

Activity-aware Topic Models to Find User Preferences of Activities from Twitter Posts

| | |
|----------|---|
| その他のタイトル | Twitterからのユーザ行動傾向を推定するための行動理解型トピックモデル |
| 学位授与年月日 | 2014-09-26 |
| URL | http://doi.org/10.15083/00007778 |

審査の結果の要旨

氏名 朱 丹丹

朱 丹丹提出の本論文は「Activity-aware Topic Models to Find User Preferences of Activities from Twitter Posts (Twitter からのユーザ行動傾向を推定するための行動理解型トピックモデル)」と題し、全 7 章より構成される。

この論文では、さまざま social media から情報を発信するユーザに焦点を当て、その中で Twitter に着目し、ユーザの発信する情報より、当該ユーザの日常生活行動を通じた行動傾向を抽出するための方法を提案する。

日常生活行動におけるユーザ行動傾向を見つけることは、そのユーザに状況に応じた適切な情報を提供する推薦システム (Recommendation System) 構築のために重要である。一般的に、情報を発信しないユーザの行動傾向を理解し、そのユーザに対して適切な推薦をすることは困難である。当該ユーザのつぶやき、フォーラムへの投稿やブログ等に代表されるユーザ生成コンテンツ (User-Generated Content, UGC) の利用は、そのようなユーザの行動傾向に間接的にアクセスする実用的な方法論となり得ると考えられる。しかしながら、一般に、ユーザ生成コンテンツの膨大かつフォーマットが標準化されていないデータに手作業ベースでアクセスすることは多大な困難を伴うことが明らかであり、数学的モデルを用いた自動的な方法が必要とされる。本論文では、人間の行動に着目する。その理由としては、ユーザ行動は能動的または受動的に維持されており、それらはユーザ自身の潜在的行動傾向により駆動されるものと仮定するためである。ユーザの行動を通じて、その行動傾向を推定することで、推薦システムは、顧客のニーズを満たすより良い情報配信が可能となると考えられる。

当該目的を達成するために、本論文では、新しいモデル開発をする必要があるが、そのベースとなる数理的な手法として、潜在的ディリクレ配分法 (Latent Dirichlet Allocation, LDA) モデルを採用する。LDA モデルは、Tweet をするユーザの内的心理過程をシミュレートするものであり、その広い実現可能性及び容易な拡張性の機能のためのクラスタリングの分野で用いられているものである。

この際、次の 2 つの問題を解決する必要がある。一つは、行動を適切に表現する形式であり、もう一つはある行動を正確に識別するための方法論の構築である。最初の問題に関しては、行動を表現する基本単位として動詞-非動詞 (verb-nonverb) ペアを提案する。第二の問題解決のために、行動を包括する一般的概念としてトピックという概念を提案し、それにより行動理解を促進することを目指す。

ここでは、トピックモデル、行動モデルおよびワードモデルを有する三層構造のモデルを提案する。このような階層型クラスタリング法を適用することにより、ツイート内の各単語は、関係する行動およびトピックを把握するために分析に利用することができ、最終的には、このユーザの行動傾向として抽出できる。

提案システムは、データマイニング部と、三層モデルクラスタという二つの主要な機能モジュールから構成される。データマイニングは、行動を表現する動詞-非動詞の選択とノイズワードの除去を行う。クラスタリング部では、2 つの新しいトピックモデルが提案されている。wpLDA モデル (word-pair LDA モデル) は、クラスタに対応するトピック層と行動を生成するために用いる。TLCG モデル (Tri-Layer Cluster Generative モデル) では、wpLDA モデルからの出力を、三層クラスタを生成するための入力として利用する。

まず、wpLDA モデルと通常の LDA モデルを比較し、主観的評価を行うことで、wpLDA モデルがより直観的に訂正であることを示している。

また、wpLDA モデルへの入力すなわち単語ペアの選択法として、任意の動詞-非動詞ペアを抽出するのではなく単語間の相互連関を利用した手法を提案している。confidence ベース (Confidence-based, C-based)、dependency ベース (Dependency-based, D-based)、dependency confidence ベース (Dependency Confidence-based, DC-based) の三種類を比較し、DC-based が最も適切な結果を得たことを示している。

さらに総合的な実験結果より、提案されたシステムがクラスタリングや行動からのユーザの行動傾向推定をうまく実現できていることを示している。これらの結果より、提案システムにより、Twitter 情報からユーザ

行動傾向を適切に推定できることがわかった。

これらの成果により、本論文は博士（工学）の学位請求論文として合格と認められる。