

コンテンツホール検索のための掲示板対話の解析

荒牧英治* 阿辺川武** 村上陽平*** 灘本 明代****

*東京大学 知の構造化センター **東京大学教育学研究科

独立行政法人 情報通信研究機構 *甲南大学知能情報学部

eiji.aramaki@gmail.com abekawa@p.u-tokyo.ac.jp

yohei@nict.go.jp nadamoto@konan-u.ac.jp

BBS Dialog Analysis for Content Hole Search

Eiji ARAMAKI* Takeshi ABEKAW**

Yohei MURAKAMI*** Akiyo NADAMOTO****

*University of Tokyo **University of Tokyo

NICT *Konan university

1. はじめに

現在、SNSやBlogのようなコミュニティ型コンテンツがインターネット上に多数存在する。このようなコミュニティ型コンテンツの場合、コミュニティ内での議論に集中するあまり視野が狭くなり、議論のテーマに対する全体像が見えなくなってしまう危険性がある。そこで本研究では、コミュニティ型コンテンツにおいて、ユーザが気付いていない情報を提示する事を目的としコミュニティ型コンテンツのコンテンツホール検索を提案する。我々は、この「ユーザが気付いていない情報を探す」ことをコンテンツホール検索と呼ぶ。

これまで我々はコミュニティ型コンテンツのコンテンツホール検索を行うために、以下の手順を提案してきた。

- (1) コミュニティ型コンテンツからのテーマの抽出
- (2) Web空間における視点情報の抽出
- (3) コミュニティ型コンテンツの視点抽出
- (4) Web空間における視点構造とコミュニティ内の視点構造を比較しその差分情報であるコンテンツホールを抽出
- (5) 抽出されたコンテンツホールの提示

これまで、上記手順(2)を提案してきた[1]。本研究では、上記手順(3)の基盤技術となるコミュニティ型コンテンツの対話解析に焦点を当てる。ここで問題となるのは、SNSやBlogなどのコミュニティ型コンテンツは通常のWebコンテンツと異なり、対話形式で構成される点である。さらに、その対話は通常の音声対話と異なり、不特定多数人によって複数のトピックが同時に議論される。このようなテキストを扱うためには、どのコメントがどのコメントと対応しているのか、その対応関係を得る処理が必須となる。

そこで、本論文では、2つのコメント間が対応しているかどうかを識別するタスクに挑戦する。このタスクにおいて我々は対応するコメント間には次の2つの関連性があると仮定している。まず、2つのコメントが内容的に類似している場合で本論文ではこれを**内容的関連性**と呼ぶ。次に、2つのコメントが応答関係になっている**機能的**な対応である。例えば、「なぜ～」といったコメントに対する応答は「～だから」というコメントで応答することが考えられる。このようなコメント間で対応する表現による関連性を本論文では**機能的関連性**と呼ぶ。

2. 二つの関連性の定式化

前者、**内容的関連性**は、文同士の類似度と近い概念である。我々は先行研究で提案されたweb上での単語の共起にもとづく類似度(WEBPMI[2])を用いて、これを計算する。後者、**機能的関連性**を得るためには、WEBPMIと同様に対応するコメント間での共起を用いる。この共起を計算するためには、大量の対応するコメントのペアが必要となるが、我々は人手に作成したパターンにより、121,699 コメントペアを収集し、共起するフレーズ対を求めた。

3. 実験

実験においては、二つのコメントを提示し、それらが応答関係にあるかどうかを判定するタスクを用い、各関連性による精度と人間の精度を調査した。その結果、人間の精度は70~79%であり、内容的関連性の精度は61%、機能的関連性は65%の精度を得ることができ、対話の応答関係において機能的関連性が重要な手がかりとなることが示唆された。

4. おわりに

本論文では、コミュニティ型コンテンツの視点抽出のための対話解析を行った。今後の展望は下記のとおりである。

- **コミュニティ型コンテンツからの視点情報の抽出**：本論文にて対話解析の結果を用いて、コミュニティ型コンテンツからの視点情報の抽出及び、視点構造グラフの生成方法。
- **コンテンツホールの抽出**：Web空間における視点構造とコミュニティ内の視点構造を比較しその差分情報であるコンテンツホールを抽出。

参考文献

- 1) 灘本明代, 阿辺川武, 荒牧英治, 村上陽平, コミュニティ型コンテンツのコンテンツホール抽出手法の提案, 日本データベース学会 Letters, vol6 No2, pp29-32, 2007.
- 2) Danushka Bollegala and Yutaka Matsuo and Mitsuru Ishizuka, "Measuring Semantic Similarity between Words Using Web Search Engines", Proceedings of 16th International World Wide Web Conference (WWW2007), pp.757-766, 2007.