

Web 議論掲示板利用者の能力推定における 関係グラフ中心性の影響分析

Analyses of Centrality of Directed Diffusion Graphs for Estimating Users' Competence in Online Discussion Bulletin Board

森尾 学 † 藤田 桂英 ‡
Gaku MORIO Katsuhide FUJITA

† 東京農工大学 工学府 情報工学専攻

Department of Computer and Information Sciences, Graduate School of Engineering, Tokyo University of Agriculture and Technology

‡ 東京農工大学大学院 工学研究院 先端情報科学部門

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

Estimating the users' competences in online discussion bulletin board is important research topics. Most existing works of estimating the users' competence are based on linguistic features. However, simple approaches based on linguistic features could miss the important features based on relationships and divisions between users. We propose new features based on linguistic characteristics and the relationship graph for classifying the users' competences; "expertise", "communication" and "organizing." The proposed method can classify the high quality and low quality users by SVM automatically. We analyze the influence of the proposed features based on linguistic and relationship graph for estimating the users' competence in online discussions.

1. はじめに

近年、主に自治体レベルの政策決定のための議論を中心としたタウンミーティングを Web 上の議論掲示板で行う試みがある [1]. 従来の対面式のタウンミーティングは、時間や場所の制約から参加者の階級や年齢が偏ったり人数を確保することが難しいという問題があった。Web 上のタウンミーティングはそのような問題を解決する手段として期待されており、実例として COLLAGREE が登場している [2, 3]. このような Web 議論掲示板は、匿名の多人数による大規模な議論が期待される。しかし、議論参加者数が多い場合、有益でない投稿が増える可能性があるため、有益な投稿を行う人物を自動的に抽出することが重要となる。

本論文では、オンライン議論に関係する参加者の 3 つの重要な能力を定義し、それらの能力の高さを自動判別する手法を提案する。従来、このような自動判別手法における素性は語彙的特徴素が中心であった。提案手法では、さらに語彙的特徴素に加えて、議論構造に焦点を当てた関係グラフ中心性による素性を提案し、導入する。その後、評価実験を行い語彙的特徴素と関係グラフ中心性の役割について議論する。

本論文の構成を以下に示す。まず、本論文の関連研究を示したのちに議論掲示板のための語彙的特徴素と関係グラフ中心性による特徴素を示す。その後、評価実験結果を示し、本論文のまとめと今後の課題を示す。

2. 関連研究

近年、議論マイニングと呼ばれる、データマイニングや機械学習、議論解析の融合分野が注目を集めている [4, 5]. 議論マイニングでは、議論を体系的に解析する試みとして語彙的特徴を中心に用いることが多い。例えば Eckerle-Kohler らは議論的言説における談話標識の役割についての研究を行った [6]. 談話標識とは、言語学において、論理的展開を示す指標となるもの (例えば, as といった前置詞や接続詞など) とされている。筆者らは議論における前提文と主張文に対する、談話標識の判

別の性質を、様々な統計的手法を用いて評価し、談話標識が前提文と主張分の差別化に有用であることを示した。

語彙的特徴だけでなく議論構造を用いた手法として Barkar らの研究がある [7]. Barkar らは、オンラインニュースの議論の解析とグラフ構造的提示のための「問題中心スキーマ」の提案を行い、読者コメントの議論的性質を捉えた要約が、どのようにグラフ構造から生成されるかを示した。

議論能力は様々な分類が提案されているが、Golder らは Usenet Newsgroup の参加者を対象としてオンラインコミュニティの参加者を Celebrity, Newbie, Flamer, Lurker, Troll, Ranter の 6 種類に分類した [8]. Golder らの参加者分類の中で合意形成に関して最も重要な人物は Celebrity である。Celebrity は発言の専門性と社会的スキルという観点でも他者を凌駕しており、コミュニティの参加者の中で最も強く影響を与える人物である。

3. 語彙的特徴と関係グラフ中心性による議論能力推定手法

本論文では自動能力判定手法に関連して、実際の能力に関する定義および特徴抽出手法の提案を行う。素性の抽出は大きく分けて「語彙的特徴素」と「関係グラフ中心性」の二種類で構成される。

3.1 議論能力

能力の自動推定にあたって、本研究では「能力の高い人物」を、Golder らのオンライン議論参加者の能力に関する定義に倣って「発言の専門性と社会的スキルの高い利用者」と考える。しかし、Golder らの定義は複合的であり、個別の能力を推定したい場合に不適切である。そこで、本論文では、Golder らの定義をさらに細分化し、朱の研究 [9] を参考に以下のように定義する。

能力定義 1 (専門性)

そのテーマに関して豊富な知見を有する。

能力定義 2 (コミュニケーション能力)

相手を尊重して相手の意見や立場を理解する。

能力定義 3 (議論を仕切る能力)

議論の流れを考慮しながら結果を出せるように努力する。

3.2 語彙的特徴素

語彙的特徴素は、Web 議論掲示板における投稿文書の語彙を用いた特徴素である。本論文では次の3つの素性を提案する。

IDF

多くの文書で出現する単語よりも少数の文章で出現する単語の難易度が高いと仮定し Inverse Document Frequency (IDF) を用いる。例えば c_i を、参加者 u_i の IDF による特徴量とすると、 c_i は u_i の Bag of Words で表される単語集合 W_i を用いて、次のように表される。

$$c_i = \sum_{w \in W_i} idf_w = \sum_{w \in W_i} \log \frac{N}{df(w)} \quad (1)$$

ここで、式 (1) の N は総文書数、 $df(w)$ は索引語 w が含まれる文書数である。本論文では式 (1) を Web 議論掲示板参加者の特徴素として提案する。

議論ユニット

議論ユニットによる素性は談話標識を用いた素性である。本論文では Ecker-Kohler らの「主張」と「前提」の談話標識による分類 [6] が、日本語文書でも同様に可能と仮定し、「前提」の出現頻度が能力指標になると考えた。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」の出現回数を特徴量として用いる。

トピック類似度

トピック類似度による素性は Web 議論掲示板参加者の発言が議論テーマに沿っているかどうかを類似度で判別する。議論テーマへの類似度を素性とする理由は、あるテーマに対しての専門性が高い人物が別のテーマに熟知しているとは限らないからである。文書間の類似度推定手法には、TF-IDF のコサイン類似度を用いることとした [10]。TF-IDF は文書ごとの単語の出現頻度 Term Frequency (TF) と IDF の積である。本論文では議論トピックが文書で与えられるものとし、議論トピックとユーザごとの投稿文書のコサイン類似度を特徴量とした。表 1 はトピック文書の例である。

3.3 関係グラフ中心性による特徴素

参加者の議論能力は参加者の発言内容、特に返信文における発言に着目することで推定することが可能である。Barkar らの研究 [7] では、問題中心スキーマによる議論構造のグラフを手動生成していたが、それではコストが大きいため実用的ではない。そこで、本論文ではユーザ間の返信関係およびその文書に着目することで、返信構造のある Web 議論掲示板に対して、自動的に素性を抽出する手法を提案する。

まず、利用者の返信関係をグラフで表す。Web 議論掲示板の参加者をノードとし、返信関係をエッジとした有向グラフ $G := (V, E)$ (ユーザグラフ) を導入する。また、 $A(i, j)$ を、ノード $i \rightarrow$ ノード j への辺が存在するときに 1、そうでないときに 0 となるバイナリ変数とする。ユーザグラフにおいて比較的単純な特徴量の一つとして中心性が挙げられる。中心性は広く用いられている指標であり、代表的なものに次数中心性 (degree centrality, [11]) がある。これは、ノードに接するエッジの本数

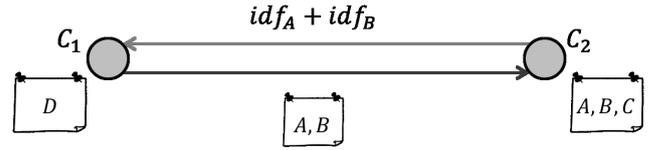


図 1: IDF を重みとしたユーザグラフの例

を中心性の指標とするモデルである。しかし、次数中心性は隣接ノードの中心性を考慮していないという欠点がある。そこで次数及び接続ノードの次数を考慮に入れた固有ベクトル中心性 (eigenvector centrality, [12]) を用いる。また、本論文ではノード i からノード j へ向かうエッジの重みを付与する。エッジの重みには、松村らの影響の普及モデル [13] を用いる。ただし、松村らのモデルは単語の重みを考慮していないため出現しやすい単語がノイズになってしまう [14]。そこで、頻度の高い単語を低く評価し、単語を特徴づけるのに優れた IDF を重みとして用いる。本論文ではこれを IDF 影響量と定義する。固有ベクトル中心性はユーザグラフのノードに関する隣接行列の転置行列の固有ベクトルを求めることで計算される。よって、IDF 影響量を重みとしたユーザノード v の固有ベクトル中心性 $c_{evidf}(v)$ は以下のように表される。

$$c_{evidf}(v) = \lambda^{-1} \sum_{t \in V} \frac{A(t, v) c_{evidf}(t) \sum_{w \in W_t \cap W_v} idf_w}{\sum_{w \in W_t} idf_w} \quad (2)$$

ただし λ は定数である。図 1 に、IDF を重みとしたユーザグラフの例を示す。

図 1 は、2人の参加者で構成されるユーザグラフであり、コーパスは「A, B, C, D」の4単語だけである。例えば、ユーザノード C_2 のコメントは「A, B, C」の3単語で構成されており、 C_1 から「A, B」の Bag of Words (BoW) で返信されている。ここで、 C_2 が C_1 へ与える影響量は、 C_2 の BoW 「A, B, C」と返信コメントの BoW 「A, B」の共通部分で表される。すなわち「A, B」の2単語の IDF 値の合計で、 $idf_A + idf_B$ と表す。

4. 評価実験

4.1 正解データ

正解データを作成するために、5人の被験者が、Web 議論掲示板利用者の3つの議論能力を評価した。評価段階については、Likert 尺度のうち最も粒度の高い7段階のアンケート調査を行った [15]。アンケート評価において、「1」を最も能力が低く「7」が最も能力が高いという基準を設けた。アノテーション対象のデータは Web 議論掲示板 COLLAGREE で議論された、六つの議論トピックのデータセット (参加者数 187 人、投稿数 1447) である。

アノテーション後の正解データに対して、一致率を評価した。3人以上の被験者によるアノテーション結果の一致率を評価する手法として、Fleiss' Kappa [16] を用いた。アノテーションされたデータに対して Fleiss' Kappa を算出したところ、専門性の一致率が $\kappa = 0.18$ であった。これは、Richard Landis らの文献 [17] によれば「Slight agreement」であるため、有効な分類問題化は期待できない。そこで、能力の高い「6,7」と能力の低い「1,2」の人物の分類を行うことにした。すなわち3つの議論能力それぞれで、以下のように2通りのパターンの分類問題を考えることにした。

表 1: 議論テーマとトピック文書の例

議論テーマ	トピック文書
名古屋の人権	人権と絆が大切にされるまち。高齢者や障害者など誰もか不安なく自立して過ごせるまち。一人ひとりがいきいきと自分らしく生活できるまち。

表 2: 各分類問題の統計情報

能力	分類問題	κ	#Positive	$\frac{\#Positive}{\#Total}$
専門性	\mathcal{H}	0.33	31	0.16
	\mathcal{L}	0.51	44	0.24
コミュニケーション能力	\mathcal{H}	0.50	35	0.19
	\mathcal{L}	0.51	54	0.29
議論を仕切る能力	\mathcal{H}	0.59	31	0.16
	\mathcal{L}	0.54	67	0.36

分類問題 \mathcal{H}

アノテーションデータの中央値のうち、「7,6」を *Positive*、「5,4,3,2,1」を *Negative* とした能力の「高い」人物の分類問題

分類問題 \mathcal{L}

アノテーションデータの中央値のうち、「2,1」を *Positive*、「7,6,5,4,3」を *Negative* とした能力の「低い」人物の分類問題

このように分類した理由は、Web 議論掲示板において、能力の高い人物と低い人物を効果的に抽出するためである。特に議論能力の高い人物は、合意形成を志向する Web 議論の主導的役割となりうる他、コミュニケーション能力などが低い人物に関しては議論へ有益でない投稿や議論に相応しくない投稿をする可能性が高いことから、ファシリテータが注意を向ける必要がある。しかし、5 人のアノテーションデータに対して中央値を計算したところ、評価値が最も高い「7」や最も低い「1」の人物が非常に少数であったため、制約を緩めた。実際、分類問題 \mathcal{H} と分類問題 \mathcal{L} のように自由度を上げたところラベルの付与数と各能力の一致率、および参加者全体に対する *Positive* ラベル数の割合が表 2 のようになった、また一致率も「Fair agreement」以上となった。よって、本論文ではこの 2 つの分類パターンを分析する。

表 2 から、能力の高い人物が Web 議論掲示板において比較的少数であることを示しており、どの能力指標についても、能力が高い人物に比べて低い人物が多数となった。特に専門性や議論を仕切る能力の高い人物は約 16%しか存在しないことが明らかになった。

4.2 実験結果

本論文で提案した特徴素を用いて、アノテーションを行なった正解データを対象に学習させ、適合率、再現率、 F_β 値を評価する。今回は適合率および再現率を等しく評価するため、 $\beta = 1$ とする。ベースラインの素性には投稿単語数を用いた特徴素である **Bag of Words(BoW)** を用いる。また、比較手法となる提案手法を以下のように定義する。

- **IDF** - IDF による特徴素
- **AU** - 議論ユニットによる特徴素

- **TS** - トピック類似度による特徴素
- **EVIDF** - 関係グラフ中心性による特徴素

分類器には非線形の Support Vector Machine(SVM) を用いる。SVM のカーネルにはガウシアンカーネルを用いる。また、過学習を防ぐため、パラメータ C と γ をクロスバリデーション(交差検証)により学習する。クロスバリデーションには、議論トピックごとにグループを分け、一つのグループをテスト用、他の全てのグループを学習用データとする、Leave one group out(LOGO) と呼ばれる手法を用いた。

提案した特徴素 (IDF,AU,TS,EVIDF) が単体および組み合わせによって有効であるかを評価した。分類問題 \mathcal{H} と分類問題 \mathcal{L} それぞれにおける、3 つの能力の再現率、適合率、F 値のうち、特に値が高かった素性の組み合わせおよびその値を表 3 に示した。

4.3 議論

表 3 より、全ての項目で提案特徴素の値がベースライン(BoW) 以上となっていることから提案手法が有用であることが分かった。それぞれの分類問題に関して議論する。

分類問題 \mathcal{H}

実験結果より \mathcal{H} について以下のことが明らかになった。

- 専門性判別においては *AU* と *EVIDF* の適合率が顕著に向上しており素性として有用である。
- コミュニケーション能力においては *EVIDF* が特に効果的に作用した。
- 議論を仕切る能力では *IDF* や *AU* の語彙的特徴素が有用である。

つまり、専門性やコミュニケーションの能力が高い人物の抽出に関しては、関係グラフ中心性が効果的な特徴素であり、語彙の普及的性質に関して影響力を持つ人物が潜在的に高く評価される。

分類問題 \mathcal{L}

実験結果より \mathcal{L} について以下のことが明らかになった。

- 専門性判別においては *EVIDF* も効果的であるが、*IDF* や *TS* といった語彙的特徴量も効果的に作用した。
- コミュニケーション能力においては *EVIDF*、*IDF*、*AU* の組み合わせが効果的である。
- 議論を仕切る能力では語彙的特徴素全般が有用であり、*EVIDF* は有用ではなかった。

以上の結果より、 \mathcal{L} のような能力の低い利用者の抽出においては、単純な語彙的特徴が最も有用である。この理由は、アノテーションデータにおいて能力が「低い」とアノテーションされた参加者の多くが、投稿が少なく、非常に短文の投稿をしているような人物であったためである。つまり、発言量が少ない参加者は評価材料の不足から低く評価される傾向があり、発言文の BoW や IDF といった単純な素性が有用である。

表 3: 各分類問題における評価実験結果 (ベースラインおよび、適合率・再現率・F 値それぞれで最も高くなった素性のみ示す)

	素性					専門性			コミュニケーション能力			議論を仕切る能力		
	BoW	IDF	AU	TS	EVIDF	適合率	再現率	F 値	適合率	再現率	F 値	適合率	再現率	F 値
H	○					0.495	0.516	0.473	0.792	0.845	0.781	0.812	0.688	0.686
		○	○		○	0.539	0.547	0.516	0.756	0.898	0.800	0.763	0.634	0.610
		○				0.524	0.544	0.508	0.813	0.826	0.771	0.854	0.721	0.718
			○			0.555	0.692	0.605	0.762	0.788	0.765	0.799	0.908	0.825
					○	0.750	0.333	0.421	0.858	0.676	0.724	0.740	0.691	0.638
L	○					0.343	0.827	0.457	0.445	0.967	0.577	0.601	0.994	0.728
		○	○	○	○	0.410	0.927	0.541	0.450	0.927	0.574	0.648	0.971	0.757
		○	○		○	0.401	0.927	0.531	0.454	0.961	0.584	0.630	0.994	0.751
		○		○	○	0.422	0.927	0.552	0.463	0.927	0.591	0.658	0.983	0.770
		○	○			0.332	0.827	0.443	0.392	0.967	0.541	0.612	0.994	0.738
		○			○	0.411	0.960	0.549	0.467	0.961	0.598	0.625	0.994	0.749
			○		○	0.321	0.827	0.430	0.383	1.000	0.531	0.573	0.989	0.700
		○				0.332	0.827	0.443	0.389	0.967	0.536	0.592	0.994	0.724
			○			0.309	0.994	0.444	0.303	0.833	0.430	0.519	0.994	0.664
					○	0.319	0.994	0.460	0.344	0.800	0.463	0.521	0.921	0.645

5. 結論

本論文では、議論参加者の3つの能力に関する自動推定に焦点をあて、従来の語彙的特徴量を用いた素性抽出手法や関係グラフに基づく特徴量を提案した。語彙的特徴量については単語の難易度を仮定した IDF、品詞に着目した談話標識の生起数、そして TF-IDF のコサイン類似度を特徴量として提案した。さらに、関係グラフ中心性を提案し、Web 議論掲示板の新たな構造的特徴量を定義した。アノテーションでは実際の Web 議論掲示板で行われた議論データの参加者を7段階で評価した。実験の結果、提案手法が有用であることが明らかになった。特に、能力の高い人物の判別に関しては多くのケースで語彙の普及を仮定した関係グラフ中心性が有用であった。また、能力の低い人物の抽出では単純な語彙的特徴量が効果的であった。

今後はさらなる精度の向上を行うため、特徴抽出器や分類器に関する研究を進める必要がある。また、自動推定アルゴリズムを参加者に公開された場合、不当に評価を上げるなどといった行為が行われる可能性があるため、今後は、能力自動評価アルゴリズムへの不当行為に対する頑健性についても研究、実証が必要である。

謝辞

本研究は、JST、CREST の支援を受けたものである。

参考文献

[1] Takayuki Ito, Y. Imi, M. Sato, Takanori Ito, E. Hideshima. Incentive Mechanism for Managing Large-Scale Internet-Based Discussions on COLLAGREE. Posters of the 3rd Collective Intelligence Conference, 2015.

[2] Takayuki Ito, Y. Imi, Takanori Ito, E. Hideshima. COLLAGREE: A Facilitator-mediated Large-scale Consensus Support System. Collective Intelligence 2014, 2014.

[3] COLLAGREE, [Online] <http://collagree.com/?locale=en>

[4] M. Lippi, P. Torroni. Argumentation Mining: State of the Art and Emerging Trends, ACM Transactions on Internet Technology. 16(2): 10, 2016.

[5] R. Mochales, M-F. Moens. Argumentation mining: the detection, classification and structure of arguments in text. In ICAIL '09: Proceedings of the 12th International Conference on Artificial Intelligence and Law, pp. 98-107, ACM, 2009.

[6] J. Ecker-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.

[7] E. Barker, R. Gaizauskas, "Summarizing multi-party argumentative conversations in reader comment on news," in Proceedings of the 3rd Workshop on Argument Mining, Association for Computational Linguistics, 2016.

[8] S.A. Golder, J. Donath, Social Roles in Electronic Communities. Presented at Association of Internet Researchers 5.0 (2004).

[9] S. Joo, H. Takeda. Analysis of Discussion Page in Wikipedia based on User's Discussion Ability. The 26th Annual Conference of the Japanese Society for Artificial Intelligence. IC2-R-5-1, 2012.

[10] R. Rinott, L. Dankin, C. Alzate, M. M. Khapra, E. Aharoni, N. Slonim. Show Me Your Evidence - an Automatic Method for Context Dependent Evidence Detection. In 2015 Conference on Empirical Methods in Natural Language Processing, 2015.

[11] L.C. Freeman, Centrality in networks: I. Conceptual clarification. Social Networks 1. 215-239, 1979.

[12] P. Bonacich, Power and centrality: a family of measures. American Journal of Sociology 92. 1170-1182, 1987.

[13] N. Matsumura, Y. Ohsawa, M. Ishizuka. Influence Diffusion Model in Text-Based Communication. Transactions of the Japanese Society for Artificial Intelligence Vol. 17 (2002) No. 3 Technical Papers / Special Issue: Web Intelligence P 259-267. Vol17, No.3, 259-267, 2002.

[14] N. Matsumura, Y. Ohsawa, M. Ishizuka. Profiling participants in online-community based on influence diffusion model, Transactions of the Japanese Society for Artificial Intelligence, Vol. 18 No. 4, pp. 165- 72, 2003.

[15] J.A. Krosnick, S. Presser. Question and questionnaire design. In: P.V. Marsden and J.D. Wright (eds.) Handbook of Survey Research, 2nd edn. Bingley, UK: Emerald, pp. 263-314, 2010.

[16] J.L. Fleiss. Measuring nominal scale agreement among many raters. Psychological Bulletin, Vol 76(5), 378-382, 1971.

[17] J. Richard Landis, Gary G. Koch. The Measurement of Observer Agreement for Categorical Data. Biometrics Vol.33, No.1, 159-174, 1977.