審 査 の 結 果 の 要 旨

氏　名　　赤 﨑　智

本論文は「Early Recognition of Emerging and Disappearing Entities in Microblogs for Entity-based Social Listening Applications（エンティティに基づくソーシャルリスニングのためのマイクロブログにおける新規および消失エンティティの早期認識）」と題し、英文6章から構成されている。ソーシャルメディアのデータを用いた社会や市場の分析（ソーシャルリスニング）を効率的に行うために、分析対象となる事物（エンティティ）の出現および消失を把握することを目的として、マイクロブログからの新規エンティティの即時的発見、消失エンティティの即時的発見、および新規エンティティのタイプ分類について、新たなタスクの設定を行うとともに、それぞれについて高精度な手法を提案し、大規模なTwitterデータを用いた実験により提案手法の有効性を論じている。

第1章は「Introduction （序章）」であり、本論文の背景、取り組む課題、および貢献について概観し、本論文の構成を述べている。

第2章は「Preliminary Knowledge （事前知識）」と題し、本論文におけるエンティティの認識、およびタイプ分類に必要となる基礎的な自然言語処理技術を概説している。

第3章は「Discovery of Emerging Entity （新規エンティティの発見）」と題し、マイクロブログから新規エンティティを早期に発見する新たなタスクを提案し、高精度にこれらを発見する手法を創出している。既存研究において、新規エンティティはWikipedia等の知識ベースに登録されていない事物として定義されることが多く、知識ベースの登録状態に依存した学習を行う必要があり、学習データを自動的に作成することが難しい。本章では先ず、マイクロブログにおいて、ある事物がそれを知らない人がいることを想定して言及される状態にある場合、これを新規エンティティとして定義することで、特定の知識ベースに依らない新しい検出タスクを設定している。その上で、新規エンティティの普及過程においてその新規性を示唆する特徴的な文脈が現れることに着目し、知識ベースおよびマイクロブログの時系列テキストから半教師あり学習を用いて訓練データを生成する time-sensitive distant-supervision（TDS）手法を提案している。実験では、本手法を大規模な英語および日本語の

Twitterデータに適用し、抽出された上位500件の適合率において英語データで73.4%、日本語データで83.2%を達成し、知識ベースを用いた新規エンティティ認識手法によるベースラインを大きく改善できることを示している。さらに、Wikipediaに登録されたエンティティについては、その登録の1年以上前に検出可能であり、Wikipediaに登録されないロングテールなエンティティについても数多く発見できることを示している。

第4章は「Discovery of Disappearing Entity （消失エンティティの発見）」と題し、死没した人物、閉店した店舗など実世界からの消失に関してマイクロブログにおいて言及されているエンティティを消失エンティティとして新たに定義し、これらをマイクロブログから高精度に発見する手法を創出している。既存研究としては人物等、特定のタイプのエンティティについてその発生と消失の時期や期間を認識するものが存在するが、本論文で対象とする多様なエンティティの消失を扱うことはできなかった。本章では、第3章において提案したTDS手法を、消失エンティティが記述される文脈や時期に適合させて改良した手法を提案している。実験では、本手法を大規模な英語および日本語のTwitterデータに適用し、オリジナルのTDS手法を用いたベースラインをF値において2倍以上上回る性能で消失エンティティを発見できることを示している。

第5章は「Typing of Emerging Entity （新規エンティティのタイプ分類）」と題し、第3章において抽出した大量の新規エンティティの分析を容易にするため、人物、施設等のタイプに基づいて新規エンティティを分類する手法を提案している。既存のタイプ分類タスクは既知のエンティティに対応する知識ベースの存在を前提に、1投稿の情報のみを用いて行われることが多く、学習データのない新規エンティティに対して同様の設定で分類を行うのは難しい。本章では、新規エンティティが出現する際に投稿のバーストが良く発生することを利用し、バースト内の複数の投稿を入力とし、新規性を示唆する文脈のみを選択して用いる新規エンティティのタイプ分類手法を提案している。大規模な英語および日本語のTwitterデータを用いた実験により、文脈選択を行わないベースライン手法をF値において大幅に上回る性能を示している。

第6章は「Conclusions and Future Work （結論と今後の課題）」と題し、本研究の成果と今後の研究課題について総括している。

以上、これを要するに本論文は、マイクロブログにおいて新規エンティティを即時的に発見するとともに、そのタイプを分類し、さらに消失したエンティティをも即時的に発見する一連のタスクを新たに提案し、それぞれについて高精度な手法を創出することで、ソーシャルリスニングにおいて分析対象となるエンティティの把握を容易にしており、電子情報学上貢献するところが少なくない。よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。