

評価文の意味的類似度を用いた テキストマイニングシステム STM の評判分析の精度向上

外山 敏彦†, 原田 実‡, 松村 冬子‡

青山学院大学大学院理工学研究科知能情報コース†,
青山学院大学理工学部情報テクノロジー学科‡

c5614133@aoyama.jp

概要 意味解析システム SAGE を用いたテキストマイニングシステム STM には評判分析機能が搭載されている。従来の STM では同じ述部（評価語）を持つ評価文でも、文脈によって Good 評価と Bad 評価の極性を適切に判別することが難しかった。そこで、同一の評価語でも評価が正しく分かれるように、評価語だけでなくその対象、属性など文脈も追加した評価文を評価表現辞書に登録し、新たな極性決定法として辞書に登録された評価文と意見文との意味的類似度をもとに評価極性を決定するようにシステムを改良した。

キーワード 評価表現, 評価語, 意味的類似度, 意味解析, 評価表現辞書

1 背景・研究目的

原田研究室では、意味解析システム SAGE[1][2]を用いた STM[3][4]というテキストマイニングシステムを数年に渡って研究している。STM は主として、文を意味解析した結果における文間の意味的類似度を元にクラスタリングする機能を中心としたテキストマイニングシステムである。STM の中には、評判分析機能が組み込まれている。これは意見文の中に評価表現があった場合に抽出する機能である。従来の STM では、評価表現辞書に登録された語と、その表記および語意が一致するものが意見文に存在していれば、それを評価表現として分類していた。しかし、同じ評価語でも文脈によっては極性が Good や Bad の評価に分かれることがある。たとえば、「安全性が高い」と「リスクが高い」では前者が Good, 後者が Bad となる。このような、同一評価語に対する適切な極性の判別を実現するために、評価表現辞書として単に評価語を登録するのではなく、評価語に係る対象、属性である語を組み合わせた短文（これを評価文と呼ぶ）を評価辞書に登録することによって、意見文に現れた評価表現が、辞書のどの評価文と意味的に最も類似しているかを調べ、その評価表現辞書における極性を、当該意見文の極性とするようにシステムを改良した。

2 STM における意味的類似度

STM では、SAGE を用いて 2 つの文を意味解析した結果の意味グラフ間の各頂点およびその辺間の意味的類似度を計算し、閾値以上の類似度からなる頂点および辺の数を数えることによって、類似共通部分グラフの

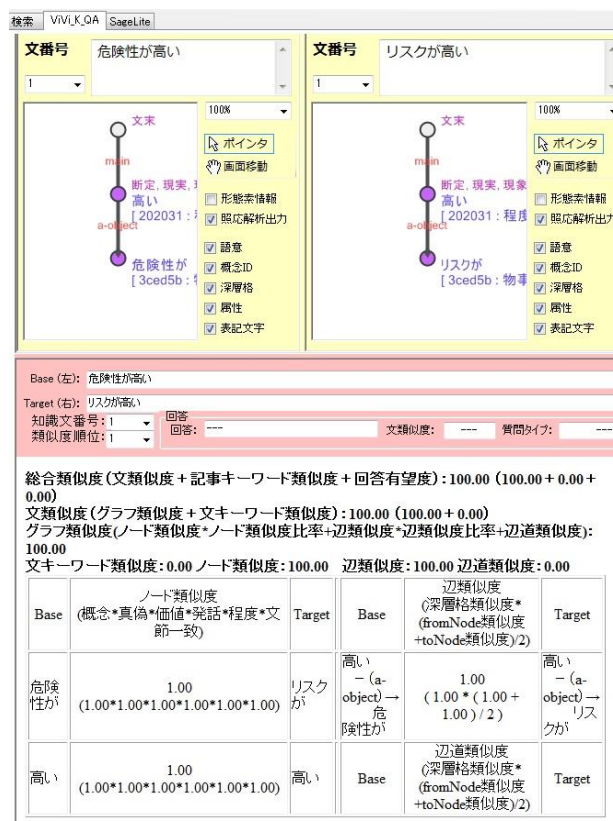


図1 「危険性が高い」と「リスクが高い」の類似度

大きさを求め、元の 2 つの意味グラフの大きさで割ることによって 2 つの文の意味的類似度を定義する。例えば「危険性が高い」と「リスクが高い」の類似度は図1のようになり、2 頂点および 1 辺の類似度が 1.0 であり、この結果グラフ類似度が 100 となっている。このようにして文中の各評価表現に最も類似している評価表現辞書中の Good, Bad, Neutral の評価極性を持つそれぞれの評価

文との類似度を求め、その合計を元にそれがどの表現であるかを分類する線形分類器を、WEKA の Support Vector Machine(SVM)を用いて作成した。

この分類器を元に、例えば、図2に示すように10個の入力文に評価分析を行うと、同じ評価語である「高い」や「大きい」に関しても文脈に応じて、正しい評価極性が付与されることが分かった。

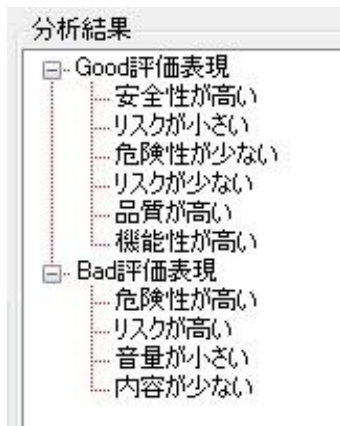


図2 10個の評価文に対する評価分析

3 評価表現辞書

評価表現辞書は、実際には評価語として 3,968 個のエントリを持つハッシュテーブルとして内部的に展開する。各評価語をエントリとしたハッシュテーブルの値としては、同じ評価語を持つ評価文の意味解析結果である意味グラフのリストを登録する。改良前は評価語のみを用いた辞書であったが、改良後は評価語に係る対象、属性である語を組み合わせた評価文を登録している。現在、この評価文の登録数は合計 14,131 個である。

4 おわりに

新しい評価極性決定法をもとに評価実験を行った。目的としては改良前と改良後との評価精度の比較を行うためである。レビューデータは、楽天みんなのレビュー、アットコスメ、Amazon などから SECRET[5]で抽出した 700 件を対象とする。評価が Good, Bad, Neutral のいずれであるかは人手で判断した正解データとの比較で行う。

4.1 評価例

ここでは、改良された事例として、「電池の交換頻度が減り、使えます。」という文を挙げる。改良前の STM では図3のように Bad 評価として判別されているが、改良後の STM では図4のように正しく Good 評価として判別されていることが分かる。これは、従来では「減る」という評価語から意見文が Bad 表現とされていたが、改良後の評価極性決定法において「交換頻度が減る」という Good 評価として登録した辞書中の評価文との高い類似度によって、極性を正しく抽出できたためである。

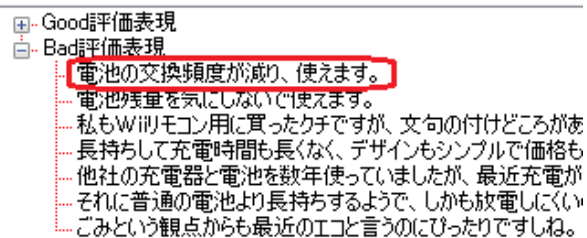


図3 改良前の STM の評価例 (Bad 評価)

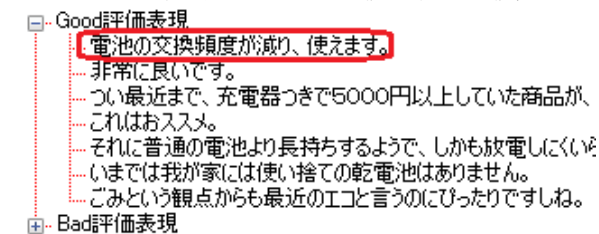


図4 改良後の STM の評価例 (Good 評価)

4.2 実験結果

改良後のシステムの実験結果は表1のようになり、表2の改良前のシステムと比べても再現率、適合率ともに上昇し、全体的な精度も上昇したことが見てとれる。

表1. 改良後の評価表現の判別精度

	Good	Bad	Neutral
再現率	0.898	0.877	0.891
適合率	0.922	0.716	0.908
F 値	0.910	0.788	0.900

表2. 改良前の評価表現の判別精度

	Good	Bad	Neutral
再現率	0.789	0.736	0.845
適合率	0.840	0.714	0.820
F 値	0.814	0.725	0.832

5 参考文献

- [1] 原田実,水野高宏: “EDR を用いた日本語意味解析システム SAGE ”, 人工知能学会論文誌, Vol.16, No.1, pp.85-93 (2001.1).
- [2] 原田実, 田淵和幸, 大野博之, "日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価", 情報処理学会論文誌, Vol.43, No.9, pp.2894-2902, (2002.9)
- [3] 西脇剛,保立哲志,原田実: “意味解析に基づくテキストマイニングシステム STM”,情報処理学会第 69 回全国大会論文集,2C-03,第 2 分冊 pp. 89-90. (2007.3).
- [4] 原田実, 泉祥太, 関谷瑠紀:“評価表現の自動抽出とテキストマイニングシステム STM への組み込み”,情報処理学会第 73 回全国大会論文集,1D-2, (2011.3).
- [5] 今井和雄,吉村友希,原田実:“Web レビューの表形式化システム SECRET の開発”,情報処理学会研究報告, Vol.2010-NL-196, No.4, pp. 1-8, (2010.5).