの予測に有効といえる。特許庁審査官に対し IPC の付与プロセスに関するヒアリングを行った所、IPC の付け方は主要部署から関連すると思われる部署に回していくというコメントが得られた。担当審査官が関連すると思われる部署に回し、想定されない部署には回らない。すな

わち、審査官の想定する範囲内でのみ技術的関連性が閉じており、その外での関連性の発想に至らないことを意味する。しかもこの傾向は上位 1 文字が異なる IPC の融合を予測するスピノフ予測においてより顕著である。本研究の結果は、審査官は今まで一度も発生しなかったような組み合わせを突然思いつくことが難しいことが現象として現れたものと考ええる。

4.5 まとめ

本研究では、技術融合の中からとくに異分野への展開に注目した技術的スピノフの予測に焦点をあてたものである。具体的な事例として炭素繊維強化プラスチックの関連特許 19,319 件、傾斜機能材料の関連特許 10,953 件を対象とした。はじめにそれぞれのデータセットに含まれる国際特許分類の共起情報に基づき月ごとの時間拡張ネットワークを作成し、学習窓に出現した IPC の全てのペアに対して重み付きリンク予測指標を算出し特徴量とした。その上で、学習期間には共起しなかったがテスト期間には共起が出現した IPC 同士の予測を行う分類モデルならびに回帰モデルを構築した。特にこれまで議論されなかった IPC の階層情報についても盛り込み、階層情報が、IPC 間のリンク予測に寄与することを示した。特に情報量の多い下位の情報は分類モデルに有効であり、情報粒度は粗いがトップダウンの情報を有している上位情報は回帰モデルが有効であることを示した。これらの IPC の付与のあり方が現れている可能性についても指摘した。技術融合の観点から、具体的に予測された IPC のペアのうち幾つかに注目した所、CFRP に類する技術が衣類に用いられるという動向や、機能性材料に類する技術がアイウェアに用いられつつあるという動向を特定することが出来た。これらの材料は、安価に調達でき製造技術もコモディティ化しつつあることからより身近な製品に用いられつつあることがわかる。本研究の成果を用いることで、特定企業が有する主軸の技術を転用可能な異分野といった技術領域の抽出に有効な手法であると考ええる。

5 第5章：サプライチェーンにおけるNグラム情報を活用した潜在的取引相手の予測

5.1 背景と目的

5.1.1 企業間関係メカニズム

企業活動において、どういった企業を顧客とするか、またどの企業から資材を調達するかといった意思決定は自社の方向性を形付ける重要な要素である。新たな顧客の創造は、すべての企業において常に命題である。多くの企業が、自社は積極的に営業開拓のための情報収集を行っていると考えている。しかし、現在おこなっている営業活動の範囲外に、自社の価値を見出してくれる潜在的な企業が存在する可能性もある。消費者ニーズの多様化に伴い、異なる業界への展開によって新たな価値創造を可能とする例も少なくない。このような潜在的取引関係を見つけることは、従来型の自社の業種に閉じた情報のみでは限界がある。現在の取引相手よりもよい相手が見つかるのであれば、すぐにでも取引をスイッチするという判断は経済合理性の観点で正しい。しかしながら、実際にはそう頻繁に取引相手をスイッチすることは起きない。合理的な取引とならない要因の一つにはサプライヤーとカスタマーの間に生じる互恵的な協力関係、すなわち「系列」と呼ばれる概念によるところが大きい [196]。Hsu(1999)によれば系列は、固定した複数のサプライヤーが継続的に中間財を供給する企業間取引の構造といえる [197]。特に、日本においてはこの「系列」と呼ばれる概念が占める企業間取引の構造におけるパワーバランスの影響は強い。欧米のオープンシステム型との取引構造と比較して、強みの源泉と言われている [198, 199]。系列に関する組織間関係論においては、幾つかの研究がなされている。特に、取引コストパースペクティブ、資源依存パースペクティブ、制度化パースペクティブといった観点における議論が活発である [200, 201, 202, 203, 204]。取引コストパースペクティブとは、組織間関係の形成メカニズムの要因にコストの観点でアプローチするものである。この観点では、取引に関するコストの総和が低い場合に、資源調達を目的として取引関係が形成されることを意味する。取引コストパースペクティブとは、その名の通り取引コストの観点から組織間形成のメカニズムにアプローチするものである。秋山(2014)によれば、取引コストには、取引相手を探すコスト、正当な契約を行うためのコスト、正しい契約を履行させるための監視コストなど、取引に関わるあらゆるコストが含まれ

る [196]。この取引コストが低い場合、企業は資源調達を目的とした取引関係の形成の意思決定を行うのである。逆に、このような取引コストが高い場合、企業は資源を内製化することが合理的である。すなわち取引コストのコントロールは、組織の境界線を決定することそのものといえる [200, 201]。組織間関係の形成をパワーバランスといった視点でとらえるのが資源依存パースペクティブである [202, 205]。多くの組織はそれぞれ資源を外部から獲得する必要があるが、強く依存しすぎてしまえば自らの存在意義や主導権を失う。例えばサプライヤーとカスタマーの生産スキルの相対的な違いがパワーバランスを決定するということは珍しくないであろう [206]。企業取引の意思決定はこのように複雑な組織間関係によってバランスされているため、単純な経済合理性のみでは説明が出来ないことがわかる。また、現実の企業間取引では、人間の意思決定において起こりやすいバイアスの影響により、必ずしも合理的な相手を選択できているわけではない。例えば、自社の製品を製造するために必要な加工技術を持っているある企業を探しているとする。もし近くに当該技術を有する企業が存在することを知らなければ、利用可能性ヒューリスティックのせいで、遠方の有名企業との取引を行わざるを得なくなる。十分な情報コストを支払うことが出来ずに起きるこの事象は、短期的には取引コストを小さく済ませることができが、長期的全体的には大きなコストを支払わなければならない場合がある。特にこのような、本来存在するべき近距離交流がおろそかであると、地域全体の経済性において機会損失ともなりうる [207, 208]。

5.1.2 潜在的なつながりの創造

企業は、販路拡大や将来的な取引断絶に備えた情報収集を常に行わなければならない。今や、単一の技術で構成される製品は稀であり、多くの企業は他業種との協業によってあらたなイノベーションを創出することが求められている [156]。現在の多くの企業はオープンイノベーションに挑戦しなければならないが、これまで自らの業種だけに依存してきた彼らは他業種の情報を得ることに苦労している。また地方自治体は、自らが受け持つ経済圏に所属する企業に対して、適切な情報提供が義務づけられる。しかしながら、昨今の情報量の爆発的増加において、情報の海に溺れてしまうことによって取引コストが増大しつつあるのが実情である。一方で計算機の高性能化、データの開示、分析技術の高度化によって情報の処理がより低コストで可能になりつつある。情報技術によって取引コストを下げることに寄与できれば、企業はより合理的な意思決定

を行うことができる。特に、現時点で取引関係が存在しないが将来的に取引関係が起りうる取引相手を特定することができることは多くの面で望ましい。このことは、企業にとって新たな販路拡大の情報源につながるだけでなく、将来的に現在の取引関係が失われた際のリスクヘッジともなる。このような背景にもとづき、繋がりという観点で企業取引関係に着目する研究が多い。規模の大きい企業や、売上を上げている企業だけではなく、つながる能力のある企業こそが当該地域において重要であるという視点が着目されている。このような潜在的な取引の予測・推薦を情報学的手法によって行うために、サプライチェーンに対して統計的機械学習を適用した手法が提案されている [209, 210, 211]。このような新たな繋がり提案の仕組みは、企業にとっての販路拡大情報となることはもちろん、自治体にとっても経済圏のマネジメントを容易にし、適切な情報提供サービスを可能とする。その観点で、企業取引の予測による推薦情報は地域活性化に資するといえる。

5.1.3 リニアサプライチェーンとネットワークサプライチェーン

サプライチェーンの研究は、伝統的にはリニア構造によるモデルによって分析が行われてきた [212]。線形に表現される伝統的なサプライチェーンモデルは、企業間の資材の流れを1方向に表すことに特徴を持つモデルである。しかしながら、このような線形モデルはサプライチェーンの実態を極端に単純化していることになり、物資と情報の流れの複雑な関係を十分に表現することは難しい [213]。そのような中、サプライチェーンの複雑性をより適切に説明するための手法として、企業間の物資と情報の流れをネットワークとして捉える分析手法が用いられつつある。特に Social Network Analysis (SNA) が用いられつつある。数学におけるグラフ分野に端を発した SNA は、複雑な社会をモデル化するために適しており、かつての線形モデルによるサプライチェーン分析よりも適切に実態を表すことができる [214]。SNA を用いる分析では、企業をノードとしてとらえ、企業間の関係性 (すなわち取引関係) をエッジとした系をネットワークと呼称する。特に、ネットワーク中心性を算出することで各企業のトポロジカルなネットワーク的位置付けを定量化することが可能な手法である。ネットワーク中心性に着目することで、地域ネットワークにおいて重要な位置付けを担うような中小企業にとって重要な視点が得られる [215, 216]。企業間取引におけるネットワーク中心性は、企業活動の成長やイノベーション創発力についても一定の関係性が有ることが分かっている。 [217, 218, 219, 220]

企業間取引の関係性をネットワークとして捉えることは、組織間関係の形成メカニズムの観点においても意義がある。例えば、次数中心性や媒介中心性は資源依存やパワーを表し、媒介中心性や構造的空隙は自立性と負の相関にある。このような数値を計算することによって、取引関係を、複数の視点によって定量的に評価することができる [196]。このような背景から、ネットワークの観点を用いて潜在的な企業取引の予測を行う研究が存在する。たとえば、Web から企業公式サイト情報をクロールし、当該企業を特徴づける技術的用語などの類似度を算出することで類似企業との潜在的取引を推薦した研究や、ネットワーク中心性を特徴として用い地域の企業取引の構造に注目することで、潜在的な取引の推薦を行う手法が提案されている [211]。しかしながら、いずれの研究もサプライチェーンの本質的要素である物資と情報の流れを十分に考慮されておらず、実務上においては実態に即した改善の余地がある。

本研究では、予測のパフォーマンスと解釈容易性を両立する手法として、ネットワークモデルとリニアモデルの双方の利点を併せ持った潜在的取引推薦手法を提案する。潜在的取引先を推薦された時になぜその企業を推薦されたのという事を容易に解釈できるようなモデルであるということは、経営の意思決定を人間が行う以上、不可欠な要素であると考え。業種の類似性を算出することで類似企業を推薦する手法では、競合他社が推薦されてしまうことからサプライチェーンの流れを予測するという本研究の目的としては議題が残る。また、ネットワーク中心性のみを特徴量としたモデルでは、何故その推薦された企業が自社にとって候補となったのかといった解釈容易性に課題が有る。

多くの分野で機械学習技術による予測モデルが提案される中、意思決定支援のための予測モデルがもつべき解釈可能性についての言及は少なくない。例えば Lakkaraju, et al. (2016) は、予測モデルを構築する際の最も重要な障害の 1 つとして、意思決定をする人間による理解・信頼に至らないという点を挙げている [221]。企業の戦略を左右する取引先の判断に於いては特に、予測モデルはあくまで人間の意思決定を支援をするツールである。その観点から、本研究では、サプライチェーンの文脈に基づき解釈容易性を有する特徴量を投入することを提案する。またその特徴量の定義上、少なからず予測に寄与するであろうと仮説づけるものである。具体的には、上流企業から下流企業へのモノや情報の流れが、潜在的取引予測に強く寄与する特徴量であると考えた。モノの流れはリニアモデルサプライチェーンの本質的な情報である。本研究の特徴は、近年主流となっているネットワークモデルに、リニアモデルが持つ物流情報を付

加することで、精度と解釈容易性を両立することにある。

5.2 手法

本研究の主な着想は、これまで十分に用いられなかったモノの流れの情報を、潜在的取引関係予測の特徴量として用いることにある。具体的には、現実にはサプライチェーンを構成している連続した3企業の主業の情報をを用いる。上流二社の主業情報が与えられたとき、3社目の企業の主業の条件付き確率を算出することができる。例えば、企業A、企業B、企業Cがそれぞれ主業 α 、主業 β 、主業 γ を有しているサプライチェーンを構成するとする。

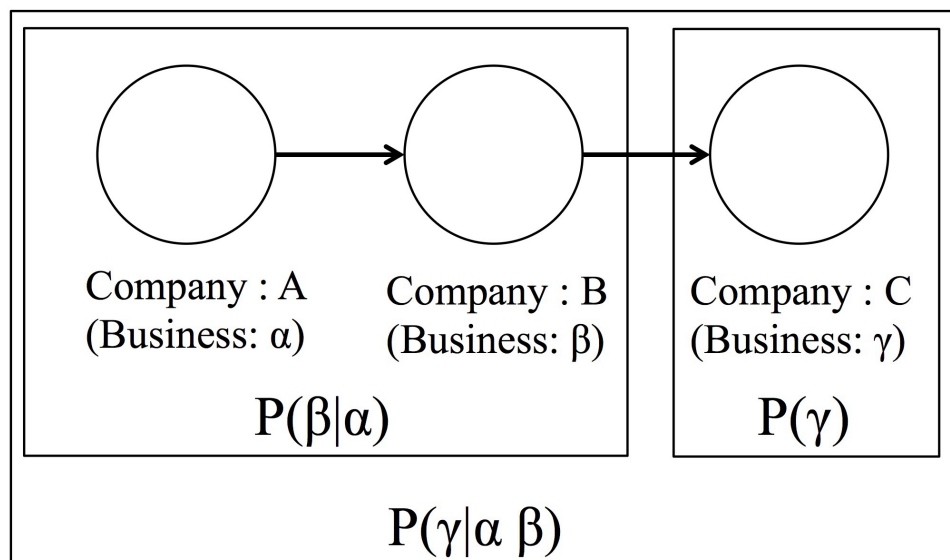


図 87. 主業の条件付き確率

この時、主業 α の企業と取引を有する主業 β の企業Bの顧客として、サプライチェーンにおいて主業 γ の企業と取引関係を有する条件付き確率 $P(\gamma|\alpha\beta)$ は

$$P(\gamma|\alpha\beta) = \frac{P(\alpha\beta|\gamma)}{P(\alpha\beta)} \quad (12)$$

で与えられる。同様に、主業 α を有する企業の次に主業 β が顧客となりうる条件付き確率 $P(\beta|\alpha)$ は次の式で与えられる。

$$P(\beta|\alpha) = \frac{P(\alpha\beta|\beta)}{P(\alpha)} \quad (13)$$

本研究ではモノの流れの確率的な表現を行うため、自然言語処理の分野で用いられる基本的な確率的言語モデルの一つである N-gram と呼称するモデルを

サプライチェーンに適用する。N-gram 言語モデルは、各単語の生起確率は、直前の (N-1) 単語までに依存することを前提とするモデルである。例えば、単語列が与えられた時その単語列の生起確率は、次で与えることができる。

$$P(w_n) = P(w_2|w_1)P(w_3|w_1w_2)\dots P(w_n|w_1\dots w_{n-1}) = \prod_{i=1}^n P(w_i|w_1\dots w_{i-1}) \quad (14)$$

例えば、“This is a cat” という文脈が与えられた時の生起確率は、下記式で計算することができる。

$$P(\text{This, is, a, cat}) = P(\text{This})P(\text{is}|\text{This})P(\text{a}|\text{This, is})P(\text{cat}|\text{This, is, a}) \quad (15)$$

この場合、これは 4-gram の言語モデルと呼ぶ。本研究では連続する 3 つの主要、すなわち 3-gram の条件付き確率を計算し各取引情報の特徴量にする。すなわち図 87 において下記式を算出し、予測モデルの特徴量とする。

$$P(\alpha\beta\gamma) = P(\alpha)P(\beta|\alpha)P(\gamma|\alpha\beta) \quad (16)$$

この確率が高いほど、そのビジネスの組み合わせの物流が起り得ることを意味する。例えば、P(金属加工業, 精密機器組み立て業, 自動車部品製造) は比較的高いことが想定できる。一方で、P(出版, 水産業, 文房具製造) は低いことが想定できる。後者のようなものの流れは起りにくい事を意味する。

5.2.1 データセット

本研究では、日本における東北地方の実際の企業間取引を記録したデータセットを用いる。東北地方は秋田県、青森県、岩手県、山形県、宮城県、福島県の 6 つの県から構成されている。特に、岩手県 宮城県 福島県の企業は、2011 年 3 月 11 日に発生した東日本大震災ならびに津波の被害によって直接的なダメージを受けた。近隣ならびに、東京を含む他の地域にも影響を及ぼす事となり、事業継続計画の必要性が議論されるきっかけともなった。本研究は、帝国データバンク株式会社が提供する企業取引データをデータ・ソースとして用いる。このデータ・ソースは、サプライヤ企業からカスタマ企業への物資の流れを表す関係を 1 つのレコードとして、エッジリストの形で格納されており、またそれぞれの企業に対し売上高、従業員数などの属性情報も記録されている。なお、企業は固有の ID で表され、具体的な企業名は匿名化されている。

5.2.2 モデル構築

潜在的取引先の予測モデルは、任意の企業間関係に対して T 年に取引は存在しなかったが T+1 年に取引が発生したという取引関係をネットワークからエッジとして抽出し、正例として学習させる教師あり学習の 2 クラス分類問題と定義する。正例のサンプルサイズと同規模の負例データをランダムに作成することで、バランスデータの構築をした。

さて、従来の企業属性単体（例えば売上や従業員数などの企業規模）を指標として取引先企業を探すという考え方は、Winner takes all を推奨することになる。ネットワーク分析の文脈では、次数の大きいノードが次数の大きいノードと隣接しやすいという傾向を正の次数相関 (Positive Assortativity) と呼ぶ [222] が、実際の企業取引ネットワークは負の次数相関 (Negative Assortativity) であることが知られている [223]。すなわち、持たざる者は持つものをつながろうとするパワーバランスが企業取引構造を支配すると考える。本研究では、サプライヤとカスタマがそれぞれ持つノード属性の比率をとりエッジに紐づく特徴量として定義する。

また、データセット中の企業に対して各種ネットワーク中心性を算出する。用いるネットワーク中心性は表 21 のとおりである。これらの中心性についても、企業属性同様に、サプライヤとカスタマが持つそれぞれのノード中心性の比率を取ることでエッジに紐づく特徴量として定義する。

次に、得られた企業取引ネットワークにおいて、それぞれの主業における 3-gram 情報を抽出する。自然言語における入力データは文章であるため、連続した 3 つの単語を抽出することは技術的に容易である。しかしながら、企業取引データは特定の 2 社間の取引関係の有無が格納されるため、3 つ以上の連続した node を直接的に入手することは難しい。本研究では、ネットワークから特定のサブネットワークを抽出する。ネットワークモチーフとは、ある条件を満たすサブネットワークである。例えば、3node のネットワークモチーフは図 88 で与えられる。本研究では図 88 におけるネットワークモチーフ ID:6 を抽出することで、連続した 3 社の主業情報を得る。得られた主業 3-gram 情報から、条件付き確率を計算し、エッジ属性として特徴量に盛り込む。

表 21. ネットワーク中心性

ネットワーク中心性	概要
次数中心性 [in/out]	隣接する辺すなわち取引数。次数が多い企業の経営決定と戦略行動が他社に与える影響は大きいと考えられている。[213]
近接中心性 [in/out][224]	ネットワーク中の全企業に対して何ステップで到達することができるかを表す。[225]
媒介中心性	任意の二社を結ぶ最短経路において当該企業が何回出現するかを表す。
固有ベクトル中心性 [226]	対象となる企業の取引先の重要性が強調される中心性 [227]
ページランク [228]	ウェブページにおいてユーザの関心を効果的に測定する手法として提案された中心性のひとつであり、頻繁に取引する企業との取引があるかどうかといった視点が強調される。
ボナチッチパワー中心性	他社の直接的な影響力と、他社が得た間接的な影響力の双方を考慮した再帰的に計算される中心性である。
クラスタ内次数係数 [229]	企業が自身のクラスタの他社とどの程度よくつながっているかを表す。
モジュール間分散度 [229]	対象企業が持つ取引が他のクラスタにどの程度分散しているかを表す。

5.2.3 モデルおよび評価基準

分類器にはランダムフォレストを使用する。ランダムフォレストは、複数の弱識別器である決定木を統合することにより汎化能力を向上させるアンサンブル機械学習アルゴリズムのひとつである。シンプルな学習方法ではあるが、一般的な決定木よりも優れた識別と予測性能を実現することが知られている。本研究では、特徴量の組み合わせに応じて、5つのモデルを提案する。Model Aは、企業属性のみを特徴量に用いたモデルである。Model Bは、ネットワーク中心性のみを特徴量に用いたモデルである。Model Cは本研究の主眼である3

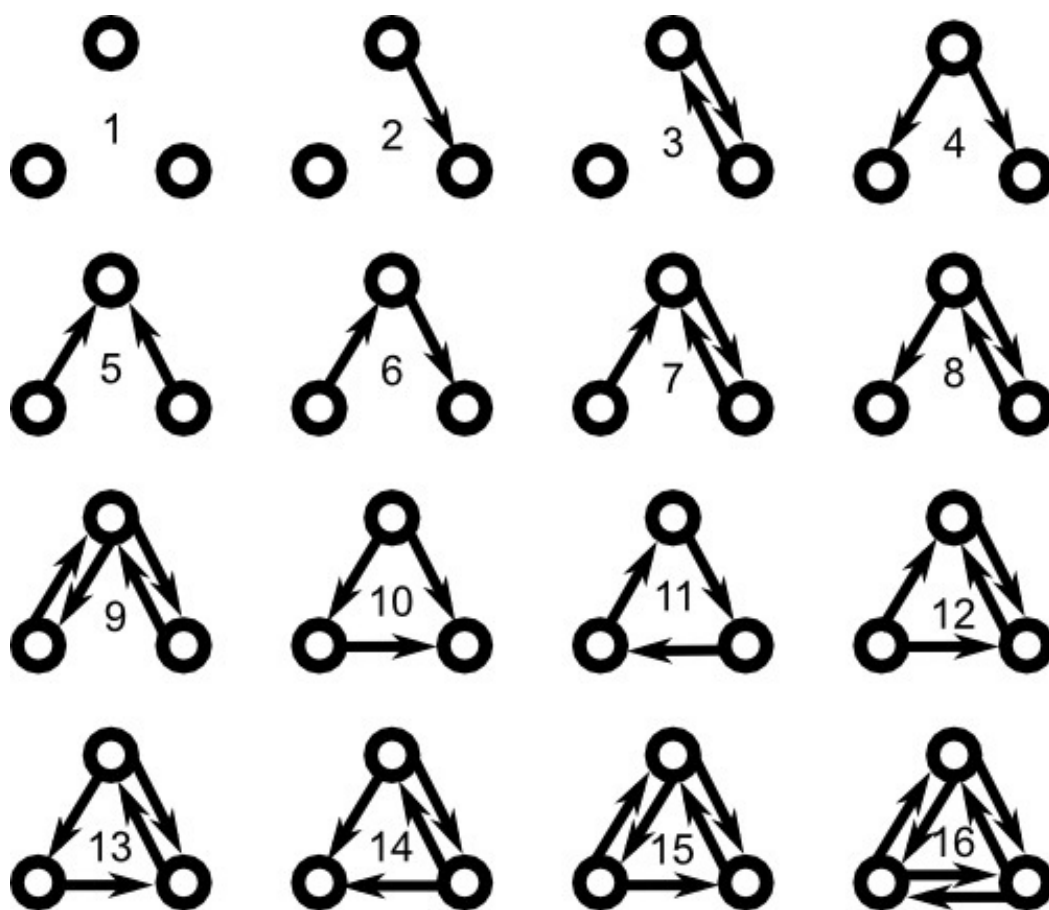


図 88. 3 ノードネットワークモチーフ (Source: Math Insight(2016))

グラム確率を特徴量としたモデルである。Model D は、企業属性及びネットワーク中心性を特徴量にしたモデルである。最後に Model E 企業属性、ネットワーク中心性、3 グラム確率を全て盛り込んだモデルである。我々はこれら 5 つのモデルを構築し、それらの予測パフォーマンスを比較する。モデルの検証には、Precision、Recall、F-measure を用いる。Precision は、潜在的取引があると予測したデータのうち、実際に翌年に取引が起こったものであるものの割合を表す。Recall は、実際に潜在的取引があるデータのうち、正しく予測されたものの割合をあらわす。F-measure は、Precision と Recall の調和平均によって得られる。全ての特徴量を用いたモデルである Model E について、いずれの特徴量が重要とされているか Impurity を算出する [230]。

5.3 結果

2010年から2014年において、東北6県のいずれかに本社の所在地を有する企業および、それらの企業と取引のある日本全国の企業全てをから抽出した結果、327,012取引(企業数131,192)を抽出した。本データにおける企業属性を表22に示す。また、各年における企業数と取引数を表23に示す。各データに紐付けられている主業情報のうち上位10に含まれるものを表24に示す。

表 22. 企業属性情報概要

企業属性	最小値	中央値	最大値
売上高 [JPY]	0	141	130.1M
利益 [JPY]	-289.4M	0	141.6M
従業員数 [人]	0	7	194,688
帝国データバンク評点	0	46	93
事務所数	0	2	24,204
工場数	0	1	117

表 23. 各年における企業数および取引数

	2010	2011	2012	2013	2014
取引数	291,808	312,864	313,762	321,780	327,012
企業数	128,890	132,616	132,075	133,926	135,257

このデータセットにおいて、年次 T 年には取引がなかったが、翌年 T+1 年に取引が初めて発生した企業の内、T 年に存在していた企業同士の取引を予測対象とする。たとえば、2010 年には取引が無かったにも関わらず 2011 年に初めて取引が発生した（かつ 2010 に存在はしていた企業同士による）関係は 53,622 取引であった。この 53,622 取引を正例とし、同量の負例の企業間関係をランダムに生成した結果、分析データセットは、107,243 取引(企業数 131,492)となった。各年次においてこのデータセットに含まれる正例と負例の概要を表 25 に示す。

取引データセットよりネットワークを構築し、企業ネットワークモチーフを

表 24. 主業上位 10

主業	Freq.
木造建築工事業	5569
土木工事業	4284
一般貨物自動車運送	3278
自動車一般整備	2469
電気配線工事業	1985
一般管工事業	1946
土木建築サービス	1804
内装工事業	1722
土工・コンクリ工事	1587
給排水・衛生工事業	1550

表 25. 各年における正例負例サンプルサイズ

Year T	2010	2011	2012	2013
Year T+1	2011	2012	2013	2014
正例	53,622	40,537	41,225	41,120
負例	53,621	40,536	41,222	41,117

抽出し、さらに主業ネットワークモチーフに変換した。これらの全ての 3 企業の組み合わせに対し、前述した 3-gram 確率を計算した。表 26 に、頻度が多い順に上位 10 組み合わせの 3-gram 確率を示す。

得られたデータセットに対して、表に対応する 5 種類のモデルを構築し、それぞれについて Precision, Recall, Fmeasure によって評価を行った結果を表 27 表 28 表 29 に示す。

また、それぞれのモデルにおいて特徴量の inpurity を算出した結果を図 89 に示す。

表 26. 出現頻度上位 10 組の 3GRAM の生起確率

主業 α	主業 β	主業 γ	頻度	$P(\alpha \beta \gamma)$
生鮮魚介卸	生鮮魚介卸	生鮮魚介卸	44,457	0.021
他の水産食料品製造	生鮮魚介卸	生鮮魚介卸	35,629	0.017
清酒製造	酒類卸	酒小売	13,679	0.007
自動車一般整備	自動車卸	一般貨物自動車運送	13,109	0.006
生鮮魚介卸	生鮮魚介卸	他の水産食料品製造	10,248	0.005
生鮮魚介卸	生鮮魚介卸	乾物卸	9,582	0.005
自動車車体整備	自動車卸	一般貨物自動車運送	8,650	0.004
医薬品製剤製造	医薬品卸	無床診療所	8,341	0.004
自動車部品付属品卸	自動車部品付属品卸	自動車一般整備	8,205	0.004
冷凍水産食品製造	生鮮魚介卸	生鮮魚介卸	8,029	0.004

表 27. 各モデルの PRECISION

	2010-2011	2011-2012	2012-2013	2013-2014
Model A	0.73	0.72	0.71	0.73
Model B	0.73	0.73	0.74	0.72
Model C	0.94	0.93	0.93	0.93
Model D	0.78	0.78	0.77	0.77
Model E	0.82	0.82	0.82	0.81

表 28. 各モデルの RECALL

	2010-2011	2011-2012	2012-2013	2013-2014
Model A	0.66	0.66	0.67	0.64
Model B	0.75	0.76	0.74	0.76
Model C	0.36	0.38	0.39	0.39
Model D	0.77	0.78	0.78	0.78
Model E	0.77	0.78	0.78	0.78

表 29. 各モデルの FMEASURE

	2010-2011	2011-2012	2012-2013	2013-2014
Model A	0.69	0.69	0.69	0.68
Model B	0.74	0.74	0.74	0.74
Model C	0.52	0.54	0.55	0.55
Model D	0.78	0.78	0.77	0.78
Model E	0.79	0.80	0.80	0.80

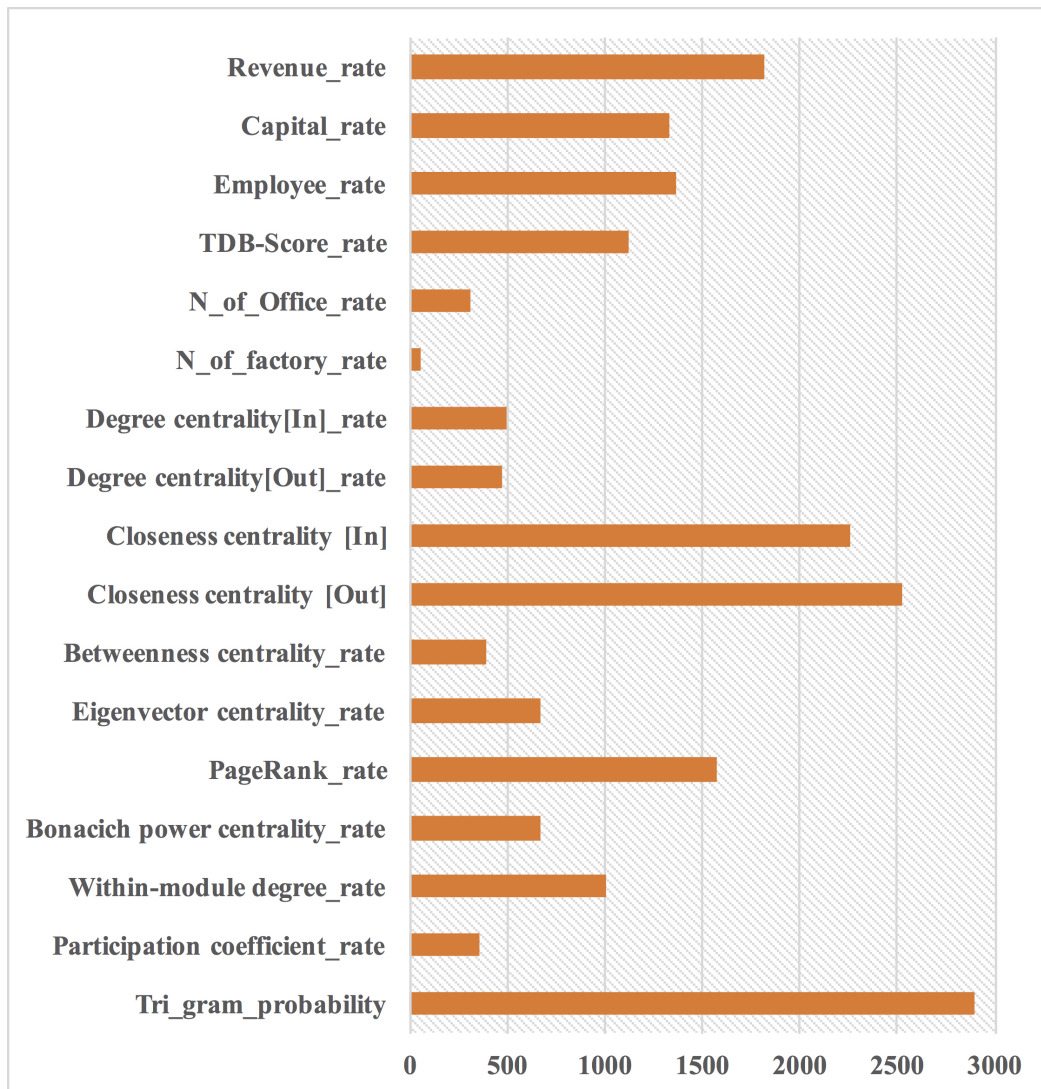


図 89. 各変数の Impurity

5.4 考察

表 23 より、年次に従って企業数取引件数ともに横ばい若しくは微増である傾向にあることがわかる。一方で、表 25 では、2011 年-2012 年のデータセットから、正例が減少していることが確認できる。正例は、T 年に存在する企業のうち T + 1 年に初めて発生した新規取引の数である。これは 2011 年に発生した東北大震災の影響によって新規取引の増分が減少したと考えられる。表 24 から、木造建築工事業や土木工事業といった建設業がこの地域の頻出業種である事がわかる。内閣府によれば、2011 年の東北大震災に起因する倒産割合の上位に建設業が挙げられるが、その理由のひとつに資材仕入が困難となったことが挙げられている。震災によって、これまで取引関係を有していたサプライヤやカスタマとの繋がりが分断してしまい新たな取引先を見つけることができなかつた為に倒産してしまったケースが多い事がわかる。震災復興工事業に加えて、東京オリンピックに向けた建設需要が高まっている。潜在的な取引を特定することは東北地方全体の経済にとって有意義である。表 26 より、この地域において最も頻度の高い連続した産業は、P(生鮮魚介卸, 生鮮魚介卸, 生鮮魚介卸) の組み合わせであることがわかる。東北地方沿岸は、日本海流と千島海流がぶつかって潮目をつくる地域であり、大量なプランクトンに多くの魚種が集まる有数の漁場であることが知られている。表 27, 表 28, 表 29 はそれぞれのモデルの Precision, Recall, Fmeasure を示す。Model A は企業属性比率のみを特徴量としたモデルである。Precision, Recall, Fmeasure 全体を通して企業属性比率のみであっても一定程度の予測精度が確認できる。厳密にはサプライヤとカスタマの企業属性の比率を特徴量としていることから、企業属性の差異がパワーバランスとして有効であることを意味する。Model B はネットワーク中心性のみを特徴量としたモデルであり、Precision は Model A と大きく変わらないかもしくはやや高い程度であるが、Recall については Model A と比較して 1 割以上優位なパフォーマンスを実現している。Fmeasure についても Model A よりも優秀であり、総合的に企業属性比率のみでのモデルよりもネットワーク中心性比率でのモデルのほうが効果が高いということが示された。これは既存研究と同様、ネットワークトポロジーが企業の取引予測に少なからず影響をすることを示唆している。Model C は 3gram 確率のみ特徴量としたモデルである。表 27 では、0.9 以上の Precision と高い精度で予測する事ができる特徴量であることが確認できる。連続する 3 つの業種の生起確率は対象とな

る地域のビジネスフローを表すものと考えることができることから、この特徴量が高い組み合わせは高い確度で取引可能性が起きるといった結果は妥当であるといえる。一方で、Recall が 0.4 を下回ることが確認できた。すなわち対象地域の典型的なビジネスフローではない流れについては、この特徴量ひとつだけでは限界がある。実際の企業取引において、Precision と Recall のいずれを重視するかを議論する場合、企業は予測された結果のうち実際に取引が起こるかどうかという点に重点を置くことから、Recall よりも Precision を重視であることが好ましい。3gram 確率は、確実な取引関係を予測する指標としては有効であると考えられる。Model D は企業属性比率とネットワーク中心性比率の双方を用いた場合の結果である。Model E は Model D に 3gram 確率の特徴量を盛り込んだモデルである。3gram 確率を投入することで Model E の Precision が向上していることが分かる。Recall については両モデルほぼ同等であることから、3gram 確率が Precision に貢献していることがわかる。図 89 は、Model E の各変数重要度を表したものである。3gram 確率が、最も変数重要度が高い事がわかる。次に重要度が高い変数に近接中心性比率が挙げられる。近接中心性は、あるノードから他のノードへの最短経路長の合計の逆数といえることから、他者に対する近さを指標化したものであるといえる。一般的に物流コストは距離とともに増加するため他社への近さは企業取引上単純に優位になる事を示す。前述の通り、ネットワーク中心性についても、サプライヤの中心性とカスタマの中心性の比率を特徴量としている。近接中心性も例外ではなく、近接中心性が高い企業と低い企業の比率が寄与している事を意味する。多くの企業とつながっているような近接中心性の高い企業同士ではなく、近接中心性の高い企業と近接中心性の低い企業の間で取引が起こりやすいということは、前述の通り企業取引における負の次数相関 (Negative Assortativity) が発生していることを意味する。これは Revenue_rate や Capital_rate も同様であり、企業のパワーバランスが特徴量として有効であることを示唆する。以上より、対象地域の典型的な物の流れ、企業のネットワーク的パワーバランスおよび企業規模のパワーバランスの差異が企業取引に有効に寄与していることが示された。

5.5 まとめ

企業活動においてのみならず、地方自治体にとっても、潜在的な取引の可能性を模索する活動は必要不可欠である。実際の取引関係は、必ずしもつねに合理的な結果によって生じたものではないことは、取引コストパースペクティブ

を始めとする組織間関係形成の観点からも説明できる。潜在的な取引を特定し推薦する仕組みは、企業の販路開拓、新たなイノベーションの機会創出、事業継続リスクヘッジ、地方自治体にとっての地域経済発展に資する政策投入の手段として注目されている。本研究では、機械学習的手法を適用することで、情報処理コストを低減し、合理的かつ新規な潜在的取引先の予測手法が可能であると考えた。近年のサプライチェーン分析では、ネットワーク的視点が主流であるが、本研究ではネットワークのみならず、伝統的な線形モデルの良い面も盛り込んだハイブリッドなモデルを目指した。これにより、複雑な取引関係の構造を考慮することができるだけでなく、人間による意思決定に必要な解釈容易性を両立した予測結果を得られると考えた。

具体的には、3つの連続する主業の条件付き確率をもとにした 3-gram 確率を特徴量とした。資材の流れが条件付き確率的に起こるという事象の定量化は、自然言語処理分野において伝統的に用いられる確率的言語処理モデルから着想を得たものである。3-gram 確率による特徴量は、起こり得ない資材の流れの組み合わせは低い数字を返す特徴量と解釈することができる。これまでのネットワークモデルでは、物資の流れについては十分に議論してこなかった。物の流れといった企業取引の本質的な情報を条件付き確率として考慮することで、現実的に起こりうる企業取引の推薦をより高い確度で実現できることを意味する。本研究において提案した手法は活用意義のあるモデルであることが確認出来た。内閣府は 2015 年より、Regional Economy (and) Society Analyzing System (RESAS) と呼ばれるシステムの運用を開始した。このシステムは、地方創生の実現に向けて地方自治体が客観的なデータに基づき地域の現状や課題を把握することを目的としており、本研究で用いた帝国データバンクが提供する企業取引情報も格納されている。API の公開や、ハッカソンイベントが行われるなど、民間を巻き込んでひとつの官民協業の形を形成し有る。API の公開や、ハッカソンイベントが行われるなど、民間を巻き込んでひとつの官民協業を形成している。2016 年 12 月には、官民データ活用推進基本法が制定されたこともあり、一層データのオープン化が進み、官民協業であらたな情報技術の活用のあり方が模索されることが考えられる。このように企業経営に資する情報として機械学習やビッグデータの活用が一層期待されるなか、本研究が提案するような手法を用いることが、潜在的な取引相手との付き合いきっかけとなったり、新たなイノベーションの創出に繋がるものとする。

6 第6章：結論

6.1 科学技術イノベーションにおける予測活動

本論文では、科学技術イノベーションのプロセスにおける科学、技術、市場の3ドメインを対象にそれぞれに必要な予測手法に基づき、イノベーション・マネジメントにおける予測のあり方についてのフレームワークの提案を行った。知識構造が動的に変化するに従ってイノベーション・マネジメントに資する情報はより複雑となっていく。上流の知識構造の変化は下流にまで影響を及ぼすなかで、自分が知らない分野を対象とした将来予測をしなければならない。

6.2 有望領域の特定

3章では、イノベーションの知識の上流のひとつである科学論文群を対象にした予測に焦点をあてた。科学のフェイズでは、つねに新たな知識が求められるがその中でも有望である知識を特定することがますます困難となってきた。具体的にはナノカーボン分野及び太陽光発電分野に関する知識構造から、萌芽な領域を予測しモデルの有効性を確認した。さらに多くは論文を単位とした萌芽性を議論するがその先の領域的議論にまで至っているケースは少ない。例えば論文単位で予測をすると多くの場合がレビュー論文になってしまう傾向にあるが、そのような結果だけでは判断が難しいという欠点がある。これはいくつかの企業担当者からも抽出された課題である。本研究では、萌芽と予測されたものを再度トピックモデルを用いて再解釈をすることで領域単位での萌芽予測を行った。萌芽論文が集中しているクラスタと全く存在しないクラスタが特定できることから、クラスタ単位で領域を特定することの有用性を示した。

6.3 有望関係の特定

4章では、技術融合に着目をしてこれまで無かった関係性を予測する手法を提案した。具体的には国際特許分類が有する階層構造について注目することで、共起の予測に寄与するとを示した。例えば、これまでに無かった初めての融合を予測するという視点においては、予測対象レイヤーの下位に当たるレイヤーの情報が有用であることが示された。この現象は予測を分類、回帰の双方

で実施したがいずれにおいても貢献することがわかった。一方で、上位のレイヤーが得意とするのは回帰モデルにおける分析であった。特許庁審査官によれば IPC のような特許分類は、明細書を主観により関係しそうな部署へ順次回覧していく。これは、人間はこれまで一度も起こっていない関連性を提案することは不得意である事を示している可能性があると考えた。一方で、徐々に起こってきた融合を観察してそれを加速させるということは得意であることを示すと考えた。データが持つ階層構造情報から特徴量を抽出するという発想は、IPC を用いた技術融合予測はもちろんのことこれまでのリンク予測問題としても取り扱われる例がすくない。本研究では学術的な貢献として、階層構造情報が異なる階層のリンク予測に有益であるということも示した。

6.4 経済主体の関係の特定

5 章では、これらで生まれた知識を活用して事業を行う経済主体の行動予測に焦点をあてた取引関係の予測である。企業の取引は、協調フィルタリングのような類似性に基づいて提案を行うと場合によっては競合企業を予測することになるという欠点がある。この章では、サプライチェーンの流れを考慮する特徴量が予測に寄与することを示すことによって、解釈性とモデルパフォーマンスの両立を図る予測手法を提案した。技術を競争優位とする企業にとっては今後、分野、地域など異なる領域への進出が不可避となる。まずは自社のサプライチェーンは熟知していてもこれから進出する領域のサプライチェーンについては未知であることが多い。そのような中で将来的な予測を行うには、取引の本質的情報を組み込んだ予測が有益であることを示した。具体的には企業が持つ業種分類に着目し、業種の流れをモノサービスの流れとした n-gram 確率を算出した。企業属性、ネットワーク属性、n-gram 特徴量を盛り込んだモデルを適用することで解釈可能性と予測パフォーマンスを両立したモデルを提案することが示された。

6.5 提案手法の組み合わせによる統合的予測手法によるイノベーション・マネジメント

知識構造が動的に変化する中、企業の研究開発戦略におけるイノベーション・マネジメントでは、今後展開すべき新たな分野の知識獲得が必要である。本研究ではこれらの手法の組み合わせにより実現できるあらたなイノベーショ

ン・マネジメントを提案する。

企業における意思決定は将来を見据えた中長期的なビジョンのもと研究開発方針を策定するが、一方でそれらビジョンに基づきつつ短期的な製品開発を行いキャッシュを得る必要がある。知識構造の変化が激しい昨今の科学技術を元にしたイノベーション・マネジメントにおいては、これらの2つの時間視点を同時に考慮しながら研究開発を行っていかねばならないという点が一層困難になる。

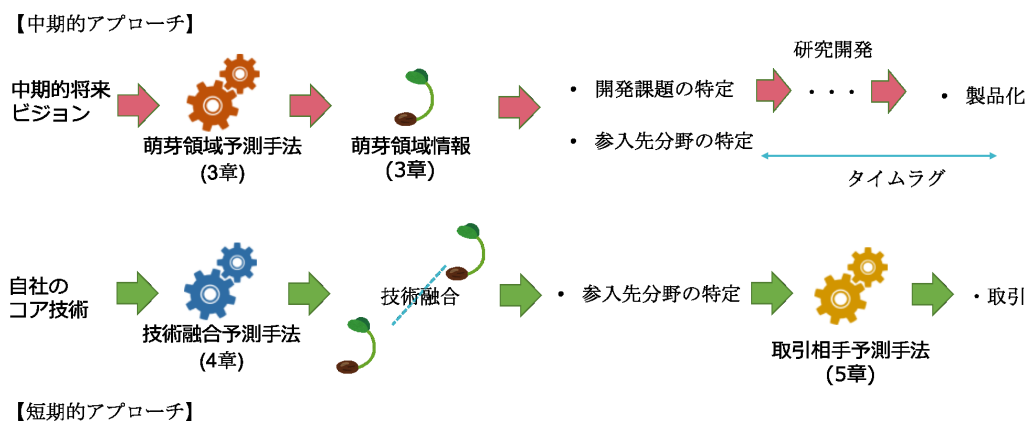


図 90. イノベーション・マネジメントプロセス（上図：中期的アプローチ。下図：短期的アプローチ。）

本研究の第3章では、科学レイヤーにおける萌芽予測手法について議論をした。この手法は、図90の上図に示すとおり、企業において中期的なアプローチに資するものである。企業は自社の中期的将来ビジョンに基づき、今後の科学知識の萌芽情報を入手しようとする。ここで言う萌芽情報は、自社の得意分野とは限らない。むしろ自社の得意分野と異なる知識が得られることの方が多い。そのような異分野の知識から今後の開発課題の特定や、参入先分野の特定を行う。したがって開発課題の特定から、研究開発、製品化に至るまでは通常の研究開発（自社の得意分野の研究開発）以上にリードタイムが必要となる。その為、参入先分野の特定との足並みが揃わないというタイムラグが発生してしまう。科学レイヤーにおける萌芽予測だけではそのような欠点が生じ、現実のイノベーション・マネジメントにおいて乖離が生じてしまう。また、本研究の第4章では、技術レイヤーにおける技術融合予測手法について議論をした。また、第5章では取引相手予測手法について議論をした。これらは、企業のイノベーションプロセスにおいて比較的下流において有効な手法である。したがって、

図 90 の下図に示すように短期的な戦略に資する手法と言える。これらの手法を用いる場合、企業はまず自社のコア技術を元に技術融合予測手法を用いて融合分野情報を獲得する。この情報により、開発課題や新たな製品提案を元に参入先分野を特定する。参入先の分野についてのサプライチェーンに関し十分な知識を持ち合わせていないことから、取引相手予測手法に基づき当該分野における取引構造の把握を実行が有効と成る。この短期的アプローチにおける技術融合予測で得られる技術知識の組み合わせは、将来生じるであろう技術知識の組み合わせである一方で、それらが有望（すなわち今後伸びる技術知識）であるかどうかについては問わない。技術融合においても一定の将来的萌芽性の要素を得るには、4章で提案した手法に一定の萌芽情報を付与する必要がある。中期的アプローチでは、未踏分野における研究開発ゆえ生じるリードタイムによるタイムラグという短所があり、短期的アプローチでは、将来的萌芽性の欠如という短所がある。図 91 ではこれらを相互に補間することで統合的予測手法によるイノベーション・マネジメントのフレームワークとして提示するものである。すなわち、技術融合で得られた知識を萌芽予測手法へ取り込むことで中期的なアプローチに用い、萌芽領域情報で得られた知識を技術融合予測手法へ取り込むことで短期的なアプローチに用いることで、それらの欠点を補間することが可能となる。

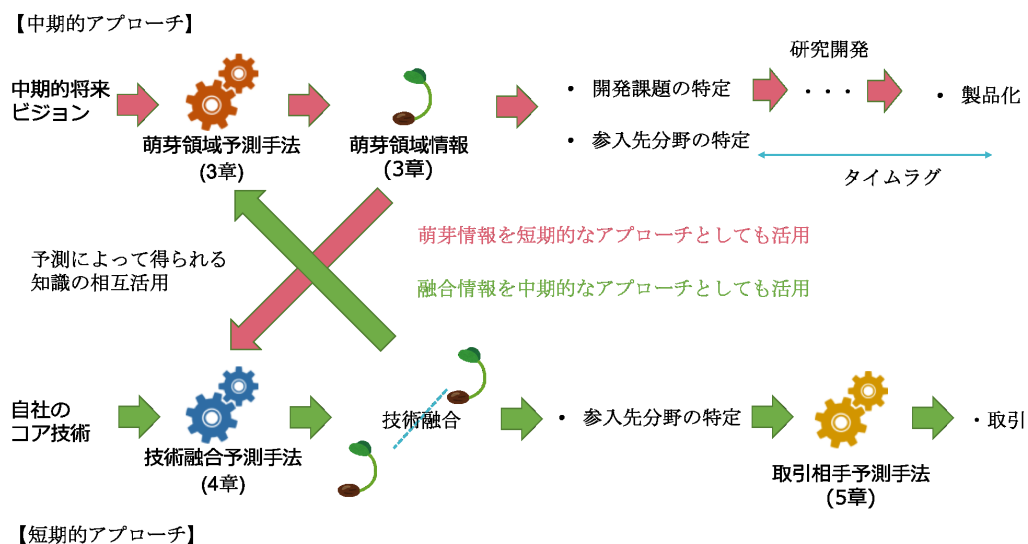


図 91. 統合的予測手法に基づくイノベーション・マネジメント

企業のイノベーション・マネジメントにおいて有効となる情報提供ツールや

予測モデルは近年多く提案されつつ有る。予測精度の観点で見れば極めて優秀なモデルであっても、実際に企業の研究開発の実務フローに単に当該モデルを組み込むだけでは現実との乖離が生じてしまう。特徴量の解釈性がその一つであり、第3章から第5章ではそれらを乗り越える手法を提案した。しかしながら、それらの手法についても単独で企業の研究開発フローにそのまま取り込むだけでは、上記で述べたような課題がある。すなわち、企業のビジョンは長期的な視点と短期的な視点の双方が存在し、片方だけの視点で情報提供ツールや予測モデルを用いるだけでは、バランスを欠いてしまう。本章では、3章から5章を総合的に用いることでこれらの二つの時間的視点における欠点を相互補完するイノベーション・マネジメントフレームワークを提案するものである。

7 第7章:まとめ

情報の増加速度が早まり、情報の相対的な質の担保が困難となり、知識同士の距離がより近くなり意志決定の求められるスピードはますます早くなる。また、科学技術知識や産業構造もダイナミックに変化しはじめており、イノベーション戦略の意志決定における課題は今後より一層大きな問題となっていく。情報技術によるスマート化をはじめとした構造変化によって、今後あらゆる技術の境界線 (Boundary) がぼやけて (Blur) いく。既に現在の自動車産業は機械のみならず電気・電子・情報工学の知が不可欠となっている。航空機の制御システムの技術の一つにフライバイワイヤがあげられる。元来、航空機の制御は油圧を中心としていた。そのような航空機の場合は気体状態によって同じ操作をしても得られる舵角が異なる。フライバイワイヤでは、気体周辺の空力情報をセンシングし、圧縮性、密度、速度の情報を演算し制御系にフィードバックする。この技術により、気体状態が異なっても同じ操作で同じ舵角が得られるようになった。これはセンシング、制御工学、情報工学、システム工学等を総合的に必要とする技術であり、既存の航空工学のみでは実現することはできなかった。この流れは自動車にも取り入れられることとなり、ドライブバイワイヤと称されている。ハードウェアの観点だけで見れば部品点数が大幅に削減される一方で、システムとしての統合はより難度があがり、必要となる知識はより高度となっている。自動運転が本格的に導入されつつあるなかで、これにもましてより一層高度な情報工学やパターン認識、機械学習の技術が導入される。2015年にはトヨタが Preferred Networks 社に 10 億円の出資をした。同社の設立は 2014 年 3 月であり、設立 1 年余りの会社への出資額としては異例とあってよい。同社が有する技術に期待をしていることのあらわれでもある。米国では既に Google がエネルギー事業をはじめ、イーロン・マスクは宇宙事業に乗り出している。このような極端な例を挙げずとも、科学技術のありかたはあらゆる場所で大きく変化している。本研究の各章ではイノベーションプロセスの上流である科学、中流としての技術、下流としての産業に注目をしてそれぞれに必要なとされる予測手法を提案し、その有効性を示した。またそれらを網羅的に用いることで、今後進出する新たな分野において、将来的に萌芽であり自社の相性のある技術分野を特定し、かつそのような分野のサプライチェーン把握から顧客特定を行う一貫したイノベーション・マネジメント戦略を提案

した。さらに、特に4章と5章では、データが持つ構造情報を活用して特徴量を構築することによる優位性を示した。すなわち、4章における知識の融合にはデータが持つ階層的概念を抽出して取り込むことが有効であることを示した。5章における経済主体の予測としては、モノやサービスの流れを取り込むことが有効であることを示した。これらはデータ自体から得られる特徴量ではなく、構造情報を元にして構築した特徴量である。膨大なデータが存在し、今後ますます機械学習を用いた予測手法は、コモディティ化する。その中で必要なのは、予測対象の本質的特徴を抽出することである。ましてや、未知の分野の知識分野を対象とした意志決定に資するべき予測活動においては、精度が高くとも複雑なモデルや抽象的な特徴量を用いているものは実際のユースケースでは積極的に活用しにくい。本研究ではそのような課題意識から、機械による判断ではなく人間の意思決定プロセスを前提とするイノベーション・マネジメントの場面に寄り添った手法の提案を行った。今後、例えば農業がIT技術を強みとした製造業となったり、これまで製造業であった企業が製造業を生業としなくなったりとすることはあらゆるシーンで起こる。分野の領域が曖昧な世界で異なる世界についての知識を獲得することはますます必要かつ困難となっていくであろう。本研究の貢献は、異なる世界において知識を獲得しビジネスチャンスを獲得するための手法の提案である。

8 参考文献

参考文献

- [1] F. Narin and K. S. Hamilton (1996) “Bibliometric performance measures,” *Scientometrics*, Vol. 36, No. 3, pp. 293–310.
- [2] J. D. Adams (1990) “Fundamental stocks of knowledge and productivity growth,” *Journal of Political Economy*, Vol. 98, No. 4, pp. 673–702.
- [3] L. Sveikauskas (1981) “Technological inputs and multifactor productivity growth,” *The Review of Economics and Statistics*, pp. 275–282.
- [4] N. Rosenberg (1974) “Science, invention and economic growth,” *The Economic Journal*, Vol. 84, No. 333, pp. 90–108.
- [5] E. Mansfield (1972) “Contribution of RandD to economic growth in the United States,” *Science*, Vol. 175, No. 4021, pp. 477–486.
- [6] M. E. Porter and S. Stern (2001) “National innovative capacity,” *The global competitiveness report*, Vol. 2002, pp. 102–118.
- [7] S. H. Zyoud, S. W. Al-Jabi, W. M. Sweileh, and R. Awang (2014) “A bibliometric analysis of toxicology research productivity in Middle Eastern Arab countries during a 10-year period (2003-2012).,” *Health Research Policy and Systems*, Vol. 12, No. 1.
- [8] S. Meo, A. Usmani, M. Vohra, and I. Bukhari (2013) “Impact of GDP, spending on RandD, number of universities and scientific journals on research publications in pharmacological sciences in Middle East,” *Eur Rev Med Pharmacol Sci*, Vol. 17, No. 20, pp. 2697–705.
- [9] S. A. Meo, A. A. Al Masri, A. M. Usmani, A. N. Memon, and S. Z. Zaidi (2013) “Impact of GDP, spending on RandD, number of universities and scientific journals on research publications among Asian countries,” *PLoS One*, Vol. 8, No. 6, p. e66449.
- [10] G. Linden, J. Hart, S. A. Lenway, and T. P. M. 1 (1998) “Flying Geese as Moving Targets: Are Korea and Taiwan Catching up with Japan in Advanced Displays?” *Industry and Innovation*, Vol. 5, No. 1, pp.

11-34.

- [11] A. Schrank (2003) “Foreign investors, “flying geese,” and the limits to export-led industrialization in the Dominican Republic,” *Theory and Society*, Vol. 32, No. 4, pp. 415–443.
- [12] I. Sakata and H. Sasaki (2013) “Scientific catch-up in Asian economies: A case study for solar cell.”
- [13] K. Hashimoto, H. Irie, and A. Fujishima (2005) “TiO₂ photocatalysis: a historical overview and future prospects,” *Japanese journal of applied physics*, Vol. 44, No. 12R, p. 8269.
- [14] D. Tryk, A. Fujishima, and K. Honda (2000) “Recent topics in photoelectrochemistry: achievements and future prospects,” *Electrochimica acta*, Vol. 45, No. 15, pp. 2363–2376.
- [15] R. N. Kostoff and R. R. Schaller (2001) “Science and technology roadmaps,” *IEEE Transactions on engineering management*, Vol. 48, No. 2, pp. 132–143.
- [16] R. Galvin (2004) “Roadmapping—a practitioner’s update,” *Technological Forecasting and Social Change*, Vol. 71, No. 1, pp. 101–103.
- [17] R. Phaal, C. J. Farrukh, and D. R. Probert (2004) “Technology roadmapping—a planning framework for evolution and revolution,” *Technological forecasting and social change*, Vol. 71, No. 1, pp. 5–26.
- [18] L. M. Branscomb and J. H. Keller (1998) “Towards a research and innovation policy,” *Investing in innovation*, pp. 462–496.
- [19] R. Galvin (1998) “Science roadmaps,” *Science*, Vol. 280, No. 5365, pp. 803–804.
- [20] P. Groenveld (1997) “Roadmapping integrate business and technology,” *Research Technology Management*, Vol. 40, No. 5, p. 48.
- [21] 小笠原敦 (2013) 「バックキャストに適した科学技術予測の方法論課題解決思考を重視した研究開発の推進」, <http://data.nistep.go.jp/dspace/handle/11035/2372> Accessed on 7th Apr. 2017, 5月.
- [22] Y. Kajikawa, Y. Takeda, and K. Matushima (2008a) “Computer-assisted roadmapping: A case study in energy research,” *Management of Engineering Technology*, 2008. PICMET 2008. Portland Interna-

tional Conference on.

- [23] N. Schwab Richard and E. Rex Walter “Lough John (1971),” Inventory of Diderot’s Encyclopedie,” *Studies on Voltaire and the Eighteenth century*, Vol. 80.
- [24] J. A. Evans (2013) “Future science,” *science*, Vol. 342, No. 6154, pp. 44–45.
- [25] A. N. Tabah (1999) “Literature dynamics: Studies on growth, diffusion, and epidemics,” *Annual review of information science and technology (ARIST)*, Vol. 34, pp. 249–86.
- [26] D. de Solla Price (1961) “Science since Babylon,” *American Journal of Physics*, Vol. 29, No. 12, p. 863.
- [27] D. de Solla Price (1965) “Networks of Scientific Papers,” *Science*, Vol. 149, No. 3683, pp. 510–515.
- [28] A. F. Van Raan (2000) “On growth, ageing, and fractal differentiation of science,” *Scientometrics*, Vol. 47, No. 2, pp. 347–362.
- [29] L. Bornmann and R. Mutz (2015) “Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references,” *Journal of the Association for Information Science and Technology*, Vol. 66, No. 11, pp. 2215–2222.
- [30] P. O. Larsen and M. Von Ins (2010) “The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index,” *Scientometrics*, Vol. 84, No. 3, pp. 575–603.
- [31] B. Statzner and V. H. Resh (2010) “Negative changes in the scientific publication process in ecology: potential causes and consequences,” *Freshwater Biology*, Vol. 55, No. 12, pp. 2639–2653.
- [32] 科学技術・学術政策研究所 (NISTEP) (2011) 「研究論文に着目した日本の大学ベンチマーキング 2011」, <http://hdl.handle.net/11035/1144> Accessed on 7th Apr. 2017.
- [33] 文部科学省「科学研究費助成事業の審査システム改革について」.
- [34] 佐野幸吉 (1985) 「経済摩擦の軽減を目ざして科学技術の現況転換を考える」, 『工業教育』, 第 33 巻, 第 4 号, 52–64 頁.
- [35] Editorial (2005) “The cost of salami slicing [editorial].”
- [36] E. Mansfield (1991) “Academic research and industrial innovation,”

- Research policy*, Vol. 20, No. 1, pp. 1–12.
- [37] G. S. McMillan, F. Narin, and D. L. Deeds (2000) “An analysis of the critical role of public science in innovation: the case of biotechnology,” *Research policy*, Vol. 29, No. 1, pp. 1–8.
- [38] B. Van Looy, E. Zimmermann, R. Veugelers, A. Verbeek, J. Mello, and K. Debackere (2003) “Do science-technology interactions pay off when developing technology?” *Scientometrics*, Vol. 57, No. 3, pp. 355–367.
- [39] O. Sorenson and L. Fleming (2004) “Science and the diffusion of knowledge,” *Research policy*, Vol. 33, No. 10, pp. 1615–1634.
- [40] L. Fleming and O. Sorenson (2004) “Science as a map in technological search,” *Strategic Management Journal*, Vol. 25, No. 8-9, pp. 909–928.
- [41] N. Fukuzawa and T. Ida (2016) “Science linkages between scientific articles and patents for leading scientists in the life and medical sciences field: the case of Japan,” *Scientometrics*, Vol. 106, No. 2, pp. 629–644.
- [42] 玉田俊平太他 (2010) 「科学の技術革新に与える影響について」, 『ビジネス and アカウンティングレビュー = Business and accounting review』, 第 5 号, 77–91 頁.
- [43] P. A. Geroski (2000) “Models of technology diffusion,” *Research policy*, Vol. 29, No. 4, pp. 603–625.
- [44] B. Godin (2006) “The linear model of innovation: The historical construction of an analytical framework,” *Science, Technology, and Human Values*, Vol. 31, No. 6, pp. 639–667.
- [45] E. H. Kessler and A. K. Chakrabarti (1996) “INNOVATION SPEED: A CONCEPTUAL MODEL OF CONTEXT, ANTECEDENTS, AND OUTCOMES.,” *Academy of Management Review*, Vol. 21, No. 4, pp. 1143 - 1191.
- [46] H. Sonnenberg (1993) “Balancing speed and quality in product innovation,” *Canadian Business Review*, Vol. 20, pp. 19–19.
- [47] G. D. Markman, P. T. Gianiodis, P. H. Phan, and D. B. Balkin (2005) “Innovation speed: Transferring university technology to market,”

Research Policy, Vol. 34, No. 7, pp. 1058–1075.

- [48] H. M. Collins and R. Evans (2002) “The third wave of science studies: Studies of expertise and experience,” *Social studies of science*, Vol. 32, No. 2, pp. 235–296.
- [49] D. Rigby and C. Zook (2002) “Open-market innovation,” *Harvard business review*, Vol. 80, No. 10, pp. 80–93.
- [50] 内閣府「科学技術イノベーション総合戦略 新時代日本創造への挑戦概要（簡略版）」。
- [51] 大岩秀和・数原良彦・淡島英輝（2017）「人工知能技術の導入判断にかかる意思決定者のバイアスとその解決に向けて」，第 31 卷，Annual Conference of the Japanese Society for Artificial Intelligence.
- [52] Z. C. Lipton (2016) “The Mythos of Model Interpretability,” 06.
- [53] B. Goodman and S. Flaxman (2016) “European Union regulations on algorithmic decision-making and a ”right to explanation”,” 06.
- [54] G. Palla, A.-L. Barabasi, and T. Vicsek (2007) “Quantifying social group evolution,” *Nature*, Vol. 446, No. 7136, pp. 664–667.
- [55] D. Greene, D. Doyle, and P. Cunningham (2010) “Tracking the evolution of communities in dynamic social networks,” in *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*, pp. 176–183, IEEE.
- [56] S. Y. Bhat and M. Abulaish (2015) “HOCTracker: Tracking the evolution of hierarchical and overlapping communities in dynamic social networks,” *IEEE Transactions on Knowledge and Data engineering*, Vol. 27, No. 4, pp. 1019–1013.
- [57] R. Mall, R. Langone, and J. A. Suykens (2015) “Netgram: Visualizing Communities in Evolving Networks,” *PloS one*, Vol. 10, No. 9, p. e0137502.
- [58] H. W. Kroto, J. R. Heath, S. C. O’Brien, R. F. Curl, R. E. Smalley et al. (1985) “C 60: buckminsterfullerene,” *Nature*, Vol. 318, No. 6042, pp. 162–163.
- [59] S. Iijima (1991) “Helical microtubules of graphitic carbon,” *nature*, Vol. 354, No. 6348, p. 56.
- [60] F. D’Souza and O. Ito (2012) “Photosensitized electron transfer pro-

- cesses of nanocarbons applicable to solar cells,” *Chemical Society Reviews*, Vol. 41, No. 1, pp. 86–96.
- [61] A. Burke (2007) “R and D considerations for the performance and application of electrochemical capacitors,” *Electrochimica Acta*, Vol. 53, No. 3, pp. 1083–1091.
- [62] K. Suzuki, M. Yamaguchi, M. Kumagai, and S. Yanagida (2002) “Application of carbon nanotubes to counter electrodes of dye-sensitized solar cells,” *Chemistry Letters*, Vol. 32, No. 1, pp. 28–29.
- [63] U. Kasavajjula, C. Wang, and A. J. Appleby (2007) “Nano-and bulk-silicon-based insertion anodes for lithium-ion secondary cells,” *Journal of Power Sources*, Vol. 163, No. 2, pp. 1003–1039.
- [64] C. Liu, F. Li, L.-P. Ma, and H.-M. Cheng (2010) “Advanced materials for energy storage,” *Advanced materials*, Vol. 22, No. 8.
- [65] B. C. Edwards (2000) “Design and deployment of a space elevator,” *Acta Astronautica*, Vol. 47, No. 10, pp. 735–744.
- [66] N. M. Pugno (2006) “On the strength of the carbon nanotube-based space elevator cable: from nanomechanics to megamechanics,” *Journal of Physics: Condensed Matter*, Vol. 18, No. 33, p. S1971.
- [67] N. M. Pugno (2007) “The role of defects in the design of space elevator cable: from nanotube to megatube,” *Acta Materialia*, Vol. 55, No. 15, pp. 5269–5279.
- [68] K. K. Franklin and J. K. Hart (2007) “Idea generation and exploration: benefits and limitations of the policy Delphi research method,” *Innovative Higher Education*, Vol. 31, No. 4, pp. 237–246.
- [69] R. Beretta (1996) “A critical review of the Delphi technique: The Delphi technique is a useful method for surveying informed opinion but researchers should be aware of all its pros and cons, suggests Ruth Beretta,” *Nurse researcher*, Vol. 3, No. 4, pp. 79–89.
- [70] Y. Kajikawa, J. Yoshikawa, Y. Takeda, and K. Matsushima (2008b) “Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy,” *Technological Forecasting and Social Change*, Vol. 75, No. 6, pp. 771–782.
- [71] A. Martin-Martin, E. Orduna-Malea, J. M. Ayllon, and E. D. Lopez-

- Cozar (2016) “A two-sided academic landscape: portrait of highly-cited documents in Google Scholar (1950-2013),” *arXiv preprint arXiv:1607.02861*.
- [72] M. Hashimoto, I. Sakata, Y. Kajikawa, Y. Takeda, and K. Matsushima (2009) “Academic landscape of innovation research and innovation policy by network Science,” *Hitotsubashi Business Review*, Vol. 56, No. 4, pp. 194–211.
- [73] Y. Takeda and Y. Kajikawa (2008) “Optics: A bibliometric approach to detect emerging research domains and intellectual bases,” *Scientometrics*, Vol. 78, No. 3, pp. 543–558.
- [74] K. Borner, C. Chen, and K. W. Boyack (2003) “Visualizing knowledge domains,” *Annual review of information science and technology*, Vol. 37, No. 1, pp. 179–255.
- [75] R. N. Kostoff, D. R. Toothman, H. J. Eberhart, and J. A. Humenik (2001) “Text mining using database tomography and bibliometrics: A review,” *Technological Forecasting and Social Change*, Vol. 68, No. 3, pp. 223–253.
- [76] K. W. Boyack, B. N. Wylie, and G. S. Davidson (2002) “Domain visualization using VxInsight for science and technology management,” *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 9, pp. 764–774.
- [77] C. Chen, T. Cribbin, R. Macredie, and S. Morar (2002) “Visualizing and tracking the growth of competing paradigms: Two case studies,” *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 8, pp. 678–689.
- [78] H. Small (2006) “Tracking and predicting growth areas in science,” *Scientometrics*, Vol. 68, No. 3, pp. 595–610.
- [79] N. Shibata, Y. Kajikawa, Y. Takeda, and K. Matsushima (2008) “Detecting emerging research fronts based on topological measures in citation networks of scientific publications,” *Technovation*, Vol. 28, No. 11, pp. 758–775.
- [80] W. Glanzel (2012) “Bibliometric methods for detecting and analysing emerging research topics,” *El profesional de la informacion*, Vol. 21,

No. 1, pp. 194–201.

- [81] J. Winnink and R. J. Tijssen (2015) “Early stage identification of breakthroughs at the interface of science and technology: lessons drawn from a landmark publication.,” *Scientometrics*, Vol. 102, No. 1, pp. 113–134.
- [82] J. Adams (2005) “Early citation counts correlate with accumulated impact,” *Scientometrics*, Vol. 63, No. 3, pp. 567–581.
- [83] W. Goffman and V. Newill (1964) “Generalization of epidemic theory,” *Nature*, Vol. 204, No. 4955, pp. 225–228.
- [84] L. Bettencourt, D. Kaiser, J. Kaur, C. Castillo-Chavez, and D. Wojick (2008) “Population modeling of the emergence and development of scientific fields,” *Scientometrics*, Vol. 75, No. 3, pp. 495–518.
- [85] C. Chen, Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino (2009) “Towards an explanatory and computational theory of scientific discovery,” *Journal of Informetrics*, Vol. 3, No. 3, pp. 191–209.
- [86] L. Li and H. Tong (2015) “The child is father of the man: Foresee the success at the early stage,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 655–664, ACM.
- [87] Y. Dong, R. A. Johnson, and N. V. Chawla (2015) “Will this paper increase your h-index?: Scientific impact prediction,” in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 149–158, ACM.
- [88] F. Davletov, A. S. Aydin, and A. Cakmak (2014) “High impact academic paper prediction using temporal and topological features,” in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 491–498, ACM.
- [89] Arxiv (2016) “arxiv.org,” March, URL: <http://Arxiv.Org>.
- [90] T. Chakraborty, S. Kumar, P. Goyal, N. Ganguly, and A. Mukherjee (2014) “Towards a stratified learning approach to predict future citation counts,” in *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 351–360, IEEE Press.
- [91] D. Wang, C. Song, and A.-L. Barabasi (2013) “Quantifying long-term

- scientific impact,” *Science*, Vol. 342, No. 6154, pp. 127–132.
- [92] 森純一郎・榊剛史・梶川裕矢・坂田一郎 (2014) 「萌芽領域特定のための大規模論文情報を用いた引用予測に関する研究」, 第 28 卷, Annual Conference of the Japanese Society for Artificial Intelligence.
- [93] H. Sasaki, T. Hara, and I. Sakata (2016a) “Identifying Emerging Research Related to Solar Cells Field using a Machine Learning Approach,” *Journal of Sustainable Development of Energy, Water and Environment Systems*, Vol. 4, No. 4, pp. 418–429.
- [94] H. Sasaki, T. Hara, and I. Sakata (2016b) “Prediction of emerging papers in nanocarbon materials-related research using a citation network,” in *Management of Engineering and Technology (PICMET), 2016 Portland International Conference on*, pp. 2636–2644, IEEE.
- [95] Y. Kajikawa, J. Ohno, Y. Takeda, K. Matsushima, and H. Komiyama (2007) “Creating an academic landscape of sustainability science: an analysis of the citation network,” *Sustainability Science*, Vol. 2, No. 2, p. 221.
- [96] S. Lizin, J. Leroy, C. Delvenne, M. Dijk, E. De Schepper, and S. Van Passel (2013) “A patent landscape analysis for organic photovoltaic solar cells: identifying the technology’s development phase,” *Renewable energy*, Vol. 57, pp. 5–11.
- [97] N. Shibata, Y. Kajikawa, and I. Sakata (2010) “Extracting the commercialization gap between science and technology—Case study of a solar cell,” *Technological Forecasting and Social Change*, Vol. 77, No. 7, pp. 1147–1155.
- [98] W. Bauhofer and J. Z. Kovacs (2009) “A review and analysis of electrical percolation in carbon nanotube polymer composites,” *Composites Science and Technology*, Vol. 69, No. 10, pp. 1486–1498.
- [99] M. Kumar and Y. Ando (2010) “Chemical vapor deposition of carbon nanotubes: a review on growth mechanism and mass production,” *Journal of nanoscience and nanotechnology*, Vol. 10, No. 6, pp. 3739–3758.
- [100] X.-M. Liu, Z. dong Huang, S. woon Oh, B. Zhang, P.-C. Ma, M. M. Yuen, and J.-K. Kim (2012) “Carbon nanotube (CNT)-based compos-

- ites as electrode material for rechargeable Li-ion batteries: a review,” *Composites Science and Technology*, Vol. 72, No. 2, pp. 121–144.
- [101] M. E. Newman (2006) “Modularity and community structure in networks,” *Proceedings of the national academy of sciences*, Vol. 103, No. 23, pp. 8577–8582.
- [102] L. C. Freeman (1978) “Centrality in social networks conceptual clarification,” *Social networks*, Vol. 1, No. 3, pp. 215–239.
- [103] P. Bonacich (1972) “Technique for analyzing overlapping memberships,” *Sociological methodology*, Vol. 4, pp. 176–185.
- [104] R. S. Burt (2004) “Structural holes and good ideas 1,” *American journal of sociology*, Vol. 110, No. 2, pp. 349–399.
- [105] D. J. Watts and S. H. Strogatz (1998) “Collective dynamics of ‘small-world’ networks,” *nature*, Vol. 393, No. 6684, pp. 440–442.
- [106] S. Brin and L. Page (1998) “The anatomy of a large-scale hypertextual web search engine,” *Computer networks and ISDN systems*, Vol. 30, No. 1, pp. 107–117.
- [107] R. Guimera and L. A. N. Amaral (2005) “Functional cartography of complex metabolic networks,” *Nature*, Vol. 433, No. 7028, pp. 895–900.
- [108] L. C. Freeman, S. P. Borgatti, and D. R. White (1991) “Centrality in valued graphs: A measure of betweenness based on network flow,” *Social networks*, Vol. 13, No. 2, pp. 141–154.
- [109] J. M. Kleinberg (1999) “Authoritative Sources in a Hyperlinked Environment,” *J. ACM*, Vol. 46, No. 5, pp. 604–632, September.
- [110] D. M. Blei, A. Y. Ng, and M. I. Jordan (2003) “Latent dirichlet allocation,” *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022.
- [111] C.-K. Yau, A. Porter, N. Newman, and A. Suominen (2014) “Clustering scientific documents with topic modeling,” *Scientometrics*, Vol. 100, No. 3, pp. 767–786.
- [112] Q. He, B. Chen, J. Pei, B. Qiu, P. Mitra, and L. Giles (2009) “Detecting topic evolution in scientific literature: How can citations help?” in *Proceedings of the 18th ACM conference on Information and knowl-*

- edge management*, pp. 957–966, ACM.
- [113] X. Liu, J. Zhang, and C. Guo (2012) “Full-text citation analysis: enhancing bibliometric and scientific publication ranking,” in *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 1975–1979, ACM.
- [114] H. Jiang, M. Qiang, and P. Lin (2016) “A topic modeling based bibliometric exploration of hydropower research,” *Renewable and Sustainable Energy Reviews*, Vol. 57, pp. 226–237.
- [115] C. Sievert and K. E. Shirley (2014) “LDAvis: A method for visualizing and interpreting topics,” in *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pp. 63–70.
- [116] R. Deveaud, E. SanJuan, and P. Bellot (2014) “Accurate and effective latent concept modeling for ad hoc information retrieval,” *Document numerique*, Vol. 17, No. 1, pp. 61–84.
- [117] Q. Zhang, J.-Q. Huang, M.-Q. Zhao, W.-Z. Qian, and F. Wei (2011) “Carbon nanotube mass production: principles and processes,” *ChemSusChem*, Vol. 4, No. 7, pp. 864–889.
- [118] Y. Lan, Y. Wang, and Z. Ren (2011) “Physics and applications of aligned carbon nanotubes,” *Advances in Physics*, Vol. 60, No. 4, pp. 553–678.
- [119] S. H. Lee, D. H. Lee, W. J. Lee, and S. O. Kim (2011) “Tailored assembly of carbon nanotubes and graphene,” *Advanced Functional Materials*, Vol. 21, No. 8, pp. 1338–1354.
- [120] S. D. Sarma, S. Adam, E. Hwang, and E. Rossi (2011) “Electronic transport in two-dimensional graphene,” *Reviews of Modern Physics*, Vol. 83, No. 2, p. 407.
- [121] X. Huang, Z. Yin, S. Wu, X. Qi, Q. He, Q. Zhang, Q. Yan, F. Boey, and H. Zhang (2011) “Graphene-based materials: synthesis, characterization, properties, and applications,” *small*, Vol. 7, No. 14, pp. 1876–1902.
- [122] R. Saito, M. Hofmann, G. Dresselhaus, A. Jorio, and M. Dresselhaus (2011) “Raman spectroscopy of graphene and carbon nanotubes,” *Advances in Physics*, Vol. 60, No. 3, pp. 413–550.

- [123] Y. Li, D. Li, and G. Wang (2011) “Methane decomposition to CO x-free hydrogen and nano-carbon material on group 8–10 base metal catalysts: a review,” *Catalysis today*, Vol. 162, No. 1, pp. 1–48.
- [124] L. Yan, F. Zhao, S. Li, Z. Hu, and Y. Zhao (2011) “Low-toxic and safe nanomaterials by surface-chemical design, carbon nanotubes, fullerenes, metallofullerenes, and graphenes,” *Nanoscale*, Vol. 3, No. 2, pp. 362–382.
- [125] V. Singh, D. Joung, L. Zhai, S. Das, S. I. Khondaker, and S. Seal (2011) “Graphene based materials: past, present and future,” *Progress in materials science*, Vol. 56, No. 8, pp. 1178–1271.
- [126] R. Leary and A. Westwood (2011) “Carbonaceous nanomaterials for the enhancement of TiO₂ photocatalysis,” *Carbon*, Vol. 49, No. 3, pp. 741–772.
- [127] H.-L. Yip and A. K.-Y. Jen (2012) “Recent advances in solution-processed interfacial materials for efficient and stable polymer solar cells,” *Energy and Environmental Science*, Vol. 5, No. 3, pp. 5994–6011.
- [128] W. Chen, M. P. Nikiforov, and S. B. Darling (2012) “Morphology characterization in organic and hybrid solar cells,” *Energy and Environmental Science*, Vol. 5, No. 8, pp. 8045–8074.
- [129] P. Kumar and S. Chand (2012) “Recent progress and future aspects of organic solar cells,” *Progress in Photovoltaics: Research and applications*, Vol. 20, No. 4, pp. 377–415.
- [130] J. Boucle and J. Ackermann (2012) “Solid-state dye-sensitized and bulk heterojunction solar cells using TiO₂ and ZnO nanostructures: recent progress and new concepts at the borderline,” *Polymer International*, Vol. 61, No. 3, pp. 355–373.
- [131] H. Zhou, L. Yang, and W. You (2012) “Rational design of high performance conjugated polymers for organic solar cells,” *Macromolecules*, Vol. 45, No. 2, pp. 607–632.
- [132] Y. Ooyama and Y. Harima (2012) “Photophysical and Electrochemical Properties, and Molecular Structures of Organic Dyes for Dye-Sensitized Solar Cells,” *ChemPhysChem*, Vol. 13, No. 18, pp. 4032–

4080.

- [133] A. Mishra and P. Bauerle (2012) “Small molecule organic semiconductors on the move: promises for future solar energy technology,” *Angewandte Chemie International Edition*, Vol. 51, No. 9, pp. 2020–2067.
- [134] L.-L. Li, Y.-C. Chang, H.-P. Wu, and E. W.-G. Diau (2012) “Characterisation of electron transport and charge recombination using temporally resolved and frequency-domain techniques for dye-sensitised solar cells,” *International Reviews in Physical Chemistry*, Vol. 31, No. 3, pp. 420–467.
- [135] L. Dou, J. You, J. Yang, C.-C. Chen, Y. He, S. Murase, T. Moriarty, K. Emery, G. Li, and Y. Yang (2012) “Tandem polymer solar cells featuring a spectrally matched low-bandgap polymer,” *Nature Photonics*, Vol. 6, No. 3, pp. 180–185.
- [136] T. Berger, D. Monllor-Satoca, M. Jankulovska, T. Lana-Villarreal, and R. Gomez (2012) “The electrochemistry of nanostructured titanium dioxide electrodes,” *ChemPhysChem*, Vol. 13, No. 12, pp. 2824–2875.
- [137] G. Calogero, A. Bartolotta, G. Di Marco, A. Di Carlo, and F. Bonaccorso (2015) “Vegetable-based dye-sensitized solar cells,” *Chemical Society Reviews*, Vol. 44, No. 10, pp. 3244–3294.
- [138] J. Wu, Z. Lan, J. Lin, M. Huang, Y. Huang, L. Fan, and G. Luo (2015) “Electrolytes in dye-sensitized solar cells,” *Chemical reviews*, Vol. 115, No. 5, pp. 2136–2173.
- [139] L. Lu, T. Zheng, Q. Wu, A. M. Schneider, D. Zhao, and L. Yu (2015) “Recent advances in bulk heterojunction polymer solar cells,” *Chemical reviews*, Vol. 115, No. 23, pp. 12666–12731.
- [140] F. Bella, C. Gerbaldi, C. Barolo, and M. Gratzel (2015) “Aqueous dye-sensitized solar cells,” *Chemical Society Reviews*, Vol. 44, No. 11, pp. 3431–3473.
- [141] P. Cheng and X. Zhan (2015) “Versatile third components for efficient and stable organic solar cells,” *Materials Horizons*, Vol. 2, No. 5, pp. 462–485.

- [142] C.-C. Chueh, C.-Z. Li, and A. K.-Y. Jen (2015) “Recent progress and perspective in solution-processed Interfacial materials for efficient and stable polymer and organometal perovskite solar cells,” *Energy and Environmental Science*, Vol. 8, No. 4, pp. 1160–1189.
- [143] Z. Liu, S. P. Lau, and F. Yan (2015) “Functionalized graphene and other two-dimensional materials for photovoltaic devices: device design and processing,” *Chemical Society Reviews*, Vol. 44, No. 15, pp. 5638–5679.
- [144] E. Singh and H. S. Nalwa (2015) “Graphene-based dye-sensitized solar cells: a review,” *Science of Advanced Materials*, Vol. 7, No. 10, pp. 1863–1912.
- [145] Z. Liang, Q. Zhang, L. Jiang, and G. Cao (2015) “ZnO cathode buffer layers for inverted polymer solar cells,” *Energy and Environmental Science*, Vol. 8, No. 12, pp. 3442–3476.
- [146] J. Albero, P. Atienzar, A. Corma, and H. Garcia (2015) “Efficiency Records in Mesoscopic Dye-Sensitized Solar Cells,” *The Chemical Record*, Vol. 15, No. 4, pp. 803–828.
- [147] A. Dasgupta, L. P. Rajukumar, C. Rotella, Y. Lei, and M. Terrones (2016) “Covalent three-dimensional networks of graphene and carbon nanotubes: synthesis and environmental applications,” *Nano Today*.
- [148] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov (2004) “Electric field effect in atomically thin carbon films,” *science*, Vol. 306, No. 5696, pp. 666–669.
- [149] E. H. Hwang (2007) “Carrier Transport in Two-Dimensional Graphene Layers,” *Physical Review Letters*, Vol. 98, No. 18.
- [150] K. I. Bolotin, K. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, and H. Stormer (2008) “Ultrahigh electron mobility in suspended graphene,” *Solid State Communications*, Vol. 146, No. 9, pp. 351–355.
- [151] S. Hitosugi, T. Yamasaki, and H. Isobe (2012) “Bottom-up synthesis and thread-in-bead structures of finite (n, 0)-zigzag single-wall carbon nanotubes,” *Journal of the American Chemical Society*, Vol. 134, No.

30, pp. 12442–12445.

- [152] S. Hitosugi, W. Nakanishi, T. Yamasaki, and H. Isobe (2011) “Bottom-up synthesis of finite models of helical (n, m)-single-wall carbon nanotubes,” *Nature Communications*, Vol. 2, p. 492.
- [153] A. Kay and M. Graetzel (1993) “Artificial photosynthesis. 1. Photosensitization of TiO₂ solar cells with chlorophyll derivatives and related natural porphyrins,” *Journal of Physical Chemistry;(United States)*, Vol. 97, No. 23.
- [154] M. Batmunkh, M. J. Biggs, and J. G. Shapter (2015) “Carbon Nanotubes for Dye-Sensitized Solar Cells,” *Small*, Vol. 11, No. 25, pp. 2963–2989.
- [155] A. Yella, A. K. Chandiran, C. Yi, C.-Y. Yeh, E. W.-G. Diao, H. N. Tsao, H.-W. Lee, M. K. Nazeeruddin, S. M. Zakeeruddin et al. (2011) 「Porphyrin-Sensitized Solar Cells with Cobalt (II/III) Y -Based Redox Electrolyte Exceed 12 Percent Efficiency」, 『*Science*』.
- [156] F. Kodama (1992) “Technology fusion and the new Research-and-Development,” *Harvard Business Review*, Vol. 70, No. 4, pp. 70–78.
- [157] J. Lind (2004) “Convergence: History of term usage and lessons for firm strategists,” in *15th Biennial ITS conference*, pp. 4–7.
- [158] J. A. Schumpeter (1939) *Business cycles*, Vol. 1: McGraw-Hill New York.
- [159] R. Nelson Richard and G. Winter Sidney (1982) “An evolutionary theory of economic change,” *Harvard Business School Press, Cambridge*.
- [160] W. S. Lee, E. J. Han, and S. Y. Sohn (2015) “Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents,” *Technological Forecasting and Social Change*, Vol. 100, pp. 317–329.
- [161] C.-S. Curran (2013) *The anticipation of converging industries*: Springer.
- [162] M. Lee, J. Lee, and Y. Cho (2009) “How a convergence product affects related markets: The case of the mobile phone,” *ETRI journal*, Vol. 31, No. 2, pp. 215–224.

- [163] J. M. Pennings and P. Puranam (2001) “Market convergence and firm strategy: New directions for theory and research,” in *ECIS Conference, The Future of Innovation Studies, Eindhoven, Netherlands*, Vol. 20.
- [164] S. Athreye and D. Keeble (2000) “Technological convergence, globalisation and ownership in the UK computer industry,” *Technovation*, Vol. 20, No. 5, pp. 227–245.
- [165] M. Callon, A. Rip, and J. Law (1986) *Mapping the dynamics of science and technology: Sociology of science in the real world*: Springer.
- [166] C. L. Palmer (1999) “Structures and strategies of interdisciplinary science,” *Journal of the Association for Information Science and Technology*, Vol. 50, No. 3, p. 242.
- [167] H. G. Small (1977) “A co-citation model of a scientific specialty: A longitudinal study of collagen research,” *Social studies of science*, Vol. 7, No. 2, pp. 139–166.
- [168] M. Zitt, S. Ramanana-Rahary, and E. Bassecoulard (2005) “Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation,” *Scientometrics*, Vol. 63, No. 2, pp. 373–401.
- [169] C.-S. Curran and J. Leker (2011) “Patent indicators for monitoring convergence—examples from NFF and ICT,” *Technological Forecasting and Social Change*, Vol. 78, No. 2, pp. 256–273.
- [170] Y. Geum, C. Kim, S. Lee, and M.-S. Kim (2012) “Technological convergence of IT and BT: evidence from patent analysis,” *Etri Journal*, Vol. 34, No. 3, pp. 439–449.
- [171] D. R. Swanson (1986) “Undiscovered public knowledge,” *The Library Quarterly*, Vol. 56, No. 2, pp. 103–118.
- [172] D. R. Swanson (1988) “Migraine and magnesium: eleven neglected connections,” *Perspectives in biology and medicine*, Vol. 31, No. 4, pp. 526–557.
- [173] D. R. Swanson (1990) “Somatomedin C and arginine: implicit connections between mutually isolated literatures,” *Perspectives in biology and medicine*, Vol. 33, No. 2, pp. 157–186.

- [174] R. S. Burt (2009) *Structural holes: The social structure of competition*: Harvard university press.
- [175] R. D. Luce and A. D. Perry (1949) “A method of matrix analysis of group structure,” *Psychometrika*, Vol. 14, No. 2, pp. 95–116.
- [176] L. A. Adamic and B. A. Huberman (2000) “Power-law distribution of the world wide web,” *science*, Vol. 287, No. 5461, pp. 2115–2115.
- [177] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman (2001) “Search in power-law networks,” *Physical review E*, Vol. 64, No. 4, p. 046135.
- [178] A. T. Stephen and O. Toubia (2009) “Explaining the power-law degree distribution in a social commerce network,” *Social Networks*, Vol. 31, No. 4, pp. 262–270.
- [179] A. Clauset, C. R. Shalizi, and M. E. Newman (2009) “Power-law distributions in empirical data,” *SIAM review*, Vol. 51, No. 4, pp. 661–703.
- [180] D. Liben-Nowell and J. Kleinberg (2007) “The link-prediction problem for social networks,” *journal of the Association for Information Science and Technology*, Vol. 58, No. 7, pp. 1019–1031.
- [181] M. Al Hasan and M. J. Zaki (2011) “A survey of link prediction in social networks,” in *Social network data analytics*: Springer, pp. 243–275.
- [182] N. Shibata, Y. Kajikawa, and I. Sakata (2012) “Link prediction in citation networks,” *Journal of the American society for information science and technology*, Vol. 63, No. 1, pp. 78–85.
- [183] F. Caviggioli (2016) “Technology fusion: Identification and analysis of the drivers of technology convergence using patent data,” *Technovation*, Vol. 55, pp. 22–32.
- [184] J.-T. . T. Chiang (1992) “Technological spin-off: Its mechanisms and national contexts,” *Technological Forecasting and Social Change*, Vol. 41, No. 4, pp. 365-390.
- [185] “Bringing NASA Technology Down to Earth,” URL: <https://spinoff.nasa.gov/>.
- [186] R. A. Goehlich, C. Blanksby, G. M. Goh, Y. Hatano, B. Pevcnik, and

- J. Wong (2005) “Space spin-offs: Making them known, improving their use,” *Space Policy*, Vol. 21, No. 4, pp. 307-312.
- [187] Y. Park, S. Lee, and S. Lee (2012) “Patent analysis for promoting technology transfer in multi-technology industries: the Korean aerospace industry case,” *The Journal of Technology Transfer*, Vol. 37, No. 3, pp. 355-374.
- [188] K. Sugiyama and M.-Y. Kan (2013) “Exploiting Potential Citation Papers in Scholarly Paper Recommendation,” in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13*, pp. 153–162, New York, NY, USA: ACM.
- [189] M. Igami (2008) “Exploration of the evolution of nanotechnology via mapping of patent applications,” *Scientometrics*, Vol. 77, No. 2, pp. 289–308.
- [190] 元田剛史・村田剛志 (2011) 「頂点間の類似度の足し合わせによるリンク予測精度の改善」, 『人工知能学会論文誌』, 第 26 卷, 第 3 号, 427–439 頁.
- [191] L. Breiman (2001) “Random forests,” *Machine learning*, Vol. 45, No. 1, pp. 5–32.
- [192] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone (1984) “Classification and regression trees. Wadsworth and Brooks,” *Monterey, CA*.
- [193] M. N. Wright and A. Ziegler (2015) “ranger: A fast implementation of random forests for high dimensional data in C++ and R,” *arXiv preprint arXiv:1508.04409*.
- [194] R. C. Team (2016) “R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014.”
- [195] R. Ihaka and R. Gentleman (1996) “R: a language for data analysis and graphics,” *Journal of computational and graphical statistics*, Vol. 5, No. 3, pp. 299–314.
- [196] 秋山高志「ネットワーク分析を用いた組織間関係の形成メカニズムに関する考察 A Study of the Formation Mechanism of Inter-organizational Relationship by the Network Analysis」, 『商学論集』, 第 3 号, 23-442 頁, 1 月.
- [197] R. C. Hsu (1999) *The MIT encyclopedia of the Japanese economy*:

Mit Press.

- [198] M. A. Cusumano and A. Takeishi (1991) “Supplier relations and management: a survey of Japanese, Japanese-transplant, and US auto plants,” *Strategic Management Journal*, Vol. 12, No. 8, pp. 563–588.
- [199] J. H. Dyer (1994) “Dedicated assets: Japan’s manufacturing edge,” *Harvard Business Review*, Vol. 72, No. 6, pp. 174–178.
- [200] R. H. Coase (1937) “The nature of the firm,” *economica*, Vol. 4, No. 16, pp. 386–405.
- [201] O. E. Williamson (1975) “Markets and hierarchies: antitrust analysis and implications,” *New York: The Free Press*.
- [202] J. D. Thompson (1967) *Organizations in action: Social science bases of administrative theory*: Transaction publishers.
- [203] J. Pfeffer and G. R. Salancik (2003) *The external control of organizations: A resource dependence perspective*: Stanford University Press.
- [204] P. DiMaggio and W. W. Powell (1983) “The iron cage revisited: Collective rationality and institutional isomorphism in organizational fields,” *American Sociological Review*, Vol. 48, No. 2, pp. 147–160.
- [205] J. Pfeffer and G. R. Salancik (1978) “The external control of organizations,” *New York*, Vol. 175.
- [206] N. Argyres (1996) “Evidence on the role of firm capabilities in vertical integration decisions,” *Strategic Management Journal*, pp. 129–150.
- [207] M. S. Granovetter (1973) “The strength of weak ties,” *American journal of sociology*, Vol. 78, No. 6, pp. 1360–1380.
- [208] I. Sakata, Y. Kajikawa, Y. Takeda, M. Hashimoto, N. Shibata, and K. Matsushima (2007) “Network dynamics in the twelve regional clusters,” Technical report, RIETI discussion paper series 07-J-023.
- [209] X. Guo, Z. Yuan, and B. Tian (2009) “Supplier selection based on hierarchical potential support vector machine,” *Expert Systems with Applications*, Vol. 36, No. 3, pp. 6978–6985.
- [210] H. Guosheng and Z. Guohong (2008) “Comparison on neural networks and support vector machines in suppliers’ selection,” *Journal of Systems Engineering and Electronics*, Vol. 19, No. 2, pp. 316–320.
- [211] Y. Zuo, Y. Kajikawa, and J. Mori (2016) “Extraction of business re-

- lationships in supply networks using statistical learning theory,” *Helveticum*, Vol. 2, No. 6, p. e00123.
- [212] R. B. Handfield, E. L. Nichols et al. (1999) *Introduction to supply chain management*, Vol. 183: prentice Hall Upper Saddle River, NJ.
- [213] Y. Kim, T. Y. Choi, T. Yan, and K. Dooley (2011) “Structural investigation of supply networks: A social network analysis approach,” *Journal of Operations Management*, Vol. 29, No. 3, pp. 194–211.
- [214] C. W. Autry and S. E. Griffis (2008) “Supply chain capital: the impact of structural and relational linkages on firm execution and innovation,” *Journal of Business Logistics*, Vol. 29, No. 1, pp. 157–173.
- [215] G. Loveman and W. Sengenberger (1991) “The re-emergence of small-scale production: an international comparison,” *Small business economics*, Vol. 3, No. 1, pp. 1–37.
- [216] H. Hoang and B. Antoncic (2003) “Network-based research in entrepreneurship: A critical review,” *Journal of business venturing*, Vol. 18, No. 2, pp. 165–187.
- [217] W. W. Powell, K. W. Koput, and L. Smith-Doerr (1996) “Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology,” *Administrative science quarterly*, pp. 116–145.
- [218] W. W. Powell, D. R. White, K. W. Koput, and J. Owen-Smith (2005) “Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences 1,” *American journal of sociology*, Vol. 110, No. 4, pp. 1132–1205.
- [219] W. Tsai (2001) “Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance,” *Academy of management journal*, Vol. 44, No. 5, pp. 996–1004.
- [220] E. Giuliani and M. Bell (2005) “The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster,” *Research policy*, Vol. 34, No. 1, pp. 47–68.
- [221] H. Lakkaraju, S. H. Bach, and J. Leskovec (2016) “Interpretable decision sets: A joint framework for description and prediction,” in

- Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1675–1684, ACM.
- [222] T. Tanizawa, S. Havlin, and H. E. Stanley (2012) “Robustness of onionlike correlated networks against targeted attacks,” *Physical Review E*, Vol. 85, No. 4, p. 046109.
- [223] A. B. Bernard, A. Moxnes, and Y. U. Saito (2015) “Production networks, geography and firm performance,” Technical report, National Bureau of Economic Research.
- [224] G. Sabidussi (1966) “The centrality index of a graph,” *Psychometrika*, Vol. 31, No. 4.
- [225] L. Freeman (1979) “Centrality in Social Networks I: Conceptual Clarification.,” *Social Networks*, Vol. 1, pp. 215-239.
- [226] P. Bonacich (2007) “Some unique properties of eigenvector centrality,” *Soc. Networks*, Vol. 29, No. 4, pp. 555-564.
- [227] P. Bonacich (1987) “Power and Centrality: A Family of Measures,” *American Journal of Sociology*, Vol. 92, No. 5, pp. 1170-1182, 3.
- [228] L. Page, S. Brin, R. Motwani, and T. Winograd (1999) “The PageRank Citation Ranking: Bringing Order to the Web.,” *Technical Report. Stanford InfoLab*.
- [229] G. Roger and A. N. A. Lu’is (2005) “Cartography of complex networks: modules and universal roles,” *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2005, No. 2.
- [230] J. D. Malley, J. Kruppa, A. Dasgupta, K. G. Malley, and A. Ziegler (2012) “Probability machines: consistent probability estimation using nonparametric learning machines,” *Methods of Information in Medicine*, Vol. 51, No. 1, p. 74.

9 研究業績

9.1 第1章に関連する研究

9.1.1 第1章に関連する査読付き論文

- H. Sasaki, Z. Liu and I. Sakata, Academic landscape of hydropower: a citation-analysis based method and its application, International Journal of Energy Technology and Policy, Vol.12, No.1, pp.84-102, 2016. DOI: 10.1504/IJETP.2016.074493
- T. Visessonchok, M. Sugiyama, H. Sasaki and I. Sakata, Detection and Introduction of Emerging Technologies for Green Buildings in Thailand, International Journal of Energy Technology and Policy, Vol. 12, No. 1, 2016. DOI: <http://dx.doi.org/10.1504/IJETP.2016.074489>
- I. Sakata and H. Sasaki, Bibliometric Analysis of International Collaboration in Wind and Solar Energy, Journal of Sustainable Development of Energy, Water and Environment Systems, Volume 1, Issue 3, pp. 187-198, 2013. DOI: <http://dx.doi.org/10.13044/j.sdewes.2013.01.0014>
- I. Sakata and H. Sasaki, Scientific catch-up in Asian countries: a case for solar cell Natural Resources 4, pp.134-141, 2013. DOI: <http://dx.doi.org/10.4236/nr.2013.41A017>
- H. Sasaki, K. Yamaguchi, G. Yoshizawa, I. Sakata and H. Shiroyama Energy Efficiency Road Mapping in Three Future Scenarios for Lao PDR. , Journal of Sustainable Development of Energy, Water and Environment Systems, Volume 1, Issue 3, pp. 172-18, 2013. DOI: 10.13044/j.sdewes.2013.01.0013

9.1.2 第1章に関連する査読付き国際学会プロシーディングス

- T. Visessonchok, H. Sasaki, and I.Sakata, (2014, July). Detecting emerging technologies and policies recommendation for energy efficiency improvement in Thailand. In Management of Engineering & Technology (PICMET), 2014 Portland International Conference on (pp. 3572-3579). IEEE.
- Z. Liu, H. Sasaki, & I. Sakata, (2014, July). Study on the academic landscape of hydropower: A citation-analysis based method and its application. In Management of Engineering & Technology (PICMET), 2014 Portland International Conference on (pp. 3572-3579). IEEE.
- I. Sakata, H. Sasaki, M. Hashimoto and K. Kayukawa, A parallel running-type growth model in Asia: A case study of clean technology Portland International Conference on Management Engineering and Technology 2014 (PICMET'14), in Kanazawa (27-31, July, 2014)
- H. Sasaki, G. Yoshizawa, H. Shiroyama, I. Sakata and K. Yamaguchi, Energy Efficiency Road mapping in Three Future Scenarios for Lao PDR, Sustainable Development of Energy, Water and Environment Systems (SDEWES2012) in Ohrid, Republic of Macedonia, (1-7, Jul., 2012)
- H. Sasaki, I.Sakata, Y.Kajikawa and S.Yamamoto, The Meta Structure of International Research Alliances in Green Technology 5th International Conference on Project Management (ProMAC2010), in Chiba, Japan, (12-15, Oct., 2010)
- H. Sasaki, I.Sakata, Y.Kajikawa and H.Tomobe, Network of International Intellectual Property Rights Research activity; For national IP policy The 10th Portland International Conference on Management

of Engineering and Technology (PICMET2010), in Phuket, Thailand, (18-22, Jul. , 2010)

- I. Sakata, H. Sasaki, Y.Kajikawa, M.Hashimoto and A.Morita, An Academic Landscape of Patent & Innovation Research for Policy Reform The 5th IEEE International Conference on Management of Innovation and Technology (ICMIT2010)
- H. Sasaki and I.Sakata, A Patent Landscape of Distribution Service Innovation The 19th International Conference for Management of Technology (IAMOT2010), in Cairo, Ezypt, (8-12, Mar., 2010)

9.2 第3章に関連する研究

9.2.1 第3章に関連する査読付き論文

- H. Sasaki, T. Hara and I. Sakata, Identifying Emerging Research Related to Solar Cells Field using a Machine Learning Approach, Journal of Sustainable Development of Energy, Water and Environment Systems, Volume 4, Issue 4, pp 418-429, 2016. DOI: <http://dx.doi.org/10.13044/j.sdewes.2016.04.0032>

9.2.2 第3章に関連する査読付き国際学会プロシーディングス

- H. Sasaki, T. Hara and I. Sakata, Prediction of Promising Papers in Nanocarbon Materials- related Research Using a Citation Network, Portland International Conference on Management Engineering and Technology 2016 (PICMET'16), in Hawaii, USA (4-8, Sep., 2016).
- H. Sasaki, T. Hara, T. Sakaki, J. Mori and I. Sakata, Identifying Embryonic Researches in Renewable Energy Fields using a Scientometric Approach: Case Study of the Solar Cell Field, Sustainable Development of Energy, Water and Environment Systems (SDEWES2015) in

Dubrovnic, Croatia, (27, Sep. - 3, Oct, 2015)

9.3 第4章に関連する研究

9.3.1 第4章に関連する査読付き国際学会プロシーディングス

- H. Sasaki, I Sakata and Y Kajikawa, Serendipitous Identification of Fields Derived from Technology Spillovers from Patent Analysis: Case Study of Material Science, Portland International Conference on Management Engineering and Technology 2015 (PICMET'15), in Portland (2-6, Aug, 2015)
- H. Sasaki, Y. Kajikawa, I. Sakata and V. Ittipanuvat, Predicting the Potential Industrial Fields of Technological Spin-offs by Using IPC in Patent Analysis, Portland International Conference on Management Engineering and Technology 2012(PICMET2012) in Portland (29,Jul.-3,Aug., 2012)

9.3.2 第4章に関連する資金獲得実績

- 文部科学省科学研究費補助金（若手B）
 - － 研究課題名： 研究開発成果の新たな適用先発見に資する、技術スピ
ンオフの予測に関する研究
 - － 代表者区分： 代表者
 - － 獲得金額： 2,990 千円
 - － 研究期間： 2015 年 4 月-2017 年 3 月

9.4 第5章に関連する研究

9.4.1 第5章に関連する査読付き国際学会プロシーディングス

- H. Sasaki and I. Sakata, Prediction of Business Partners Using an n-gram-based Approach that Combines a Network Model and Linear Model of a Supply Chain, Portland International Conference on Man-

agement Engineering and Technology 2016 (PICMET'17), in Portland, USA (8-13, Jul., 2017). Accepted.

- H. Yamano, H. Sasaki and I. Sakata, Metabolism of Inter-Firm Transactions in Regional Network, Portland International Conference on Management Engineering and Technology 2016 (PICMET'17), in Portland, USA (8-13, Jul., 2017). Accepted.
- T. Oka, H. Sasaki, and I. Sakata, Identifying Influencers of Corporate Performance in Interfirm Networks, Portland International Conference on Management Engineering and Technology 2016 (PICMET'17), in Portland, USA (8-13, Jul., 2017). Accepted.
- H. Sasaki, I. Sakata and Y. Kajikawa, Focus on Boundaries of Industrial Network of Kanto in Japan, The International Society for Professional Innovation Management 2012 (ISPIM2012) in Barcelona (17-20, Jun. , 2012)
- H. Sasaki, I. Sakata and Y. Kajikawa Scope of Multilayered Network for Regional Cluster Policy in Japan. The 21th International Conference for Management of Technology (IAMOT2012), in Hsinchu, Taiwan (18-22, Mar. , 2012)
- H. Sasaki, Y. Kajikawa and I. Sakata, Multiple layered Network Structure of Regional Alliances in Japan.-Implication for Regional Cluster Policy-Portland International Conference on Management Engineering and Technology 2011 (PICMET'11), in Portland, USA (31, July- 4, August, 2011)
- H. Sasaki, I.Sakata and Y.Kajikawa,The Structure of Regional Alliances in Nagano Prefecture -Implication for Regional Cluster Project-, 5th International Conference on Project Management (ProMAC2010), in Chiba, Japan, (12-15, Oct., 2010)

9.5 その他の研究業績

9.5.1 その他の査読付き論文

- H. Sasaki, I. Sakata, W. Wangjiraniran and S. Phrakonkham, Appliance Diffusion Model for Energy Efficiency Standards and Labeling Evaluation in the Capital of Lao PDR, Journal of Sustainable Development of Energy, Water and Environment Systems, Volume 3, Issue 3, pp. 269-281. 2016. DOI: <http://dx.doi.org/10.13044/j.sdewes.2015.03.0021>
- I. Sakata, H. Sasaki, H. Nakamura and Y. Kajikawa, Maps of International Research Collaboration in Clean Energy Journal of Energy and Power Engineering, 7, pp480-490, 2013. DOI: <http://www.davidpublishing.com/show.html?11991>
- I. Sakata, H. Sasaki, M. Akiyama, Y. Sawatani and Y. Kajikawa, Bibliometric analysis of service innovation research: identifying knowledge domain and global network of knowledge, Technological Forecasting and Social Change, 80(6), pp. 1085-1093, 2012. DOI: [10.1016/j.techfore.2012.03.009](http://dx.doi.org/10.1016/j.techfore.2012.03.009)
- I. Sakata, H. Sasaki and Y. Kajikawa, Identifying Knowledge Structure of Patent and Innovation Research Journal of Intellectual Property Association of Japan 8(2) 56-67, 2012. DOI: <http://www.ipaj.org/bulletin/backnumber/JIPAJ8-2/p56-67.html>
- 坂田一郎、梶川裕矢、佐々木一、藤末健三、「クリーンエネルギーに関する科学技術協力の構造分析」、開発技術 17, pp67-77, 2011.

9.5.2 その他の査読付き国際学会プロシーディングス

- H. Yamano, H. Sasaki and I. Sakata, Detecting Candidate Combination of Organ-Material- Technology in Regenerative Medicine, Portland International Conference on Management Engineering and Technology 2016 (PICMET'16), in Hawaii, USA (4-8, Sep., 2016).
- H. Sasaki, I. Sakata., Seino, M., & Hashimoto. (2015, September). Off-grid electrification scenarios for rural electrification in Myanmar, Sustainable Development of Energy, Water and Environment Systems (SDEWES2015) in Dubrovnic, Croatia, (27, Sep. - 3, Oct., 2015)
- K. Hayashima. H. Sawamura. Y. Matsumoto, H. Sasaki and I. Sakata, Analysis of Scientific Research in Singapore using Bibliometrics and Network Analysis for Understanding Their Characteristics of R&D: A Case Study of Biomedical Field, The IEEE International Conference on Industrial Engineering and Engineering Management 2014 (IEEM2014), in Malaysia, (9-12, December, 2014).
- H. Sasaki, K Yamaguchi, W Wangjiraniran, S Phrakonkham and I. Sakata, Energy Efficiency Standards and Labeling Forecasting Model in the Capital of Lao PDR, Sustainable Development of Energy, Water and Environment Systems (SDEWES2013), in Dubrovnic, Croatia, (22-27, July., 2013)
- I. Sakata and H. Sasaki, Scientific Catch-up Process in Asian Countries: A Case Study of Solar Cell, IEEE International Conference on Industrial Engineering and Engineering Management 2012 (IEEM2012) in Hong Kong, (10-13, Dec. , 2012)
- I. Sakata, H. Sasaki, T. Inoue, Structure of International Research Collaboration in Wind and Solar Energy, The 2011 IEEE International Conference on Industrial Engineering and Engineering Man-

agement (IEEM2011) in Singapore (6-9 Dec. , 2011)

- I. Sakata, H. Sasaki, H. Nakamura and Y. Kajikawa, Maps of international research collaboration in renewable energy 6th Dubrovnik Conference on Sustainable Development of Energy Water and Environment Systems(SDEWES2011), in Dubrovnic, Croatia (25-29, Sep. , 2011)
- I. Sakata, H. Sasaki, N. Shibata and Y. Kajikawa, Specifying International Research Network Diagram in Growing Field Portland International Conference on Management Engineering and Technology 2011 (PICMET'11), in Portland, USA (31, July- 4, August, 2011)
- I. Sakata, H. Sasaki, M. Akiyama, Y. Sawatani and N. Shibata, Bibliometric analysis of service innovation research Portland International Conference on Management Engineering and Technology 2011 (PICMET'11), in Portland, USA (31st, Jul. - 4th, Aug. , 2011)
- I. Sakata, H. Sasaki, K. Fujisue and Y. Yajikawa Geopolitical structures of international academic collaboration The 20th International Conference for Management of Technology (IAMOT2011), in Miami, USA (10-14, Apr. , 2011)
- I.Sakata, H. Sasaki, Y.Kajikawa, K.Fujisue and A.Morita Systematic identification of academic knowledge in patent & innovation research,IEEE International Conference on Management Science and Information Engineering 2010 (ICMSIE2010) in ZhengZhou, China (17-19, Dec., 2010)
- H. Sasaki and I.Sakata, Detecting the valley of international academic collaboration in renewable energy the 2010 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM2010), in Macao, (7-10, Dec. , 2010).

- H. Nakamura, H. Sasaki, Y. Kajikawa, N. Shibata, I. Sakata, and S. Suzuki, Science and Technology Map Analysis of A Multi-Disciplinary Field -Case Study of Aerospace Engineering- IEEE International Conference on Industrial Engineering and Engineering Management 2010 (IEEM2010), in Macao, (7-10, Dec. , 2010).

9.5.3 その他の査読付き国内学会プロシーディングス等

- 佐々木一、梶川裕矢、坂田一郎「次世代電池に関わる国際的な共同研究の構造」情報知識学会 第 18 回年次大会 (2010)
- 佐々木一、坂田一郎、梶川裕矢、柴田尚樹「サービスイノベーションにおける特許俯瞰」研究・技術計画学会第 24 回年次学術大会 (2009)
- 佐々木一、坂田一郎「公開特許からみたサービスイノベーションと人的能力」しごと能力研究学会 第 2 回全国大会 (2009)
- 佐々木一、坂田一郎、梶川裕矢、友部博教「知的財産権に関する学術研究の俯瞰」第 7 回日本知財学会学術研究発表会 (2009)

9.5.4 その他の査読なし国際会議発表

- H. Sasaki, Y. Kajikawa, and I. Sakata, Detecting knowledge transitions between science and technology for forecasting growing fields. The Institute for Operations Research and the Management Sciences Annual Meeting 2014 (INFORMS 2014) in San Francisco, USA (09-12, Dec., 2014).

9.5.5 その他の寄稿論文

- 佐々木一、森純一郎「ネットワーク構造からみる産学連携仲介機能」産業立地 7月号 (2009)

9.5.6 その他の書籍

- 「5～10年先の市場を獲得するための未来予測による研究開発テーマの決め方」第1章 第2節「機械学習・人工知能による科学技術動向の予測」、2016、技術情報協会
- スティーブ・マントン（著）、屋代菜海（翻訳）、佐々木一（翻訳）、「統合化された知的資産マネジメント—組織の知的資産を活用、保護するためのガイドブック」2007、発明協会
- サーチナ総合研究所（共著）、「中国知的財産権白書—サーチナ中国白書〈2006-2007〉」2006、サーチナ

9.5.7 その他の資金獲得実績

- 公益信託マイクロソフト知的財産研究助成基金
 - － 研究課題名：萌芽的領域の知財情報遷移モデルの特定とその手法の提案に関する研究
 - － 代表者区分：代表者
 - － 獲得金額：1,000 千円
 - － 研究期間：2012年9月-2013年8月
- 文部科学省科学研究費補助金（若手（B））
 - － 研究課題名：イノベーションに資する学術研究領域の特定と知識伝搬に関する研究
 - － 代表者区分：代表者
 - － 獲得金額：1,820 千円
 - － 研究期間：2012年4月-2014年3月

10 謝辞

まずは主査をお引き受け頂きました坂田教授に御礼を申し上げます。指導教官として本研究におけるストーリー構成、結論、各章の全体での位置付けなどについてご指導を頂戴し、おかげでようやくこの形にすることができました。なかなかまとめようとしないうちに私に対して、はやく実績を博士論文にまとめましょうと根気強く根気強く仰って頂いた先生には大変ご面倒をおかけしました。複数のプロジェクトを経験させていただいたなかで、対象分野としても研究手段としても多様な経験をさせていただきました。本研究にそれら全てを盛り込むことはできませんでしたが、知識を獲得するという点でそれらは生きているのだとこの場になって改めて思います。本当に有難うございました。

副査として審査をお引き受け頂きました、技術経営戦略学専攻の影山教授、茂木准教授、森准教授、政策ビジョン研究センターの古月教授に多大なる感謝を申し上げます。御多用の中本当に有難うございました。わたしは学生時代を含めると、技術経営の分野に飛び込んでようやく10年ほどになります。遅々とした進捗ではありましたが、博士論文を通じてようやくひとつの一里塚とすることができました。今後もこの分野の中で研究者として一步踏み出せればと思います。今後ともどうぞよろしくお願い致します。

なお、本研究における第2,3,4章はクラリベイトアナリティクス社による学術情報ならびに特許情報をデータセットとして用いました。また、第5章は株式会社帝国データバンクより貸与いただいた取引データをデータセットとして用いました。両社に心より感謝申し上げます。

以降、学内外問わず、これまでの研究活動を通じて少なからずお世話になった方々へ御礼の意を表します。東工大イノベーション・マネジメント研究科の梶川准教授には、東大以来本当にお世話になりました。今となっては笑い話ですが、実は東大に来て一番最初に衝突したのが梶川先生でした。その後、冗談も言えるまでになりましたが、今考えると当時は言葉や思考の浅さを見抜かれていたんだろうと反省している次第です。東工大MOTへ移られてからも、アポなしで訪問するなど本当にご迷惑をおかけしましたが、常に親身に対応くださいました。外部講演の機会をくださるなど含め、大変良い経験となりました。科研費を含む研究予算申請書では、何が求められているかを見抜き、重複無くひとつのストーリーを紡ぐ事の重要性を教えてくださいました。

イノベーション政策研究センターの石原絢さん、佐藤妙子さんには本当にお世話になりました。石原さんとは一番長い間お世話になったのではないかなと思います。思えば松島研究室の時からになります。公私含めいろいろと他愛もない話をさせていただきました。佐藤さんには経費やプロジェクトの手続き面でいろいろご面倒をおかけしました。政策研究センター職員小林さん、村上さん、佐藤さん方々には沢山のお手間を掛けしました。事務職員の方々には本当に感謝申し上げます。

株式会社データアーティスト山本覚さんとはいつの間にか思いのほか長いおつきあいになってます。たぐいまれなプレゼン能力や人を引きつける力にますます磨きがかかって、勉強させてもらえばなしです。小塩さん、またアイディアソンや面白い事業の話をしましょう。落ち着いたらまたお邪魔しますね。株式会社ホットリンク榎 剛史さんには、年度末かつ小さいお子さんが居てお忙しい中、最後のモデル評価や疑問点などのディスカッションにお付き合い頂きどうもありがとうございました。4人で実施していた定例打ち合わせは、相互に異なる風を吹き込むことが出来、刺激のあるミーティングでした。しばらくご無沙汰ですがまた機会があったらやりましょう。蜂巢健一さんには公私ふくめ大変お世話になりました。ご自身も論文で忙しいにも関わらず一度もそのようなところを見せないところに大きな尊敬の念を抱きます。うちの新キャラが増えた頃にはまたご家族とご一緒させてください。山野泰子さん、暫く一緒にの部屋で研究をさせていただいたなかでご迷惑をおかけしたこと多々あったかと思えます。たまにお茶を入れてくれるのがとてもほっこりしました。ウェディングキャンプにも参加いただきありがとうございました。引き続き研究の議論などもさせていただけると嬉しいです。山下君、磯沼君、には最後にモデルの検証やディスカッション、必要と思われる追加情報についてもコメントをいただきました。この場を借りて感謝申し上げます。

両親へ。美容師の長男として生まれ、紆余曲折ありつつも何を思ったか学問の世界に身をおくことになってそろそろ10年になります。この博士論文は、博士号という学位を取るための最後の論文です。学問の世界において博士号より上の学位は存在しないそうです。一方で、最後の論文と書きましたが、博士号自体は研究者としてのスタート地点としてのひとつの成果です。なので、ようやく研究者としてひとりで歩いていけますということを示すものです。これからようやくスタートです。同時に父親としてのスタートも始まります。不安でいっぱいですが、2人も僕が生まれたときは不安でいっぱいだったんだろうな

あとと思います。お前も親になれば分かる、という言葉の意味がようやく分かり始めてきたかもしれません。おじいちゃん、おばあちゃん2人に、僕の子を見せてあげられるのが楽しみです。いままで色々不安をかけた時期もありました。二人のおかげでここまでこれました。支えてくれてありがとう。健康に気をつけて長生きしてください。これからもよろしく。

翼へ。こんな頼りない兄ですが、おかげで一段落つきました。君の兄は晴れて博士です、良かったですね。父親としては少し先輩になると思うので色々教えてもらうことがあるかもしれません。美容師として一人前になっていく様を見て嬉しいです。挫けそうな時、愛ちゃん、葵ちゃんにも癒やしをもらいました。仲良く素敵な家庭を築いてください。

愛する妻恵理子へ。毎日毎日支えてくれてありがとう。お陰でようやく論文を形にすることが出来ました。共に生活をはじめて2年経ちます。以前は僕自身不規則な生活を過ごしてきたこともあり、一緒に暮らすと決めたときは色々両立できるかあまり自信がありませんでした。今、この論文をひと通り書き終わって思います。恵理子がいなかったら、きっとここまで形にすることは出来ませんでした。日々の暮らし、食事、どこでも笑いが絶えないのが何よりの支えでした。本当にありがとう。これからもよろしく。妊娠がわかってから暫くはあまり実感がありませんでした。自分が本当に親になるのだろうか、なれるのだろうかなど考えていました。以前話しましたが、世の中の親は急に親になったわけじゃなくて、それまでは何も変わらない普通の暮らしをしていたなかで、徐々に親になっていくんだなあと。自分の親もそうだったんだろうと思います。実感がどうであっても、子にとっては唯一の親です。力を合わせて家族を作っていきましょう。これを書いているのは2017年3月29日の昼下がりに。あっと言う間に結婚1周年です。昨日までの寒さが嘘のようで、突然暖かくなりました。鉢植えの桜もきれいに咲きました。沖縄の天気はどうかわかりませんが行ってみてのお楽しみ。よい結婚記念日になるといいですね。ほんの少しだけ筆を休めて結婚記念日を楽しめればと思います。

お義父さん、お義母さん。和音さん、大輔さん。皆様のおかげもあり、博士学位取得にこぎつけました。このタイミングでもろもろが重なり、身重の恵理子さんに関しましては色々ご不安をおかけしました。恵理子さんの支えがとても大きかったので、帰省の際はまたいろいろお世話になってしまいますが、少しの時間親子兄弟水入らずの時間をゆっくり過ごしてください。まだまだご面倒をおかけするかもしれませんがよろしく願いいたします。

最後に、まだ見ぬ我が子へ。きみのお母さんから、きみを宿したかもしれないと聞いたのは2016年12月4日のお昼過ぎだったでしょうか。電話をもらって、すぐに飛んで帰りました。この論文も、きみが生まれるまでには形にしなればという事で筆を進めることができました。きみのおかげです。ありがとう。無事に生まれてくれて健やかに育ってくれることを祈っていますこの文章を書いています。少し物心がついたら色んなお話をしましょう。この世界は楽しいことや、まだ誰も知らないこと、不思議なことが満ち満ちています。生まれたからには、沢山のものに興味を持って、自分が好きなことには目いっぱい取り組んで、この世界を楽しんでください。精一杯ちからになります。君が生まれた瞬間から、ぼくも父親0才です。不手際があると思いますが、どうぞよろしく。沢山思い出を作ろうね。

研究を遂行するにあたっては数え切れないほど多くの方のご協力を頂戴しました。この場を借りて御礼申し上げます。みなさま、本当に有難うございました。