

発話意図を用いた Web における人同士の関与の強さ推定手法

Relationship Activity Estimation by Using Utterance Intention in Conversation on the Web

西原陽子*1 砂山渡*2 谷内田正彦*1
NISHIHARA, Yoko Sunayama, Wataru YACHIDA, Masahiko

*1 大阪大学基礎工学研究科 *2 広島市立大学情報科学部
Graduate School of Engineering Science, Osaka Univ. Faculty of Information Science, Hiroshima City Univ.

We often gather information with e-mail, BBS and SNS service on the Web. We need to grasp our own human relationships on the Web for effective information gathering. We are able to know the strength of our friend's relationship from their conversations. However, it is difficult to grasp all relationship strength. The relationship strength is not worked on information gathering on the Web. This paper proposes a system for relationship strength estimation. The system identifies the utterance intention from particles and auxiliary verbs that are included in the utterance text. The strength of relationship is estimated with the combination of the particles and auxiliary verbs. Experimental results proved that the correlation between A and B remains as high as 0.7.

1. まえがき

電子メールや電子掲示板, ソーシャルネットワークサービスを利用した Web 上での議論, 情報収集が行われるようになった。議論, 情報収集を効果的に行うには, 自らの Web 上での人間関係を把握することが重要である。人は他人同士の人間関係の強さを彼らの会話から予測することができ, 自分と他人との人間関係の強さも無意識のうちにおおよそ把握できる。しかし自らの人間関係であっても, 全ての関係の強さを把握することは難しい。そのため現状では Web 上でのやりとりにおいて, 自らの人間関係を活かしきれていない。本稿では発話文の発話意図を用いて, 人間関係の強さを推定する手法を提案する。提案手法は発話文に含まれる助詞・助動詞の意味から発話意図を同定し, 発話意図の組合せから関係の強さを推定するものである。

2. 関連研究

2.1 発話意図の同定

人の発話は意味内容から, 命題と文の述べ方に対する発話者の態度に分けられる。文の述べ方は主に文末表現に現れ, 発話意図は文と文の間に現れる [1]。この文末表現から発話意図の一部が取れることが確認されている。

熊本らが開発した対話型ヘルプシステムはユーザが出した質問文の発話意図を推定し, 質問文に対する最も適切と思われる答えを返す [2]。発話意図の推定には文末表現, 発話文中の動詞, それらの出現順序を用いる。発話意図は 12 種類あり, ユーザが質問文に含めると予想される発話意図が用意されている。人の発話意図は無数に存在するが, 美馬らの手法においてもあらかじめ 5 種類の発話意図が定義されている [3]。

本研究では従来研究と同様に文末表現を用いて発話意図を推定する。具体的には文中に含まれる助詞助動詞の意味と発話意図の対応を示した発話意図データベースを用意し, データベースを用いて発話意図を推定する。また扱う発話は研究者のコミュニティにおける会話で出される文とし, そのコミュニティで交わされると予想される発話意図を用意する。

2.2 人間関係の解析

電子メール, SNS, 掲示板を通じて構築される対人関係は直に会って構築される対人関係に比べ全体像を把握しにくい。見えにくい人間関係を視覚化する研究を紹介する。

Web 上での人間関係をエージェントが解析するために開発された言語に FOAF がある [4]。FOAF は自分の名前, 属する組織, ホームページ, チャット ID, 友人などを記述できる言語である。FOAF を用いると友人のプロフィールだけでなく, 友人の友人の情報も見ることができる。FOAF での友人関係は知ってるか否かの 2 値であるため, 友人関係の強さを正確に表現するための言語として RELATIONSHIP [5] が用意されている。FOAF, RELATIONSHIP での友人関係は人手で記述されるものであり, 全ての友人関係を書くことは大きな負担となる。

Web 上にある情報を用いて自動的に人間関係を解析するシステムも存在する。Tyler らはメールのやりとりから組織の中の人間関係をグラフで表すシステムを開発した [6]。人をノード, やり取りがある場合にはエッジをはり, 重要なノードには色をつけ人間関係を視覚的に分かりやすく表現している。松尾らは研究者の協働関係を検索エンジンでの名前のヒット数から評価して, 協働関係をグラフで表現するシステムを開発した [7]。2 人の研究者の氏名をクエリとして検索したときのヒット数が多いほど関連が強いとしている。

Tyler や松尾らの手法は Web 上の情報を用いて人間関係を取り出すため人にかかる負担は多くない。しかし, Tyler の方法であればメールが集中する上司が重要な人物となり, その周りにいる人間との関連が他の人同士の関連よりも強くなる。松尾らの方法でも有名な人との検索ヒット数は他よりも大きくなるため, 必ずしも正しい関与の強さを表してるとはいえない。

本研究では発話に含まれる意図の違いが関与の強さを表すという仮定を置き, 発話意図の種類数から正確な関与の強さを測る。

3. 関与の強さ推定システム

発話文から 2 人の人間の関与の強さを推定する方法を説明する。提案システムの処理の流れを図 1 に示す。

システムは入力に 2 人の人間の会話からなる発話文集合 (表 1) をとる。入力された発話文を形態素解析し, 文末に助詞助動詞が含まれる発話文のみを取り出す。続いて発話意図データ

連絡先: 西原陽子, 大阪大学基礎工学研究科, 560-8531
豊中市待兼山町 1-3, 06-6850-6363, 06-6850-6341,
yoko@yachi-lab.sys.es.osaka-u.ac.jp

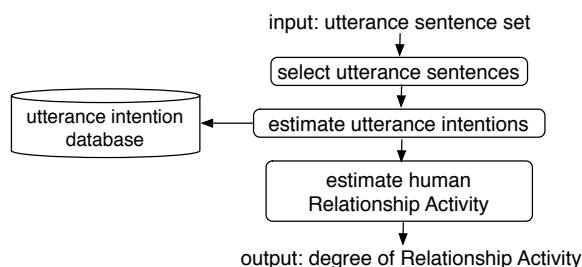


図 1: 関与の強さ推定システム

表 1: 入力発話例 (5章実験のメールデータから抜粋)

話者 1	6月27日(月)にMさんと打合せをする ことになりました。
話者 2	Nさん、ご興味があればご連絡ください。 その頃は大阪にいますので、用事がなければ行 きます。 不参加の場合は連絡します。
話者 1	午後2時ちょうどくらいに現地集合にしましょ うか。
話者 2	現地集合ということで了解しました。 デフォルトでは行くということで、行けなくな れば連絡します。
話者 1	13時20分くらいにA駅の改札あたりで待ち 合わせして、そこから一緒に行きましょうか?

ベースを参照し、含まれている助詞助動詞の組み合わせから発話文の発話意図を同定する。次に推定された発話意図の種類数から2人の関与の強さRAを算出し、最後にRA値を出力する。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を各人の発話文集合に分けて、助詞助動詞が含まれる発話文を選別する。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を選別する。

入力された発話文集合 S_{init} を発話者 $h1$ と $h2$ の発話集合 $S1_{init}$, $S2_{init}$ に分類する。続いて $S1_{init}$, $S2_{init}$ の各発話文 s に対して、茶筌[9]を用いて形態素解析を行う。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それぞれ $S1$, $S2$ とする。

3.2 発話意図データベース

助詞助動詞の意味から発話意図を同定するための発話意図データベースを説明する。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がおさめられている。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[8]に掲載されている意味を採用する。表2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する。初めに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る。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る。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

ころでそのクラスタに最適な発話意図を表すラベルを付ける。次に不足している発話意図を探す。ここではSWBD[10]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る。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、発話意図データベースにおさめる。

3.3 発話意図の同定

発話文集合 $S1$, $S2$ の各発話文 s の発話意図を表1の発話意図データベースを用いて同定する。ここでは各発話文に対して表1の発話意図からなる発話意図ベクトルを出力する。発話文に含まれる助詞助動詞の意味と発話意図データベースでの対応回数をベクトルの要素の値とする。発話文 s の発話意図ベクトル I_s を式(1)で表す。

$$I_s = (x_1, x_2, \dots, x_{23}) \quad (1)$$

発話文 s_i に含まれる助詞助動詞の集合を P_{s_i} とする。 P_{s_i} 内の助詞助動詞 p_j の意味 $m_{p_j}^k$ から式(2)を用いてベクトルの値を求める。

$$I_{s_j}(x_i) = \left(\sum_{p_j \in P_{s_i}} freq(x_i, m_{p_j}^k) \right) conf(x_i) \quad (2)$$

式(2)の $freq(x, m)$ は発話意図データベースにおいて x と m の対応があり、 m が文末に存在する助詞助動詞 p から得られた場合は1、それ以外の場所から得られた場合は0.5を返す関数である。式(2)では、文末に存在する助詞・助動詞と対応がある発話意図の値を高くする。これは発話意図は文と文の間に現れるものであり[1]、文末で用いられている助詞・助動詞が、文中で用いられているものよりも発話意図に関係すると思われるからである。 $conf(x)$ は同定された発話意図の信頼度を返す関数であり、ここでは予備実験によって求めた推定の適合率(2)を返す関数とする。

続いて発話意図ベクトル I_{s_j} のノイズを除去する。ここでは発話意図ベクトル I_{s_j} の要素の値の上位3つとそれとの類似度が閾値を超えるものを残し、それ以外の要素の値を0とする。発話意図 x_i, x_m の類似度 $correlation$ は式(3)で与える。

$$correlation(x_i, x_m) = \frac{T(x_i)T(x_m)}{|T(x_i)||T(x_m)|} \quad (3)$$

式(3)において、 $T(x)$ はSWBD[10]の発話タグを要素とするベクトル^{*2}であり、 $|T(x)|$ はベクトルの大きさである。

3.4 関与の強さの推定

同定した各発話文の発話意図を用いて2人の人間の関与の強さRAを測る。RAは発話集合 $S1$, $S2$ に多数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いければその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いるとみなせるため、双方の発話意図の種類数が高いほどRAが高いとする。式(4)を用いて2人の発話者 $h1, h2$ の $RA(h1, h2)$ を評価する。

*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。

*2 ベクトル $T(x)$ の要素は、発話意図 x 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。

表 2: 発話意図と助詞助動詞の意味の対応

発話意図	対応する助詞助動詞の意味
x_1 心情	感動, 意志, 自発, 可能, 不可能, 希望, 詠嘆, 強調
x_2 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
x_3 補足	他との区別, 添加, 並立, 選択, 不確定, 列挙, 強調
x_4 価値判断	打消当然, 不確かな気持ち, 理由
x_5 知識獲得	確認, 疑問, 質問, 希望, 願望, 不確かな気持ち, 問掛
x_6 知識提供	強調
x_7 教える	類推, 推量, 打消推量, 推定, 婉曲, 打消, 強調
x_8 教えられる	確認, 疑問, 質問, 願望, 依頼, 不確かな気持ち, 問掛
x_9 依頼	疑問, 質問, 希望, 願望, 依頼, 不確かな気持ち
x_{10} 確認	不確かな気持ち, 確認
x_{11} 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
x_{12} 発言要求	婉曲, 確認, 疑問, 質問, 意志, 念押し, 呼びかけ, 同意要求, 勧誘, 命令, 問い掛け, 許可
x_{13} 提案	他との区別, 反語, 意志, 希望, 強調
x_{14} 礼	尊敬, 自発
x_{15} 謝罪	尊敬, 意志, 自発
x_{16} 賛成	詠嘆, 適当, 当然, 全面否定, 全面肯定
x_{17} 反対	打消, 打消意志, 打消強調, 非難, 全面否定, 全面肯定, 反対
x_{18} 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
x_{19} 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
x_{20} 納得	婉曲, 確認, 断定, 不確かな断定, 過去
x_{21} 理由陳述	理由, 価値判断
x_{22} 保持	不確かな気持ち, 確認, 列挙
x_{23} 話題転換	依頼, 意志, 質問

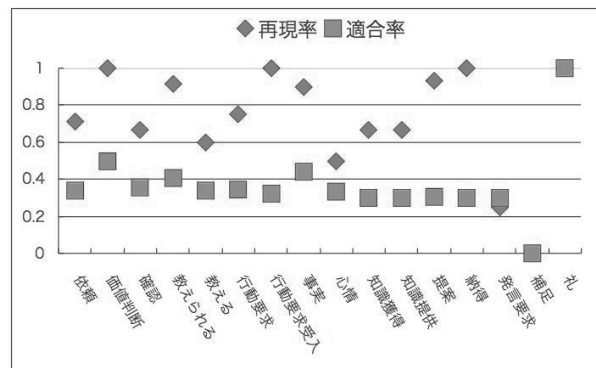


図 2: 発話意図ごとの適合率と再現率

$$RA(h1, h2) = \frac{type(h1)type(h2)}{num} \quad (4)$$

式 (4) において $type(h)$ は発話者 h の発話集合 S の発話 s_j から得られた発話意図ベクトル I_{s_j} を発話者ごとに足し合わせ, 要素値の合計が閾値を超えた発話意図の数を表す. num は発話意図が同定された発話数である.

4. 関与の強さ推定実験

4.1 実験手順

発話意図を用いた関与の強さ推定実験を行い, 提案手法の有効性を確認した. 発話文集合を入力し, 助詞助動詞が含まれる発話文の発話意図を出力, 出力した発話意図から RA を算出した. 比較評価のために 2 人の入力発話集合中の (a). 発話数の合計値, (b). 単語の種類数, (c). 助詞助動詞のべ数, (d). 式 (4) による値の 4 種類の RA を算出し, 出力した RA と人の主観による RA の相関を測った.

4.2 使用データ

用いたデータの内容を表 3 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り, 全て研究に携わる人同士の会話である. 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最も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 3 に示す通りである, 正解となる RA 値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の合計とし, 合計値が高いほど RA が高いとした.

4.3 実験結果

表 4 にメールデータでの RA の推定値と RA 値の算出に使った意図とその値を示す. メールデータでは各ペアの発話数に差はなかった. 表 4 の正解 RA と発話意図の数を見ると正解 RA が高いほど各ペアの発話意図の種類数も多くなっていることが分かる. 各人の発話意図の種類数やその和には大きな差はないが, 種類数の積をとることで正解の RA の順と近いものが得られた. このことから式 (4) の手法が RA の推定に有効であることが分かる.

表 4 ではどの発話者においても式 (2) の値が 0 の発話意図が 6 種類存在した. これは助詞助動詞との対応が少ない発話意図は出力されにくいためである. だが表 4 の出力された発話意図を見ると各ペアで出力発話意図は異なっており, 全く同

表 3: 関与の強さ推定実験データ

データ	内容	相手人物数	平均発話数	最小発話数	最大発話数
メール	研究全般	3	59	42	72
掲示板	教育	5	43	5	110
チャット 1	授業内容	4	22.5	9	33
チャット 2	wikipedia の管理	9	75.8	14	156

表 4: メールデータの発話ペアごとの RA 算出に用いた発話意図 (M1, M2, M3 は相手人物)

発話意図	M1		M2		M3	
	相手	基本	相手	基本	相手	基本
心情	3.46	0	2.77	4.85	3.77	4.62
事実	16.1	14.6	8.84	24.4	11.2	5.49
補足	0	0	0	2.53	0	0
価値判断	0	0	0	0	0	0
知識獲得	0	3.13	2.44	0	0	0
知識提供	0	0	2.3	0	0	0
教える	2.3	0	2.99	5.29	0	0
教えられる	0	4.41	2.45	0	2.21	2.21
依頼	2.02	5.18	4.03	0	2.88	2.59
確認	0	0	0	3.31	0	0
行動要求	0	4.05	2.32	0	2.32	3.09
発言要求	0	2.91	0	0	0	2.08
提案	0	2.95	0	3.4	0	0
礼	0	0	2.00	2.00	0	2.00
謝罪	0	0	0	0	0	0
賛成	0	0	0	0	0	0
反対	0	0	0	0	0	0
行動要求受入	3.42	3.75	2.14	3.42	2.25	0
発言要求受入	4.10	2.4	2.2	4.4	2.3	0
納得	2.14	0	2.21	0	0	0
理由陳述	0	0	0	0	2.5	2.66
保持	0	0	0	0	0	0
話題転換	0	0	0	0	0	0
意図数	7	9	12	9	8	8
評価 RA		0.09		0.15		0.10
正解 RA		8		15		7

一の発話意図が出力されたわけではない。このことから各ペアでの会話内容の違いをとらえられたことが分かる。会話全体の内容が異なると RA の大きさにも差が生じる。発話内容の違いを発話意図からとらえて RA の推定を行えたことから、RA を推定する本手法の有効性を確認できる。

表 5 に (a) から (d) までの 4 つの手法で算出した RA と正解 RA の相関を示す。チャット 1 データ以外の 3 つのデータにおいて提案手法が最も高い相関値を示した。表 5 より掲示板データやチャット 2 データでは発話数による相関値も高い値を示しているが、その他の値は RA の大きさが発話数や単語種類、助詞助動詞の種類数と必ずしも関係がないことを示している。表 5 の結果は発話数の量によらず適切な RA 値を提案手法によって算出できることを示している。

チャット 1 データでは発話数による相関値が最も高くなった。チャット 1 データは他のデータと比べ平均発話数が低く、発話意図の同定が十分に行われなかったためである。チャット 1 での 1 人あたりで出力された平均発話意図数は 1.7 つでありこれは他のデータと比べても少ない。また全体的に発話数が少ないデータでは、互いに話す回数が多いほど RA も高いと考える人が多い。そのため発話数による相関値が高くなった。このことから提案手法が扱えるデータは発話数に下限があることが分かる。

5. むすび

本稿では、会話テキストから人間関係の強さを推定するシステムを提案した。提案システムは 2 人の人間の会話テキスト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

表 5: 4 種類の手法による RA 推定評価値と正解評価値の相関

データ	メール	掲示板	チャット 1	チャット 2
(a) 発話数	0.21	0.80	0.75	0.68
(b) 単語種類数	-0.99	0.82	0.72	0.32
(c) 助詞助動詞数	-0.95	0.88	0.73	0.24
(d) 提案手法	0.84	0.95	0.72	0.73

図の種類数から人間関係の強さを推定するものである。実験結果から提案手法の有効性を確認した。

今後は出力発話意図の違いから RA の種類の違いを表現する方法を考え、上下関係など他の関係も推定できるようにしたい。また、人間関係の推定によって、特定の人間からの情報のみを出力する情報フィルタリングシステムの構築を目指す。

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] Tadahiko KUMAMOTO, Akira ITO and Tsuyoshi EBINA: “Recognizing User Communicative Intention in a Dialogue-Based Consultant System — A Statistical Approach Based on the Analysis of Spoken Japanese Sentences —”, *Systems and Computers in Japan*, Vol.25, No.14, pp.91-105 (1994).
- [3] Hideki Mima, Kazuaki Ando and Jun-Ichi Aoe: “A Robust Method for Understanding NL Interface Commands in the ICI Intelligent Command Interpreter”, *Inf. Sci.*, Vol.85, No.1-3, pp.185-198 (1995).
- [4] the foaf project, (URL) <http://www.foaf-project.org/>.
- [5] RELATIONSHIP: A vocabulary for describing relationships between people, (URL) <http://vocab.org/relationship/>.
- [6] Josh Tyler, Dennis Wilkinson, and Bernardo A. Huberman: “Email as Spectroscopy: Automated Discovery of Community Structure within Organizations”, HP Laboratories, Palo Alto, CA 94304.
- [7] Yutaka Matsuo and Yuki Yasuda: “An Analysis of Researcher Network Evolution on the Web”, ANDI’05 (2005).
- [8] infoseek マルチ辞書 (URL) <http://jiten.www.infoseek.co.jp/>.
- [9] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka, Masayuki Asahara: “Japanese Morphological Analysis System ChaSen version 2.2.”, Dec (2000).
- [10] D. Jurafsky, L. Shriberg and D. Bisca: Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, <http://www.dcs.shef.ac.uk/nlp/files/bib/ics-tr-97-02.pdf> (1997).