

## 議論マイニングによる議論掲示板利用者の能力推定

森尾 学†

藤田 桂英‡

†東京農工大学 工学部 情報工学科

‡東京農工大学大学院 工学研究院 先端情報科学部門

### 1 はじめに

近年、WEB 議論掲示板が有益な意見交換の場として注目されている。場所や時間を問わず誰でも発言を書き込むことが出来るため、多人数による大規模な議論が期待される。しかし、投稿数に比例して、議論に対して有益でない情報が大量に投稿されるという問題がある。そのような状況を防ぐために投稿者の議論的能力を適切かつ自動で判別することが必要である。本論文では議論に参加する人物の専門性やコミュニケーション能力などの議論的能力を自動推定することを目的とする。特に議論構造に関する研究領域である議論マイニング (Argument Mining) [1] の概念や、グラフ構造や影響量の普及モデルを取り入れた特徴素を提案する。さらに、提案した特徴素を用いてサポートベクターマシンによる学習を行い、有用性を評価する。

### 2 提案手法

#### 2.1 語彙的特徴量

特徴素には、議論マイニングにおいて主流である語彙的特徴素を用いる。本論文では宮崎らの知識量推定 [2] にならい、語彙的特徴量として IDF (Inverse Document Frequency), 異なり語数, 対話的特徴量, 質問に関する特徴量を用いる。IDF は単語の専門性の指標であり、式 (1) で求まる。

$$idf_t = \log \frac{N}{df(t)} \quad (1)$$

式 (1) の  $N$  は総文書数、 $df(t)$  は索引語  $t$  が含まれる文書数であり、本論文では IDF を正規化した値を特徴量とする。異なり語数はユーザのユニークな発話数であり、対話的特徴量は感動詞や助詞、助動詞の出現回数を正規化した特徴量である。質問に関する特徴量は「質問」が利用者の持ち合わせる情報量が反映されていると考え、質問文の出現回数を正規化した特徴量である。また、新たに、議論ユニットによる特徴素を提案する。

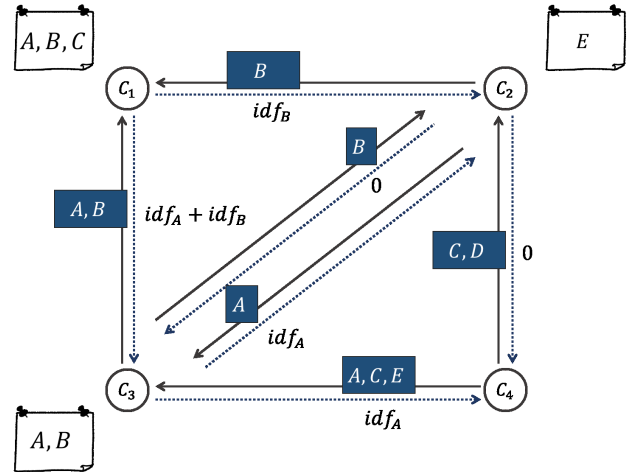


図 1: IDF を重みとしたユーザネットワークの例

Eckle-Kohler らは議論的言説における談話標識が「前提」文と「主張」文の分離に有効であることを示した [3]。そこで「前提」の談話標識に関連しやすいと思われる品詞「接続詞, 副助詞, 係助詞, 接続助詞, 並立助詞, 連体化, 詞類接続」に着目し、これらの品詞について利用者毎の正規化された出現回数を特徴量として用いる。

#### 2.2 IDF 影響量に基づく固有ベクトル中心性

本論文では議論掲示板の返信関係と投稿内容の語彙的普及性に着目し、松村らの影響の普及モデル [4] を、固有ベクトル中心性の重みとした特徴量を提案する。さらに、平易な単語の特徴量を吸収し単語を特徴づけやすい IDF を重みとして用いる。図 1 に、IDF を重みとしたユーザネットワークの例を示す。

IDF を重みとしたユーザネットワークは、議論掲示板利用者をノードとし、返信関係をエッジとした有向グラフ  $G := (V, E)$  の一例である。コーパスは「A,B,C,D,E」の 5 単語のみで、例えばユーザノード  $C_1$  のコメントは「A,B,C」の 3 単語で構成されており、 $C_1$  が  $C_3$  へ与える影響量は  $C_1$  の語彙集合「A,B,C」と返信コメント「A,B」の積集合で、その値は  $idf_A + idf_B$  である。ここで、ユーザノード  $x$  に対してユーザノード  $y$  がコメントを返信する構造において、 $x$  のコメントの Bag of Words 表現された単語集合を  $W_x$  とすると、IDF を重みとした固有ベクトル中心性は

Estimating User Competence based on Argument Mining for Discussion Bulletin Board

†Department of Computer and Information Sciences, Faculty of Engineering, Tokyo University of Agriculture and Technology

‡Division of Advanced Information Technology and Computer Science, Institute of Engineering, Tokyo University of Agriculture and Technology

$$c_e(v) = \lambda^{-1} \sum_{t \in V} \left( A_{t,v} \cdot \frac{\sum_{w \in W_t \cap W_v} idf_w}{\sum_{w \in W_t} idf_w} \cdot c_e(t) \right) \quad (2)$$

と表す。本論文では式(2)をIDF影響量に基づく固有ベクトル中心性と定義する。なお、今回は時系列を考慮しないため、 $v$ から $t$ への返信構造が複数存在した場合、1つの返信構造として扱うものとする。

### 3 評価実験

#### 3.1 議論データへのアノテーション

本論文では朱の定義[5]を参考に、「専門性 (Expertness)」「コミュニケーション能力 (Communication)」「議論を仕切る能力 (Argument)」の3能力を定義した。専門性はそのテーマに関して豊富な知見を有するかどうか、コミュニケーション能力は相手を尊重して相手の意見や立場を理解しているか、議論を仕切る能力は議論の流れを考慮しながら結果を出せるように努力しているか、をそれぞれ評価観点とする。

Web 議論掲示板 COLLAGREE[6]上で行われた6つのトピック(総計187人, 1447投稿)に対し、3つの議論能力を7段階で5人の大学生がアノテーションを行った。得られたアノテーションデータを「7,6」「5,4」「3,2,1」の3段階に自由度を上げたところ Fleiss' Kappa 一致率が0.321(一応の一致)以上となった。

#### 3.2 実験結果

「7,6」「5,4」「3,2,1」の3段階による三値分類実験を行い、提案手法が議論掲示板における能力推定手法として有効であるかを調べた。提案手法の比較対象となるベースライン(B)は宮崎らの「IDF, 異なり語数, 対話的特徴量, 質問に関する特徴量[2]」とする。また、提案手法である、議論ユニットによる特徴素をF1, IDF影響量に基づく固有ベクトル中心性による特徴素をF2とする。本研究ではベースラインにF1とF2を追加する形で有用性を評価した。SVMの学習にはガウシアンカーネルを用い、過学習を防ぐため、議論トピック毎にグループを分け、1つのグループをテスト用データ、他の全てのグループを学習用データとする Leave one group out(LOGO)と呼ばれる手法を利用した。表1に、各議論能力における適合率, 再現率およびF値の平均値を示す。

表1より、専門性については、既存手法に対してIDF影響量に基づく固有ベクトル中心性を導入したモデルが顕著な精度向上を達成しており、語彙的特徴量であるF1による精度と比べても高い水準となった。以上から、専門性が語彙的特徴量だけでなく関係グラフに

表1: 各分類における、適合率, 再現率, F値

| 能力            | 特徴素    | 適合率          | 再現率          | F値           |
|---------------|--------|--------------|--------------|--------------|
| Expertness    | B      | 0.575        | 0.612        | 0.506        |
|               | B + F1 | 0.558        | 0.611        | 0.519        |
|               | B + F2 | <b>0.639</b> | <b>0.673</b> | <b>0.614</b> |
| Communication | B      | 0.740        | 0.741        | 0.720        |
|               | B + F1 | 0.753        | 0.747        | 0.725        |
|               | B + F2 | <b>0.761</b> | <b>0.758</b> | <b>0.744</b> |
| Argument      | B      | <b>0.850</b> | <b>0.847</b> | <b>0.842</b> |
|               | B + F1 | 0.829        | 0.810        | 0.815        |
|               | B + F2 | 0.840        | 0.833        | 0.828        |

よる中心性に大きく影響されることが明らかとなった。コミュニケーション能力においても同様の傾向を見せており、判別においてF1とF2いずれも有用であることを示している。しかし、議論を仕切る能力の結果はベースラインが最も高くなったことから、単純な語彙的特徴素が優れる傾向にあることが明らかになった。

### 4 まとめ

本論文では、従来の語彙的特徴量を用いたモデルの拡張手法やグラフ構造, 影響の普及モデルを取り入れた手法を提案した。さらに、評価実験において、多くのケースで提案手法がWeb議論掲示板利用者の能力推定モデルに有用であることを確認した。一方、議論を仕切る能力の分類については単純な語彙的特徴量が良く作用していたことから改善の余地がある。今後は議論を仕切る能力の推定モデルについて着目していきたい。

### 参考文献

- [1] Lippi, M. Torroni, P. Argumentation Mining: State of the Art and Emerging Trends, ACM Transactions on Internet Technology. 16(2): 10, 2016.
- [2] 宮崎 千明, 東中 竜一郎, 牧野 俊朗, 松尾 義博. コールセンタ対話における話者の知識量推定. 言語処理学会第18回年次大会発表論文集, 2012.
- [3] J. Eckle-Kohler, R. Kluge, I. Gurevych. On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015.
- [4] 松村 真宏, 大澤 幸生, 石塚 満. テキストによるコミュニケーションにおける影響の普及モデル. 人工知能学会論文誌. Vol17, No.3, 259-267, 2002.
- [5] 朱 成敏, 武田 英明. 議論の妥当性モデルを用いたオンライン議論の参加者分類. 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎 113(283), 13-18, 2013.
- [6] 伊藤 孝行, 奥村 命, 伊藤 孝紀, 秀島 栄三. 多人数ワークショップのための意見集約支援システム Collagree の試作と評価実験: ~議論プロセスの弱い構造化による意見集約支援~. 日本経営工学会論文誌 66(2), 83-108, 2015.