

自己中心性に基づく質問回答サイトにおける意思疎通のズレの表現

鈴木信雄[†] 津田和彦[‡][†]KDDI 株式会社[‡]筑波大学大学院 ビジネス科学研究科

1.はじめに

近年、インターネットの急激な普及と共に、様々な課題を相互に解決し合う質問回答サイトが多く利用されてきている。このようなサイトは、課題解決の場であると共に、有望な集合知の集積手段であるので、計算機処理を用いた解析による知識の抽出が期待されている。しかし、多数の発言者が同じサイト上で発言するため、必ずしも良好な意思疎通が行われているとは言い難い。質問者からの意図とは外れた意味の無い発言が繰り返されるために、知識の自動抽出が困難となると共に、サイト自体の運営も危機に晒される場合もある。そのため、本研究では、人間の自己中心性を質問回答サイトの発言に見られる言語表現から抽出し、それに基づいて発言の意思疎通のズレを抽出することを試みた。本研究では、文章の意味を扱わずに表層的な言語表現によって意思疎通のズレを表現可能であることを示す。具体的には、先の研究にて求めた言語表現から自己中心性を推定する手法を用いて、対話形式を取る質問回答サイトのテキストデータにおける意思疎通のズレを推定する。

自己中心性に着目した研究としては、Weblog に関するものがいくつか行われている。沼らは、単語の類似度と FOAF に基づく距離の自分からの近さを自己中心性と意味づけ、情報検索手法を提案している^[2]。また、松岡らは、Weblog からの自分を中心としたメタデータやトラックバックなどのリンク情報を基に実際の人間関係をモデル化している^[3]。これに対して、本研究では、自己中心性をサイト間の関係ではなく、実際の人間の情緒的な性質と位置づけ、その言語的特徴により対話のズレを抽出するものである。一方、質問回答サイトのテキストデータから対話の特徴を分析した研究としては、荒牧らの呼応表現や質的関連語を用いたコメント間の対応関係を抽出するものがある^[4]。この研究では、コンテンツではなくメタデータのみを使って記事(発言)間の構造的関係性を抽出している。これに対して、本研究では、記事のコンテンツを解析することにより、対話中における意思疎通のズレという意味的關係性を抽出している。

2.自己中心性の表現

自己中心性は、人間の基本的で情緒的な性質であり、多くのサイトのテキスト上に現れる。先の研究では、このような言語表現を多数収集し、n グラムの形態素より自己中心性の分類と推定手法を提案した^[1]。本研究では、これを利用して、質問回答サイトにおける意思疎通のズレを推定する。そのために、まず、自己中心性の各分類に対する強度を示す重み付けを表 1 のように定義した。

The expression method for misunderstanding on QA sites based on egocentrism

SUZUKI Nobuo[†] and TSUDA Kazuhiko[‡]

[†]KDDI Corporation

[‡]Graduate School of Business Sciences, University of Tsukuba

表 1 自己中心性分類と重み付け

分類	細分類	重み付け
他者攻撃	不満	12
	軽蔑	11
共感性の低さ	突き放し	10
	皮肉	9
自己都合優先	開き直り	8
	制限	7
自己正当性	断定	6
	押し付け	5
	言い訳	4
自益的帰属	要求	3
推測	推測	2
	質問	1
同意	お礼	0

ここで、先の研究に加えて次のような追加の検討を行った。先の研究では、「同意」の分類は含まれていなかった。しかし、対話において同意は特に重要な意味を持ち、意思疎通の良好さを表すことから、今回新たに追加した。また、先の研究では、「質問」に分類される言語表現は、単に疑問符を手がかりに分類していた。しかし、意思疎通においては、単純な質問だけでなく、質問の形を使って何らかの意見を示す場合がある。そのため、質問文の名詞が回答文の中に繰り返されている場合には、用語の説明を要求するような単なる質問として、本解析の対象から除外する処理を行った。

次に、発言単位の自己中心性の強度を次の手法により求める。まず、質問回答サイトのテキストデータを取得し、形態素解析を行う。一つの発言は複数の文から構成されており、発言の最終文が自己中心性の言語表現であれば、その自己中心性の強度を l と置く。これは、多くの発言においては、最終の文が発言全体を表現する場合が多いことによる。また、「同意」については、当該発言中に一つでも該当する言語表現があれば、その発言は「同意」とすることとした。これは、発言中に一度同意が存在すれば、それ以降は補足説明であることが多いことによる。一方、最終文が自己中心性の言語表現ではない場合、当該発言中にある N 個の自己中心性の言語表現を含む文について、各自己中心性の強度 x_i の平均を求める。この l と平均の大きい方の値を、その発言を代表する自己中心性の強度 S とし、式(1)に示すように定義する。

$$S = \max(l, \frac{\sum x_i}{N}) \quad (1)$$

3.対話における意思疎通のズレの表現

ここでは、前章で求めた発言毎の自己中心性の強度を用いて質問回答サイトの発言における意思疎通のズレを

表現する手法について述べる。

まず、前章で求めた発言毎の自己中心性の強度を各質問回答の一連の記事(スレッド)についてグラフ上にプロットする。このグラフの形状の分類により、対話のズレの推定が可能となる。具体的には、横軸に発言の回数を、縦軸に当該発言の自己中心性の強度をプロットする。このようにして作成したグラフは、以下のような4つのパターンに分けることができる。

(1)最終的な発言の強度がゼロ

質問に対して適切な回答が得られ、最終的に質問者の満足が得られている。これは、対話における意思疎通のズレが小さいことを示している(図1)。

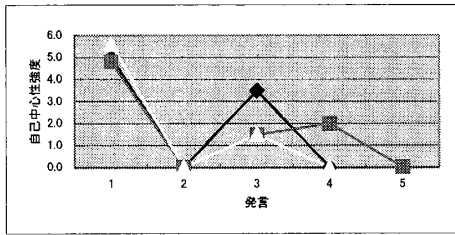


図1 最終的な発言の強度がゼロの例

(2)最終直前の発言の強度がゼロ

質問の回答に対して同意が図られ、その後、補足説明が行われている。これも、意思疎通のズレが小さいことを示している(図2)。

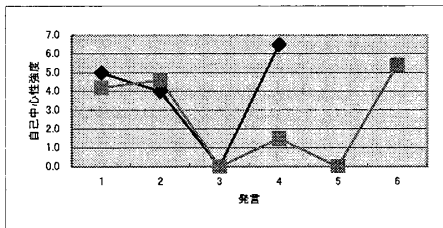


図2 最終直前の発言の強度がゼロの例

(3)全体的に強度が高止まり

質問に対する回答に同意が得られず、お互いに一方的な発言が多い。そのため、意思疎通のズレが大きいことを示している。具体的には、強度5以上を指す(図3)。

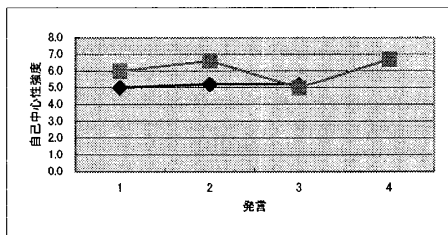


図3 全体的に強度が高止まりの例

(4)その他

質問に対する回答に同意が得られず、一方的な発言、情報の補足、提案などが行われている。そのため、意思疎通のズレが大きいことを示している(図4)。

4. 評価実験

これまで述べた手法を実際のデータで検証するために、携帯電話に関する質問回答サイトから 1,585 スレッドの

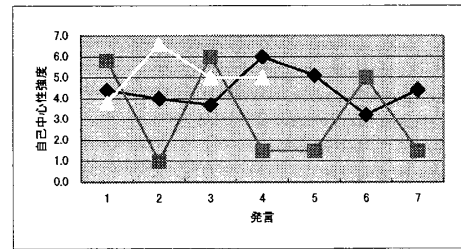


図4 その他の発言の例

テキストデータを収集した。この中で、自己中心性の言語表現を持つ発言が3回以上現れるスレッドは41あった。これらのデータを用いて、意思疎通のズレを調査したところ、全てのスレッドにおいて、今回の手法にて推定した結果と同一であることが確認できた。表2に実際の発言の例を示す。

表2 実際の発言例

分類	発言内容	判定
最終的な発言の強度がゼロ	<ul style="list-style-type: none"> ・ホワイトとピンクで迷ってるんですが,43Hと43HIIに違いはありますか? ・値段的な変わりも機能も変わりません。【押し付け】 ・回答ありがとうございます【お礼】 ・値段も変わりませんし,性能もかわりません。【押し付け】 ・気に入った色を選ぶのが一番ですが,私は白がオススメです。 ・お答えありがとうございます。【お礼】 	ズレ小
全体的に高止まり	<ul style="list-style-type: none"> ・着うたフル 350 円は高いです。音楽だけなら 100 円のサイトをたまに見つけますが着うたフルは見つからないです。無いのでしょうか? ・着うたフルだと着信音にすることもできないといけなないので,その手間を考えると自然です。【断定】 ・CD などの音源を所持しているならば合法ですが,そうでない場合は違法行為になるので注意が必要です。【断定】 ・機種によっては再生できないものもあるのであしからず。【突き放し】 	ズレ大

5. おわりに

本稿では、自己中心性の分類と強度から、質問回答サイトの対話における意思疎通のズレを推定する手法を提案した。本方式は、対話文の深いコンテキストを解析することなく、表層的な言語表現から推定するため、高速処理が可能であり、サイトの監視などに応用することが期待できる。今後は、より多くのサイトにおける評価を実施し、推定精度を向上させる予定である。

参考文献

[1]鈴木信雄,津田和彦, "n-gram を用いた自己中心性の推定手法", 情報処理学会 第 70 回全国大会, 3D-4 (2008)
 [2]沼見介,大向一輝,濱崎雅弘,武田英明, "Weblog におけるエゴセントリック検索の提案と実装", 人工知能学会研究会 SIG-SWO-A401-06 (2004)
 [3]松岡寿延,瀬下仁志,岡野真一,荒川則泰,加藤泰久, "エゴセントリックコミュニケーション指向の Weblog 分析", 信学技報 KBSE2004-57 (2004)
 [4]荒牧英治,阿辺川武,村上陽平,瀬本明代, "BBS 対話における発言間の応答関係の判定", 言語処理学会 第 14 回年次大会 (2008)