

発話意図を用いた人同士の関与の強さの推定

Relationship Activity Estimation by Using Utterance Intentions

西原陽子^{1*} 砂山渡² 谷内田正彦¹

Yoko NISHIHARA¹, Wataru SUNAYAMA² and Masahiko YACHIDA¹

¹ 大阪大学大学院基礎工学研究科, ² 広島市立大学情報科学部

¹ Graduate School of Engineering Science, Osaka Univ., ² Faculty of Information Science, Hiroshima City Univ.

1 まえがき

電子メールや掲示板などネットワーク上の連絡手段が発達し、遠く離れた所に住んでいる人と知り合う機会が増えた。ネットワーク上での人間関係は目で把握しにくい。この見えにくい対人関係を計算機によって視覚化させる動きもある [2, 3] が、多くの手法では関係の種類も強さも全て同一と見なしている [4, 5]。

対人関係における関与の強さは正しく表示しなくてはならない。自分自身の対人関係であれば可能かもしれないが、自分の友人の対人関係となると全てを把握することはできない。だが、人は友人の対人関係の関与の強さを友人同士の会話から推定できる。我々は会話の中身を解析することで関与の強さを推定しようと考えた。

本稿では会話中の発話文の発話意図を推定し、発話意図を用いて2人の人間の関与の強さ Relationship Activity (RA) を推定する手法を提案する。発話文から発話意図を機械的に知ることは難しいが、文末表現から意図の一部が取れることが確認されている [1]。本研究でも文末表現を用いて発話意図を推定する。

2 発話意図データベースの構築

提案手法ではRAの推定に発話文中の発話意図を用いる。発話意図は用意した発話意図データベースを用いる。発話意図データベースには助詞助動詞の意味と発話意図の対応を示したルールをおさめている。助詞は係助詞、終助詞、副助詞の25種類で助動詞は22種類あり、それらの意味は日本語辞書大辞林 [6] に掲載されていた意味を採用した。の3種類を用いた。表1に発話意図と助詞助動詞の意味との対応を示す。

発話意図データベースは以下の手順で構築した。まず助詞助動詞が持つ意味を似ているものでまとめて、意味クラスタを作った。続いて意味クラスタをそれらが使われる状況が似ているものでまとめて、上位のクラスタを作った。順に使われる状況が似ているクラスタをまとめていき、まとめきれなくなったところでその

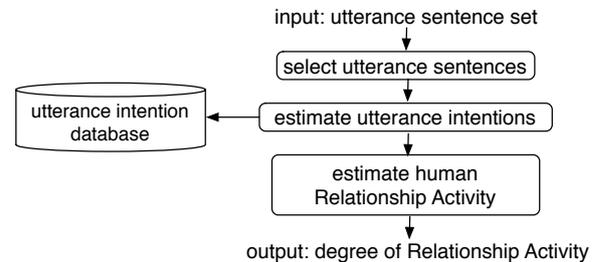


図 1: Relationship Activity 推定システム

クラスタに最適な発話意図を表すラベルを付けた。次に不足している発話意図を探した。ここではSWBD [8] で採用されている発話タグ¹を発話意図に割り当てていき、割り当てられることなかった発話タグが使われると考えられる発話意図を新しく作った。新しく作った発話意図を表すのに使われると考えられる意味のクラスタを割り当て、データベースにおさめた。

3 関与の強さ推定手法

提案システムの処理の流れを図1に示す。以下では各処理の詳細を述べる

3.1 発話文の選別

入力された発話文集合から助詞助動詞が含まれる発話文のみを取り出す。ここでは発話意図は文末に現れるということを考え [1]、文末に助詞助動詞が含まれる発話文のみを取り出す。

入力された発話文集合 S_{init} の各発話文 s に対して、形態素解析を行う。形態素解析には茶筌 [7] を用いる。文末に助詞助動詞が含まれる発話文のみを選び、選ばれた発話文からなる新たな発話文集合を S とする。

*連絡先: 大阪大学大学院基礎工学研究科システム創成専攻
〒560-8531 大阪府豊中市待兼山町1-3
E-mail: yoko@yachi-lab.sys.es.osaka-ua.c.jp

¹発話の種類を表すもので、疑問文、平叙文、提案文、賛成文など66種類存在する。

3.2 発話意図の同定

発話文集合 U の各発話文の意図を表 1 の発話意図データベースを用いて推定する

U の各発話文 s に含まれる助詞・助動詞の集合を P とする．図 1 の発話意図データベースを参照し， P の各助詞助動詞 p から考えられる意図 i を全て出力する．各意図 i には式 (1) で重要度をつける．

$$importance(i) = freq(i) * conf(i) \quad (1)$$

$$freq(i) = \sum corres(i, p) * place(p) \quad (2)$$

$$place(p) = \begin{cases} 1 & \text{from sentence end} \\ 0.5 & \text{otherwise} \end{cases} \quad (3)$$

式 (1) において， $freq()$ はその意図が出力された元の助詞助動詞の重み付き和であり， $conf()$ は意図の信頼度を表す．式 (2) において， $corres(i, p)$ はデータベース内に意図 i と助詞助動詞 p の対応があれば 1，なければ 0 を返す関数であり， $place(p)$ は p が文末に存在するならば 1，文中であれば 0.5 を返す関数である．

各発話文の意図として式 (1) の重要度の上位 3 つとそれらとの類似度が高いものを出力する．意図 i, j の類似度は式 (4) で与える．

$$correl(i, j) = \frac{v(i)v(j)}{|v(i)||v(j)|} \quad (4)$$

式 (4) において， $v()$ は [8] の発話タグを要素とするベクトルであり， $|v()|$ はベクトルの大きさである．ベクトル $v(i)$ の要素は，意図 i を表すのに使われると思われる発話タグを 5 名の人間がそれぞれ与えたときに，過半数の人間が与えた発話タグのみとした．

3.3 関与の強さの推定

推定した発話意図を用いて 2 人の人間の関与の強さ RA を測る．RA は発話集合 S に含まれた発話意図の種類数から決定する．発話意図の種類数が多いほどさまざまな会話を行っており，発話意図の種類数が多ければその 2 人の RA も高いと考えられる．また，発話者の一方の発話意図の種類数が他方より多いならば一方的に喋っており RA は低いと考え，双方の発話意図の種類数が高いほど RA が高いとする．式 (5) を用いて 2 人の発話者 $h1, h2$ の RA を評価する．

$$RA(h1, h2) = \frac{type(h1) * type(h2)}{num} \quad (5)$$

式 (5) において $type()$ は発話意図の種類数であり， num は発話意図がとれた発話数の合計である．

表 1: 作成した発話意図と助詞助動詞の意味の対応

発話意図	助詞助動詞の意味
心情	感動, 意志, 自発, 可能, 不可能, 希望, 詠嘆, 強調
事実	打消, 確認, 質問, 添加, 並立, 選択, 不確定, 列挙, 反語, 程度, 限定, 軽視
補足	他との区別, 添加, 並立, 選択, 不確定, 列挙, 強調
価値判断	打消当然, 不確かな気持ち, 理由
知識獲得	確認, 疑問, 質問, 希望, 願望, 不確かな気持ち, 問掛
知識提供	強調
教える	類推, 推量, 打消推量, 推定, 婉曲, 打消, 強調
教えられる	確認, 疑問, 質問, 願望, 依頼, 不確かな気持ち, 問掛
依頼	疑問, 質問, 希望, 願望, 依頼, 不確かな気持ち
確認	不確かな気持ち, 確認
行動要求	意志, 希望, 願望, 依頼, 呼びかけ, 禁止, 同意要求, 勧誘, 不確かな気持ち, 命令, 許可, 原因
発言要求	婉曲, 確認, 疑問, 質問, 意志, 念押, 呼びかけ, 同意要求, 勧誘, 命令, 問掛, 許可
提案	他との区別, 反語, 意志, 希望, 強調
礼	尊敬, 自発
謝罪	尊敬, 意志, 自発
賛成	詠嘆, 適当, 当然, 全面否定, 全面肯定
反対	打消, 打消意志, 打消強調, 非難, 全面否定, 全面肯定, 反対
行動要求受入	打消, 打消当然, 確認, 質問, 意志, 自発, 可能, 不可能, 打消意志, 念押, 願望, 依頼, 過去
発言要求受入	比例, 様態, 伝聞, 例, 例示, 類推, 推量, 打消推量, 推定, 婉曲, 打消, 打消当然, 確認, 質問
納得	婉曲, 確認, 断定, 不確かな断定, 過去
理由陳述	理由, 価値判断
保持	不確かな気持ち, 確認, 列挙
話題転換	依頼, 意志, 質問

4 意図推定精度評価実験

4.1 実験手順

発話文集合を入力し，文末に助詞助動詞が含まれる発話文を抜き出し，発話意図データベースを用いて重要度の上位 3 つ，およびそれらとの類似度が高い意図を出力した．出力した意図とあらかじめ人手で与えておいた正解意図を比較し，出力意図の中に正解意図が含まれるかどうかを確認した．本実験では式 (1) の $conf(i)$ を全て 1 とした．

4.2 使用データ

用いたデータは研究者同士の間で交わされたメールの本文である．データ数は 3 組であり，全ての組において一方の人物は同じである（以下基本人物，他方を

表 2: 予備実験データの意図推定発話文数 (カッコ内は入力発話文数)

データ	基本人物	相手人物
Data1	25 (70)	16 (61)
Data2	14 (42)	15 (72)
Data3	21 (62)	17 (47)

表 3: 式 (1) の $conf$ の有無での意図推定精度

$conf$	適合率	再現率
有	0.35	0.74
無	0.19	0.43

相手人物と呼ぶ)。入力した発話文数を基本人物、相手人物ごとに表 2 に示す。

正解意図であるが 3 名の人間に文末に助詞助動詞が含まれる発話文を示して、一つの文から読み取れる発話意図を表 1 に示す中からあるだけ選んでもらい、選ばれた全てを正解意図とした。正解意図の平均数は一発話文あたり 1.2 つであった。

4.3 実験結果

初めに意図ごとの推定精度を図 2 に示す。図 2 から意図ごとに適合率再現率が異なっており、意図推定の信頼度に差があることが分かる。そこで意図推定で得られた適合率を式 (1) の $conf$ の値として用いて、正確な RA 推定を行うことにした。

図 3 適合率を信頼度としたときの発話文の意図推定精度を表 3 に示す。推定意図の平均出力数は 2.3 つであった。表 3 からは出力意図を信頼度で絞り込むことで、適合率再現率が向上したことが分かる。この結果から信頼度を加えた意図推定手法の有効性を確認した。

表 4: 関与の強さ RA の推定評価のデータ

データ	内容	相手人物数
メール	研究全般	3
掲示板	教育	5
チャット 1	授業内容	4
チャット 2	wikipedia の管理	9

5 関与の強さの推定実験

5.1 実験手順

発話文集合を入力し、助動詞助詞が含まれる発話文の意図を推定、推定した発話意図から RA を算出した。出力した RA と人の主観による RA を比較した。

5.2 使用データ

用いたデータの内容を表 4 に示す。使用データはメール、掲示板、チャットを通じて行われた会話である。本実験では基本人物として各データにおいて発言回数が多い人物を選択した。基本人物と会話した相手人物の数は表 4 に示す通りである。

正解となる RA は以下の手順で用意した。被験者に 2 人だけの会話からなる発話集合を読んでもらい、各データ内のペアをお互いの関係を維持向上させようとしていると思う順に並べてもらった。正解 RA は 5 名の被験者が並べた順序の順位合計とし、合計値が高いほど RA が高いとした。

5.3 実験結果

表 5 に式 (5) による RA と正解 RA の相関を示す。いずれのデータにおいても相関値 0.7 以上の強い正相関が得られた。使用した会話データは各ペアの発話数に差があり、発話数と正解 RA には必ずしも相関がないことを確認した。各ペアの発話数に差があっても提案手法は適切な RA を付与できることが分かる。

表 5 のチャットデータでの RA 相関値が他のものより低くなった。チャット 1 の相関値が低い理由は相手人物の推定発話数が基本人物の推定発話数に比べ、差があったためである。両者の入力発話文数は 15 文、12 文であったが意図が推定できた発話文は 14 文、5 文であった。このため正しい RA を推定できなかった。

チャット 2 の相関値が低い理由は意図推定ができた発話数が少ない人物を正しく評価ができなかったためである。表 6 には正解 RA の順に RA と意図推定ができた発話割合を示した。表 6 からは C1 の意図が推定できた発話割合が 0.23 と低いことが分かる。C1 は基本人物とくだけた口調で会話していたため文章の省略が多い。人は省略された部分も想像して RA を評価したと考えられ、意図がとれない発話数が多いと適切な RA を付与できないことが分かる。表 6 からは意図が推定できる発話数が 0.4 以下の人物 (先頭に*印) では正解 RA と評価 RA がずれることが分かる。この結果から正しい RA 値が推定できるのは文末に助詞助動詞を含む発話文が入力文に対して 4 割を超えるときであると分かった。

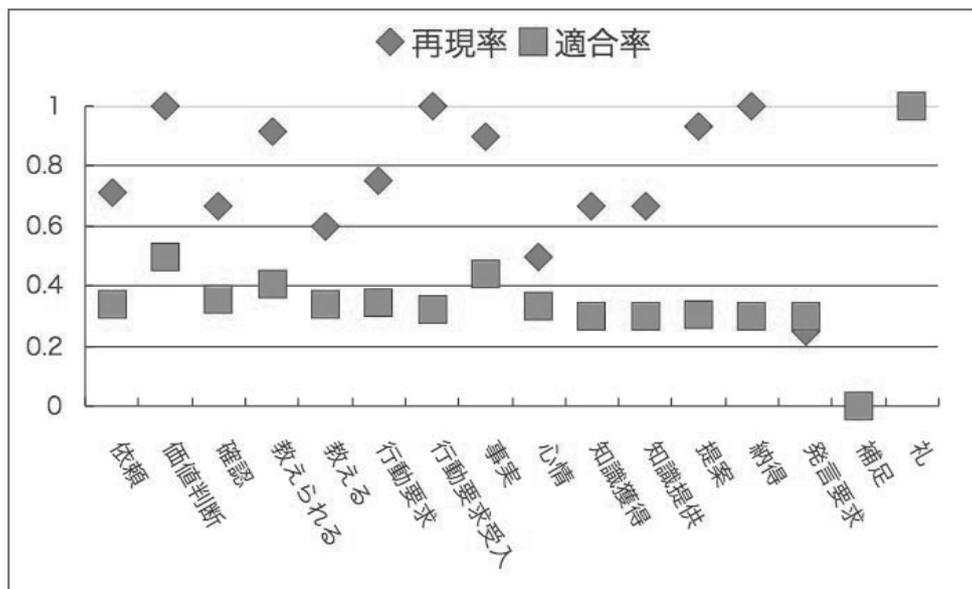


図 2: 意図ごとの推定適合率と再現率

表 5: RA 推定評価値と正解評価値の相関

データ	相関値
メール	0.84
掲示板	0.95
チャット 1	0.73
チャット 2	0.75

表 6: チャット 2 での RA と意図推定された発話割合

相手人物	RA	意図推定された発話割合
*C1	2.06	0.23
C2	3.33	0.52
C3	3.68	0.41
*C4	2.04	0.37
C5	2.35	0.53
C6	1.68	0.53
C7	1.13	0.40
C8	1.84	0.45
C9	0.50	0.40

6 むすび

本稿では発話意図からの関与の強さ Relationship Activity 推定手法を提案した。提案手法は 2 人の人間の発話文の発話意図を助詞助動詞の意味から推定し、発話意図の種類数から関与の強さを推定するものである。実験結果から入力された発話文のうち、文末表現省略文が 4 割以下であれば RA を推定できることを確認した。

今後の課題には RA が高い人間からの情報のみを入力する、情報フィルタリングシステムの構築があげられる。

謝辞

本研究の一部は日本学術振興会（特別研究員奨励費 17・9881）の助成による。

参考文献

- [1] 益岡隆志: “モダリティの文法”, くろしお出版 (1991).
- [2] the foaf project, (URL) <http://www.foaf-project.org/>.
- [3] RELATIONSHIP: A vocabulary for describing relationships between people, (URL) <http://vocab.org/relationship/>.
- [4] Josh Tyler, Dennis Wilkinson, and Bernardo A. Huberman: “Email as Spectroscopy: Automated Discovery of Community Structure within Organizations”, HP Laboratories, Palo Alto, CA 94304.
- [5] Yutaka Matsuo and Yuki Yasuda: “An Analysis of Researcher Network Evolution on the Web”, ANDI’05 (2005).
- [6] infoseek マルチ辞書 (URL) <http://jiten.www.infoseek.co.jp/>
- [7] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka, Masayuki Asahara: “Japanese Morphological Analysis System ChaSen version 2.2.”, Dec (2000).
- [8] D. Jurafsky, L. Shriberg and D. Binasca: Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, <http://www.dcs.shef.ac.uk/nlp/files/bib/ics-tr-97-02.pdf> (1997).