

ツイートテキストからの Q&A 型知識の抽出

Extraction of Q&A from Tweet Texts

中渡瀬 秀一¹⁾, 大山 敬三¹⁾²⁾

NAKAWATASE Hidekazu¹⁾, OYAMA Keizo¹⁾²⁾

1) 国立情報学研究所, 2) 総合研究大学院大学

1) National Institute of Informatics, 2) The Graduate University for Advanced Studies

キーワード: Q&A 型知識, ミニブログ, SNS

1. 研究の背景と目的

近年, ミニブログ (またはマイクロブログ) と呼ばれる短文メッセージを投稿・共有する Web サービスが世界中に広く普及している. 中でも世界最大のミニブログサイトである Twitter における投稿数の増加が著しく, 2012 年 6 月 7 日現在, 1 日に 4 億件以上のメッセージが投稿されている (うち日本語の投稿は約 4000 万件). このような背景を受け, このデータ資源を活用しようとする研究も盛んである. それらでは Tweet 中に含まれる単語や属性値を集計することによって Tweet の分類・ランキング・分析・推薦や現実事象の予測, 事件・イベントの検知などを試みてきた. しかし個々の Tweet は内容が希薄なものが多いため知識の抽出対象には適していない.

そこで本研究では複数 Tweet のセットを対象にして Q&A 型の知識を抽出することを目的とした. ミニブログ中にも QA の対話は存在し, そこには日常生活におけるニーズも表現されている. それらは質問者だけでなく第三者にとっても有用な知識であろう. したがって, この質問と回答を知識として抽出できるようになれば, それらは有益な知識源となることが期待される.

2. 研究内容

本研究では, ツイートデータからの Q&A 型知識抽出 (絞り込み) 実験とその結果に対する評価, そして得られた質問の分類を行った. 抽出実験においてはデータ取得日に日本語で投稿された Twitter の全ツイート (24 時間分) を対象に Q&A 型知識となる候補 (質問と回答のペア候補) の絞り込みを行った. 次に絞り込み結果から評価用サンプルを抽出し, これらに対して Q&A の出現状況調査 (質問の出現数, 回答の対応状況・Q&A の有用度) を行い評価した. 最後に得られた Q&A に対して先行研究に基づく, 質問意図を基準にしたタイプ分類を行った. 本発表ではさらに Yahoo! 知恵袋 (QA サイト) との比較や Twitter 中における知識の特徴についても説明する.

3. まとめ

本研究の結果, 以下の結果が得られた.

- ツイートデータから Q&A 候補を絞り込むために, ツイートとその返信 (URL を含む) のペアを対象に抽出したところ, その中の約 15% が Q&A であり, それら Q&A のうち約 48% が有用な質問であった. さらに有用な質問の約 65% には回答が投稿されていた.
- 質問タイプによる分類の結果からツイートでは事物の名称, 目的実現のための手段や方法を問う質問が多く, 意見や体験を尋ねる質問は少数であることが判明した.