

Twitter と論文との自動対応付け

Automatic Alignment of Tweets with Research Papers

中橋光 難波英嗣 竹澤寿幸
Hikaru Nakahashi Hidetsugu Nanba Toshiyuki Takezawa

広島市立大学大学院 情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

1. はじめに

論文間の参照関係はこれまで様々な目的に利用されている。そのひとつに、ある論文(被参照論文)を参照する複数の論文の中で被参照論文について言及した個所(参照個所)を抽出し、まとめることで、被参照論文の読解支援をする、という試みがある。しかし、この方法は、その論文が参照されるまでは適用できないという問題点がある。

そこで、本研究では、Twitter に着目する。Twitter は、近年、急速に普及が進んでいるコミュニケーションツールである。Twitter には、ユーザの日常生活に関するツイートだけでなく、学会等で発表された論文への意見や疑問を投げかけるツイートや論文の評価に関するも数多く投稿されている。学会で発表されている論文とこのようなツイートを対応付けることができれば、その論文に対するユーザの評価、意見、議論など論文からだけでは得ることができない多様な情報を得ることができると考えられる。

本研究では論文とツイートを自動的に対応付ける手法を提案する。Twitter はリアルタイム性に優れているため、学会終了直後にも評価等を知ることができる。ここで、ツイートの中には、意見や評判など論文の理解に役立つもの他に、発表の様子を実況中継するツイートなど論文理解にはあまり貢献しないものも存在する。そこで、本研究では、ツイートをいくつかのタイプに自動分類し、有益なツイートの検出を目指す。

2. 提案手法

2.1. ツイートと論文の対応付け

本研究では、タイプ分類の手法として、機械学習に基づく手法を提案する。提案手法の流れを以下に示す。

- ① ツイートの投稿時間と論文の発表時間を用いて、ツイートに対応する論文の候補を選定する。
- ② 対象のツイートと①で選定した論文間の類似度を算出する。類似性尺度として、DP マッチングを用いる。
- ③ ツイートに対して①で選定した論文の中から類似度が最も高い論文をそのツイートに対応付ける。

また、ユーザは学会において複数回ツイートすることに着目する。本研究では、類似度を測る際に用いるツイートから、前後 20 分に投稿された同一ユーザのツイートを追加して類似度を測る手法を行う。この手法を「ユーザ情報あり」と呼ぶ。

2.2. ツイートのタイプ分類

本研究では、タイプ分類の手法として、機械学習に基づく手法を提案する。本研究では以下の 4 つのタイプに分類を行う。

1. type S: 論文について評価しているツイート
2. type B: 論文について実況中継しているツイート

3. type D: 論文について議論しているツイート

4. type O: その他のツイート

機械学習に用いた素性を以下に述べる。

素性: 評価表現辞書[1], 手がかり語, 論文との類似度, 文タイプ[2], リツイートの有無

3. 実験

実験には、学会に関する全 2,269 ツイート、及び 134 件の論文を用いた。評価尺度には、精度、再現率、F 値を用いた。対応付けでは、提案手法として、2.1 節で述べた類似性尺度に DP マッチングを用いた場合と DP マッチングにユーザ情報を追加した場合を用いる。ベースラインには、ツイートと発表時間により対応付けを行う。対応付けの実験結果を表 1 に示す。

表 1: ツイートと論文の対応付けの実験結果

	精度	再現率	F 値
DP マッチング (ユーザ情報あり)	0.508	0.485	0.496
DP マッチング	0.494	0.466	0.480
ベースライン	0.458	0.388	0.420

実験の結果、DP マッチング(ユーザ情報あり)でベースラインに比べて F 値が 0.076 向上した。

タイプ分類では、提案手法として、2.2 節で述べた手法を用いる。ベースラインでは、全単語の出現頻度を素性として機械学習を行う。機械学習には TinySVM を用いる。タイプ分類の結果を表 2 に示す。

表 2: ツイートのタイプ分類の実験結果

	精度	再現率	F 値
提案手法	0.578	0.453	0.506
ベースライン	0.582	0.439	0.500

実験の結果、提案手法がベースラインに比べて F 値 0.006 向上した。

4. おわりに

本研究では、Twitter と学会の発表論文を用いて、対応付けとタイプ分類の手法を提案した。実験の結果、DP マッチングのユーザ情報を追加した場合で比較手法に比べて精度が F 値 0.076 向上し提案手法の有効性を示した。タイプ分類では、提案手法が比較手法に比べて F 値 0.006 向上し提案手法の有効性を示すことができた。

参考文献

- [1] 鍛冶伸裕, 喜連川優, “自動構築した評価文コーパスからの評価表現辞書の構築,” 日本データベース学会 Letters, Vol.6, No.1, pp.41-44, 2007.
- [2] 横山憲司, 難波英嗣, 奥村学, “Support Vector Machine を用いた談話構造解析” 情報処理学会研究報告.自然言語処理, 2003(23), pp.193-200, 2003.

発表論文

中橋光, 難波英嗣, 竹澤寿幸, 高須淳宏: Twitter と論文との自動対応付け, Soc2013, 2013.