

議論システムにおける賛成／反対意見の生成手法の検討

Study on generating supporting/opposing opinion in debate system

古舞千暁^{1*} 有木康雄¹ 滝口哲也¹

Furumai Kazuaki¹ Arika Yasuo¹ Takiguchi Tetsuya¹

¹ 神戸大学システム情報学研究科

¹ Graduate School of System Informatics, Kobe University, Department of Information Science

Abstract: In recent years, the development of information networks and the diversification of platforms have caused information to flood, so the cost to obtain knowledge users want has increased. Therefore, interfaces using dialogue have been drawing attention in transit guidance systems and information guidance systems and so on are studied. In this study, we constructed a debate system to support users understanding about news. The system aims to deepen the understanding of the news through debate by presenting the counter opinion against the user's opinion. In the debate system, we estimate the user's claim and reason using Convolutional Neural Networks, and generate counterargument opinions based on that. In this paper, we investigate a vectorization method of utterance used for calculating similarity, and conducted basic experiments on opinion generation using it.

1 はじめに

本研究では、ニュースに関するユーザの主張に対して、システム反論意見を提示して、議論をすることで、ニュースに関する理解を深めることを目的としている。本稿では、意見の類似度計算に用いる意見文のベクトル化手法の検討と、そのベクトル化手法を用いた反論生成手法の提案を行う。

を推定し、推定したユーザの主張に対して反論する意見（ユーザの主張が賛成なら反対の意見）をデータベース内で探し、発話内容が最も類似した意見を反論意見として提示する。もし類似度が十分に高い意見が見つからなかった場合は、ユーザ発話と類似する意見を用いて反論意見を新たに生成し、議論を行う。あらかじめ、WWW上で議題に対する意見を井上らの手法 [1] で自動収集し、データベースを構築している。

2 議論システム

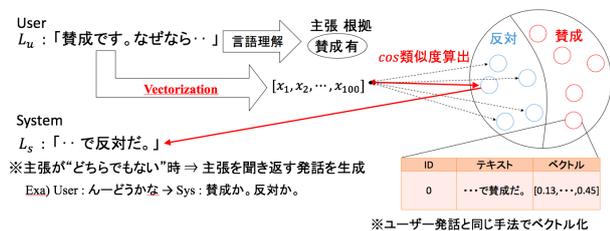


図 1: Debate management

議論システムの概略を図 1 に示す。まず、ユーザの主張 (賛成/反対/どちらとも言えない) と根拠 (有/無)

2.1 意見のベクトル化手法

本システムでは、ユーザ発話と収集した意見群をベクトルとして扱う。推定されたユーザの主張ベクトルが賛成の場合は、収集した反対意見群の中から cos 類似度が最も高いものを、反論意見候補として選択する。cos 類似度の計算に用いるベクトル化手法については、以下の 4 つの手法を比較・検討した。

- TF-IDF に基づいた Bog-of-Words によるベクトル化
- Word2vec を用いた特徴ベクトル平均によるベクトル化
- Doc2vec によるベクトル化
- Sparse Composite Document Vectors (SCDV)[2] を用いた特徴ベクトル平均によるベクトル化

*連絡先：神戸大学システム情報学研究科
〒 657-8501 兵庫県神戸市灘区六甲台町 1-1 自然科学
総合研究棟 3 号館 805 号室
E-mail: kazuaki.furumai@stu.kobe-u.ac.jp

2.2 類似意見を用いた反論生成

ユーザ意見と類似した意見がデータベース内に見つからない (\cos 類似度が低い) 場合には、図 2 のようなモデルを用いて反論意見を生成する。ユーザ意見に \cos

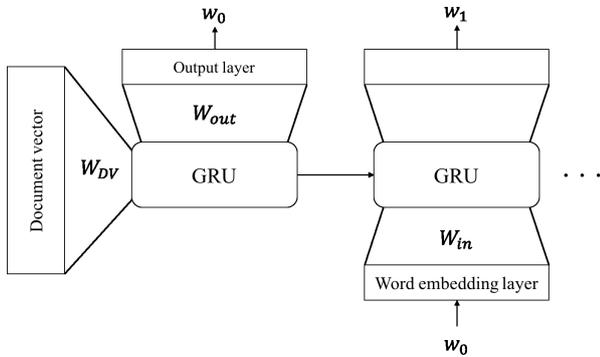


図 2: Opinion generator

類似度が近い意見ベクトル N 個の平均ベクトルを初期の隠れ状態に入力し、開始記号から文章の生成を行う。ここで、 w_t は各単語、 W_{in} , W_{out} , W_{DV} はそれぞれ、単語の埋め込み層から GRU への重み、GRU から出力層への重み、隠れ層の初期値としての意見ベクトルの入力重みである。

3 実験

ベクトル化手法の検討については、それぞれの手法でベクトル化された意見群を用いて LightGBM による分類実験を行なった。反論生成については、分類実験で最も精度の良かった手法によってベクトル化された単一の意見を、初期の隠れ状態への入力に用いて、BLEU スコアによる評価を行なった。収集した意見数については表 1 の通りである。

表 1: Number of opinions

	賛成	反対
カジノ法案	38	68
死刑制度	158	183
原発再稼働	54	220

721 個の意見のうち、547 個で LightGBM[3] の学習を行い、174 個の意見を分類した。結果を表 2 に示す。

SCDV が最も良い精度を示した為、次に、SCDV によるベクトル化を用いた文章生成の実験を行なった。本稿では、単一意見のベクトルから文章を復元できるかについて BLEU スコアを用いた実験を行なった。Document vector を用いず、先頭単語を入力として文末記号が現

表 2: Document classification with LightGBM

	Accuracy	Precision	recall	f1
BoW	0.867	0.864	0.864	0.863
Word2vec	0.878	0.874	0.878	0.874
Doc2vec	0.626	0.623	0.626	0.615
SCDV	0.952	0.953	0.952	0.952

表 3: BLEU score

	BLEU-1	BLEU-2	BLEU-3	BLEU-4
SCDV-RNN	13.82	4.41	1.51	0.52
RNN	6.51	2.2	0.70	0.21

れるまで生成を行う RNN による言語モデルとの比較を行なった結果を、表 3 に示す。

BLEU-1~4 は、それぞれ、1-gram のみ、1-gram と 2-gram、1-gram から 3-gram、1-gram から 4-gram を用いて計算したものである。

4 考察・今後の取り組み

意見のベクトル化手法については、SCDV が最も良い精度を示した。言語モデルによる文章生成についても、初期の隠れ状態への入力に SCDV を用いた提案手法の方が良い結果を示したが、まだスコアは高いとは言えず、モデル構造の見直しを行い精度向上を目指す。また、複数意見のベクトルを用いた意見生成実験を行う必要がある。

謝辞

本研究の一部は、JSPS 科研費 JP17K00236 の助成を受けたものである。

参考文献

- [1] 井上 結衣, 藤井 淳: Web 世論からの意見抽出と賛否に基づく分類, 言語処理学会論文集, p. 364-367 (2008)
- [2] Dheeraj Mekala et al.: SCDV : Sparse Composite Document Vectors using soft clustering over distributional representations, *Proc of EMNLP*, 2017
- [3] Guolin Ke et al.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *Proc of NIPS*, pp. 3146-3154, 2017