

理工学専攻修士論文要旨

提出年度：2007年度
提出日：2008年1月21日
専修コース：知能情報コース
学生番号：3567056
学生氏名：伊藤美帆
研究指導教員：原田実

(論文題目)

評価表現の自動抽出

(内容の要旨)

近年、掲示板、Weblog、口コミサイトなど Web 上で製品や商品の評価や意見を簡単に知ることが出来る様になった。これにより人々は購入する製品や商品を詮索する際の一判断材料として貴重な情報となりえる。これらの情報は Web 上に膨大な数存在しており、実際にどの商品が自分にとって一番良い商品なのかを調べるのは大変労力のいる作業となってしまふ。このためこのような評判情報を如何に迅速かつ正確に取得するかがカギとなっている。

そこで本研究では、こうした Web 上の情報から評価表現を抽出する EEE(Evaluation Expression Extraction)System を研究・開発した。EEE-System は共起関係や単語の意味を解析することにより、評価表現を正確に抽出することを可能にしている。EEE は評価表現を抽出する機能と文の評価を分類する 2 つの機能に大別でき、それぞれ EEE-Evidence, EEE-Concept としている。

EEE-Evidence では、“良い”、“満足”、“おいしい”、“悪い”、“嫌い”、“最低だ”等の単純に文の評価を決定すると思われる確証表現を抽出する。例えば、「値段が安いので学生の私にはうれしいです。」という文が存在した場合には、「うれしい」という確証表現を抽出し、うれしい根拠である「値段が安い」という情報を評価表現として抽出する。また確証表現に、(* ^ - ^ *), (^ - ^), (笑), ♪, (T o T), (>_<.), (泣)といった顔文字を含めることにより、上記の確証表現で拾い切れないが評価に値する表現を抽出し、評価表現辞書を補強する。これは「リップクリームなのに、口紅っぽい容器も可愛かったです (^ - ^) /」の様に、顔文字出現する直前には評価となる表現が含まれることが多いという統計的な判断からである。

EEE-Concept の機能は、アンケートデータから評価語、属性、対象の文節を抽出する。肯定的評価・否定的評価を決定付ける評価語の概念の上位にある概念を手動で選択する。抽出した評価語、属性、対象、評価語の修飾の各要素の文節品詞、上位概念の有無、文のムード、顔文字の有無を SVM で学習する。実際に分類を行う際には EEE-E で構築した評判辞書を使用することにより、分類精度を向上させることが可能となっている。

実装した自動辞書 EEE-Concept・EEE-Evidence によるシソーラスの充実により、新語登録や更新の簡略・簡便化、記号や絵文字を含む若者系コトバ、文節のムード、共起する言葉の大幅な曖昧性解消、等々に可能なかぎりの解決を見、本 EEE システムにより、ラベル付きデータを学習に組み込む教師あり学習手法をもって、従来より、粒度の高い評価極性抽出法の開発をなした。

E-Questionnaire から収集したコスメデータに対するラベル付き評価極性抽出は、確証表現を用いた正解率が 75.0%、顔文字を用いた正解率が 66.7%、の各正解率を得られることがわかった。