

# 情報交流サイトにおける会話分析に基づく情報推薦システムの提案

府川 和樹<sup>†</sup> 王 元元<sup>††</sup> 河合由起子<sup>†</sup> 角谷 和俊<sup>†††</sup>

<sup>†</sup> 京都産業大学コンピュータ理工学部 〒603-8555 京都府京都市北区上賀茂本山

<sup>††</sup> 山口大学大学院創成科学研究科 〒755-8611 山口県宇部市常盤台 2-16-1

<sup>†††</sup> 関西学院大学総合政策学部メディア情報学科 〒669-1337 兵庫県三田市学園 2 丁目 1 番地

E-mail: <sup>†</sup>{g1445142,kawai}@cc.kyoto-su.ac.jp, <sup>††</sup>y.wang@yamaguchi-u.ac.jp, <sup>†††</sup>sumiya@kwansei.ac.jp

あらまし 本研究では、情報交流サイトにおけるユーザ発言の活性化を目指し、ユーザの積極的な質問を促進させる情報推薦システムを提案する。具体的には、ユーザの質問に対して、応答時に質問者の潜在的質問を推定し、応答結果に潜在的質問内容に関する質問を付与し返答する。また、応答システムはユーザの質問に関して、関心の高い他ユーザを発見し、そのユーザが返答可能な形式に拡張し、質問を転送する。これによりユーザの発言が増加し、情報交流サイトの質の向上も期待できる。本論文では、提案する潜在的質問の推定手法に基づき情報推薦システムを構築し、検討する。

キーワード 機械学習, 会話分析, 情報交流サイト, 深層学習, 応答システム

## 1. はじめに

近年、Yahoo!知恵袋をはじめとした QA サイト情報交流サイトが普及し、多くの人々が情報交流サイトを通じて、オンライン上でつながりを作っている。情報交流サイト上では、ユーザ同士の質問・回答を通じて、コミュニケーションできるが、質問者には「何を聞けばいいのか」、「どのように聞けばいいのか」、「誰に聞けばいいのか」という知識の差が存在し、これにより円滑なコミュニケーションが進まないことがある。そこで、本研究では、情報交流サイトにおけるコミュニケーション促進を目的にユーザの積極的な質問促進、それに対する適切な応答結果を返答することを目指す。

本論文では質問促進として、具体的には、ユーザの質問情報に対して、応答時に質問者の潜在的質問を推定し、応答結果に潜在的質問内容に関する質問を推薦する応答システムを提案する(図 2)。推薦される潜在的質問を提示することで、ユーザの積極的な質問を促進させることが期待できる。また、提案する応答システムはユーザの質問に関して、関心の高い他ユーザを発見し、そのユーザに質問を転送する。これにより質問を転送されたユーザの返信が期待でき、ユーザの積極的な質問促進と合わせて発言が増加し、情報交流サイトにおけるユーザ満足度の向上や対話の活性化につながる。

## 2. 関連研究

チャットや QA サイトにおける対話分析に関する研究として、田中ら [7] や原田ら [10] は、ユーザ間の対話からユーザの発話意図を読み取り、システムに自動的に問い合わせを行うシステムを提案している。また、松尾ら [9] や Mladenic [10] は、ユーザの履歴情報から興味やキーワードを抽出し、ブラウジングを支援するシステムを提案している。また、ユーザの振る舞い情報に基づく情報推薦として、Liu ら [6] は、ニュースサイトにおいて、ユーザの振る舞いに応じたページを推薦するシステムを開

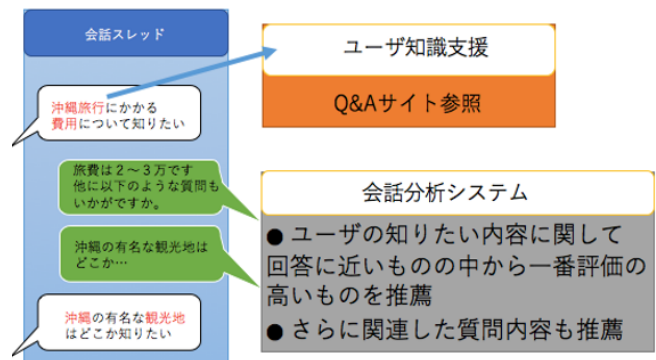


図 1 質問促進のための情報推薦システム

発している。これらはいずれもユーザの振る舞いに着目しており、本研究では質問間および回答間の類似性を学習する点、さらにそれらを新たな質問として回答と合わせてユーザへ返信する点で異なる。

ユーザへの質問回答は知識支援技術とも関連しており、Sakura ら [3]、武吉ら [4]、さらに Zhou ら [5] は、知識支援として Web サイトの文章に自動でリンク生成するシステムを開発している。本研究では、知識支援につなげるために回答に新たな質問を加える点、さらにそれに基づいた会話促進を目的としてユーザを推薦する点が異なる。

## 3. システム概要

提案する質問促進のための情報推薦システムは、会話となる質問に対する回答との関係性を分析する。分析のため、会話に出現する質問キーワードと関連する質問と回答のペアを QA サイトより取得する。それら質問から会話の質問と類似性の高い質問と回答のペアを取得し、回答と回答集合との類似性を算出し、類似性の高い回答に対する質問を取得する。また、質問も同様に、質問集合との類似性の高い質問を取得し、取得した 2

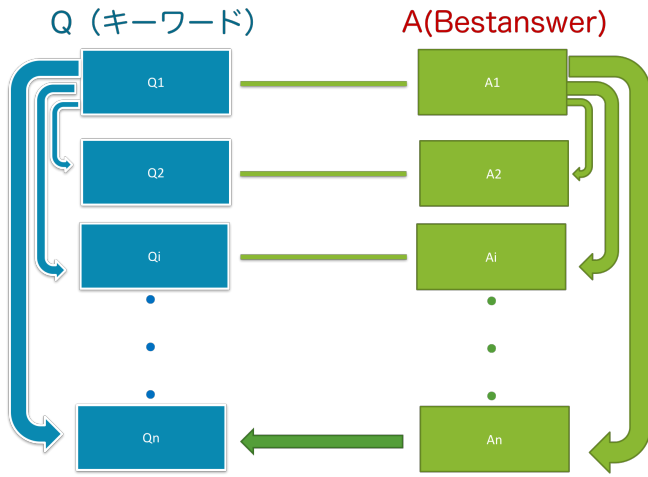


図2 質問とベストアンサーの類似性に基づく関連質問抽出

つの質問を会話の質問に対する関連質問として回答と合わせてユーザへ返信する(図2)。

### 3.1 質問抽出と応答生成

まず、ユーザの質問(発言)に対して、Yahoo!WebAPIのキーフレーズ抽出<sup>(注1)</sup>を用いて重要度付き単語を特徴語として抽出する。次に、各特徴語の重要度の高い特徴語を質問キーワードとしてYahoo!知恵袋APIより、質問およびベストアンサーのペアをN件取得する。取得した最上位のベストアンサー $A_1$ をユーザの質問に対する回答とする。

次に取得したN件の回答から再度キーフレーズ抽出より単語をM個取得し、下記のTF-IDF手法に基づき各回答を1つのドキュメントとして、質問キーワード以外の各単語に重みを付与する。

$$TF = \frac{\text{単語 } i \text{ の出現回数}}{\text{単語総数}}$$

$$IDF = \frac{\text{回答総数 } N}{\text{単語 } i \text{ が出現した回答数}}$$

この重みを要素とする各ベストアンサー $A_i$ のベクトル $va_i = (w_1, \dots, w_i, \dots, w_m)$ を生成し、各ベストアンサーベクトル間の類似度を下記より算出する。

$$sim(a_i, a_j) = \frac{\sum_{j=1}^{|M|} a_i \cdot a_j}{\sqrt{\sum_{i=1}^{|M|} (a_i)^2} \cdot \sqrt{\sum_{j=1}^{|M|} (a_j)^2}} \quad (1)$$

$sim(a_i, a_j)$ の最も高い $A_j$ のベストアンサーに対する質問 $Q_j$ を関連質問として抽出し、 $A_1$ と合わせて応答としてユーザへ返信する。

### 3.2 会話促進のためのリンク生成

ユーザ間の類似度に基づいたユーザの投稿情報の語彙にリンク情報を自動的に張り付ける。ここでリンク生成を行うリンク先として、知識支援情報と興味喚起支援情報がある。知識支援被リンク情報は、受信者の不明な語彙に対する情報を補完することを目的としており、受信者のユーザ特性として抽出した特

徴語のうち、重要度の低い語彙を選択する。選択された語彙を用いた検索結果ページとWikipediaページを被リンク情報とする。また、受信者と関係性が高い友人を式(2)より発見し、それら友人の投稿情報に語彙が含まれている場合に被リンク情報とする。

$$Sim(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{|V|} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{|V|} (x_i)^2} \cdot \sqrt{\sum_{i=1}^{|V|} (y_i)^2}} \quad (2)$$

$\vec{x}$ はユーザ $x$ の特徴ベクトル、 $\vec{y}$ はユーザ $y$ の特徴ベクトル、 $|V|$ は特徴ベクトルの次元数である。

受信ユーザが発信ユーザへ周知させたい興味喚起支援被リンク情報は、受信者の興味喚起による会話促進を目的としており、先程とは逆に重要度の高い語彙を選択する。そして、この場合は受信者ではなく、この語彙を話題にした発信者と関係性が高い他の友人(ただし、受信者ともネットワークがある)を式(2)より発見し、語彙が含まれている投稿情報を受信者の被リンク情報として提示する。

たとえば、ユーザA, B, C, Dがあり、(1)AとB、(2)BとC、(3)BとDが友人関係であり、AがBへ情報を発信した場合、(2)と(3)の友人関係により、受信者Bにとって知識情報は、Bと関係あるCとDとの関係性が高い方が発信している同じ話題を提供する。ここでは、式(2)より $Sim(\vec{B}, \vec{C}) < Sim(\vec{B}, \vec{D})$ と仮定しているため、Bには、Dが投稿した情報のうち、Aの発信情報から抽出された語彙を含む重要度の高い投稿情報、検索結果ページやWikipediaページをリンク先とする。このことより、受信者Bの知識を深めることができる。一方、受信者Bにとって発信者Aへ提供したくなる情報は、受信者Bと関係あるCとDのうち、発信者Aとも関係が深い方が発信している同じ話題である。ここでは、式(1)より $Sim(\vec{A}, \vec{C}) > Sim(\vec{A}, \vec{D})$ と仮定しているため、Bには、Cが投稿した情報のうち、Aの発信情報から抽出された語彙を含む重要度の高い投稿情報や講義映像をリンク先とする。このことより、発信者Aと受信者Bの友人Cとの会話を促進することができる。

## 4. 環境構築

本論文では、実際のデータを用いて情報交換サイトにおける質問の特徴語抽出と関連質問抽出を検証するための環境構築を行った。

本稿では、Yahoo!JAPANが提供する知恵袋WebAPI<sup>(注2)</sup>を用いて、質問応答データを取得する。今回、車に関する質問応答を対象とし、2,000件の質問応答データを取得した。検索キーワードより、質問とベストアンサーのペアを100件取得可能である。なお、形態素解析にはMeCabを用いた。また、ベストアンサー間のコサイン類似度を算出し、類似度の最も高いベストアンサーを抽出した。その結果、提案手法に基づき質問に対して類似する質問の抽出を確認した。今後、取得したデータに対して提案手法により推薦された新たな質問提示について検証を行う予定である。また、車以外のカテゴリを含めた検証が必

(注1) : <http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.htm>(注2) : <http://developer.yahoo.co.jp/>

要である。

## 5. ま と め

本論文では、情報交流サイトにおける質問促進として、ユーザの質問情報に対して、応答時に質問者の潜在的質問を推定し、応答結果に潜在的質問内容に関する質問を推薦する応答システムを提案した。また、ユーザの質問に関して、関心の高い他ユーザを発見し、会話促進のためのりんくせいせいそのユーザに返答可能な形式を拡張した。評価実験では、ユーザの質問情報から、*TF-IDF*を用いて各質問の特徴語を抽出し、*LSI*に基づき抽出した関連質問内容を確認した。今後、質問を推薦する応答システムの有用性を検証する予定である。

## 謝 辞

本研究の一部は、JSPS 科研費 15K00162, 16H01722 の助成を受けたものである。ここに記して謝意を表す。

## 文 献

- [1] 長谷川聡, 安井明代, 山口宗芳, “SNS の教育利用とソーシャルラーニング”, 名古屋文理大学紀要 第 13 号 (2013), pp. 51-58, 2013.
- [2] 村上正行, “ソーシャルメディアを活用した大学教育”, リメディアル教育研究, 7(2), 189-195, 2012.
- [3] Tomoya Sakusa, Motomichi Toyama, “Supporting Web Content Development using Web Index, Proc. of the 19th International Database Engineering & Applications Symposium”, pp. 204-205, 2015.
- [4] 武吉朋也, 服部元, 小野智弘, 滝嶋康弘, “Web ページのリンク自動生成方式の提案と評価”, DEIM Forum 2010, F9-1, 2010.
- [5] Dong Zhou, Mark Truran, Tim Brailsford, Helen Ashman, Amir Pourabdollah, “LLAMA-B: Automatic Hyperlink Authoring in the Blogosphere”, Proc. of ACM HT '08, pp. 133-137, 2008.
- [6] Jiahui Liu, Peter Dolan, Elin Ronby Pedersen, “Personalized news recommendation based on click behavior”, Proc. of the International Conference on Intelligent User Interfaces, pp.31-40, 2010.
- [7] 田中貴志, 小山聡, 角谷和俊, 田中克己, “ユーザ間の対話からのユーザプロフィール抽出と動的情報探索”, 電子情報通信学会技術研究報告 102(207), pp. 31-36, 2002.
- [8] 原田修平, 丸山広, 高嶋章雄, 中村太一, “チャットを用いたグループワークにおける人の行動の分析方法の提案”, 電子情報通信学会技術研究報告 111(282), pp. 85- 90, 2011.
- [9] 松尾豊, 福田隼人, 石塚満, “ユーザ個人の閲覧履歴からのキーワード抽出によるブラウジング支援”, 人工知能学会論文誌 18 巻 4 号 E, pp. 203-211, 2003.
- [10] Dunja Mladenic, “Using Text Learning to help Web browsing”, Proc. of the 9th International Conference on Human-Computer Interaction, 2001.