

質問回答サイトにおける質問タイプの分類に関する研究

渡邊 直人

近年、Yahoo!知恵袋や教えて！goo など、質問回答サイトの利用が増加している。そこでは、日常の疑問や困りごとなどを質問として投稿し、それに対して別の利用者が自身の知識や経験、価値観から回答を投稿することで利用者に役立っている。多くの質問回答サイトでは、質問の内容に応じたカテゴリが用意されており、回答者は自身の興味あるカテゴリから、興味や回答可能かどうかなどによって質問を選択し、回答を行う。しかし、同一カテゴリ内でも、知識や経験を有していないと回答できない質問から、主観的な意見を述べるの質問までが混在している。この混在は、回答者による質問の選択を困難にし、質問者が多くの回答を期待しても得られる回答の数が僅かであったり、専門的な知識を問う質問にも関わらず、主観的な回答がされてしまうといった問題を生じさせている。

本研究では、質問回答サイトにおける回答者の質問選択を容易にすることを目的とし、新たな質問タイプを提案し、その分類方法を検討する。質問者が回答に期待するタイプを質問タイプと定義し、従来のカテゴリ分類と組み合わせることで、投稿された質問のさらなる細分化を行い、回答者が自身が回答できる、あるいは回答をしたいと思う質問の選択を容易にすることを目指す。

提案する質問タイプは、回答の正解の有無と回答の性質から、factoid, non-factoid, 経験, 提案, 意見, その他の 6 タイプである。質問から抽出した品詞別の形態素とその組み合わせを素性として機械学習を行う。また筆者が経験によって得た質問タイプの分類に有効な語を素性でも機械学習を行い、質問を分類する実験を行った。

一方、タイプ分類については、Yahoo!知恵袋の質問 400 件を 2 人の判定者がそれぞれを手作業で分類し、その一致度を算出した。その結果、中程度以上の一致が認められ、質問回答サイトの質問を提案する質問タイプに分類することが可能であることが判明した。機械学習による分類では、経験則から選択した語を素性として用いた場合、品詞別に抽出した語の組み合わせを素性とした場合と比較し、精度と再現率から算出した F 値の平均値が、2 倍程度高くなるという結果を得られた。以上の結果から、質問タイプはある特定の単語によって特徴付けられるという知見を得た。

今後の課題として、さらなる分類性能の向上のために学習データの追加、経験則による単語の選択方法の定式化や単語の追加が挙げられる。

(指導教員 佐藤 哲司)