

# オンライン教育におけるユーザ特性を考慮したコミュニティ生成の提案

林 利憲<sup>†</sup> 王 元元<sup>††</sup> 河合由起子<sup>†††</sup> 角谷 和俊<sup>†</sup>

<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

<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> 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

E-mail: <sup>†</sup>{anc93184,sumiya}@kwansei.ac.jp, <sup>††</sup>y.wang@yamaguchi-u.ac.jp, <sup>†††</sup>kawai@cc.kyoto-su.ac.jp

あらまし 本研究では、オンライン教育における利用ユーザ間のコミュニケーションを促進することで、学習の質を向上させることを目的とし、会話受信時にユーザの知識や興味に基づきリアルタイムでテキストに多様な注釈（リンク）自動付与手法を提案する。被リンク情報を提供することで、受信ユーザに対して相手ユーザへの返答を促進させることができ、会話の活性化と議論の理解向上につながることを期待できる。オンライン教育におけるコミュニケーション促進として、1) 受信側ユーザの知識支援ならびに、2) 受信ユーザが発信ユーザへ周知させたい興味喚起による知識提供支援となる2種類のリンクを生成する。知識支援に対する被リンク情報は、受信ユーザにとって不明な語彙に対する検索結果ページや Wikipedia ページ、他ユーザの会話とする。また、受信ユーザが発信ユーザへ周知させたい被リンク情報は、その語彙を話題にした他の友人達の会話や関連講義映像とする。これにより、オンライン教育を利用しているユーザそれぞれに適応した知識提供を行うことができる。本稿では、Facebook の教育講座ページを対象に、受講ユーザ特性に基づいた特徴語抽出に関して検証する。

キーワード リアルタイムリンク生成, ユーザ特性, 特徴語抽出, コミュニケーション, オンライン教育

## Community Generation based on User Characteristics in Online Education

Toshinori HAYASHI<sup>†</sup>, Yuanyuan WANG<sup>††</sup>, Yukiko KAWAI<sup>†††</sup>, and Kazutoshi SUMIYA<sup>†</sup>

<sup>†</sup> Kwansei Gakuin University 2-1, Gakuen, Sanda-shi, Hyogo, 669-1337, Japan

<sup>††</sup> Yamatuchi University 2-16-1, Tokiwadai, Ube, Yamaguchi, 755-8611, Japan

<sup>†††</sup> Kyoto Sangyo University Motoyama, Kamigamo, Kita-ku, Kyoto, 603-8555 Japan

E-mail: <sup>†</sup>{anc93184,sumiya}@kwansei.ac.jp, <sup>††</sup>y.wang@yamaguchi-u.ac.jp, <sup>†††</sup>kawai@cc.kyoto-su.ac.jp

### 1. はじめに

近年、オンライン教育プラットフォームが普及し、学校に行かないと受けることができなかった授業がインターネットを通じて、誰でもいつでも好きな時に受けれるようになった。しかし、オンライン教育における受講ユーザのモチベーション維持が難しいという点がある。また、Facebook をはじめとした Social Network Service(以下 SNS) が普及し、受講中に多くの人が SNS を通じたオンライン上でインタラクティブな学習が可能である。長谷川ら [1] と村上 [2] の研究では、SNS を大学教育の実践に活用する事例を紹介した。オンライン教育では、受講ユーザ同士が投稿スレッドを通じて、コミュニケーションできるが、個人それぞれの知識の差で円滑なコミュニケーションが進まないということがある。そこで、本研究では、オンライン教育におけるユーザ間のコミュニケーションの低さに着目

する。オンライン教育におけるユーザの投稿情報から各ユーザの特性を分析し、投稿した文章にユーザ特性に基づいたリンクを生成する手法を提案する。被リンク情報を閲覧することで、相手への返答を促進させることができ、会話支援と理解向上につながることを期待される。

本稿ではオンライン教育における会話促進として、1) 受信側ユーザの知識支援ならびに、2) 受信ユーザが発信ユーザへ周知させたい興味喚起による知識提供支援となるリンク生成提供手法について述べる。1) 知識支援に対する被リンク情報は、受信ユーザにとって不明な語彙に対する検索結果ページや Wikipedia ページ、他ユーザの SNS の会話とする。また、2) 受信ユーザが発信ユーザへ周知させたい被リンク情報は、その語彙を話題にした他の友人達の会話や関連講義映像とする。本稿では、特に Facebook の教育講座ページを対象に、受講ユーザ特性に基づいた特徴語抽出に関して検証する。

第2章では、オンライン教育の現状と関連研究について述べ、第3章では、提案する本システムの構成、ユーザ特性分析と、リンク先である知識支援被リンク情報、興味喚起支援被リンク情報について述べる。そして第4章では、検証とその考察を述べ、最後に第5章で本研究のまとめと今後の課題について述べる。

## 2. オンライン教育の現状と関連研究

### 2.1 オンライン教育

オンライン教育とは、教育をインターネット上で受けることである。アメリカでMOOC(Massive Open Online Course)と呼ばれるオンライン教育システムが登場した。有名なMOOCのプラットフォーム Web サイトとして Udacity<sup>(注1)</sup>、Coursera<sup>(注2)</sup>、edX<sup>(注3)</sup>などが挙げられ、それらと提携する大学は世界中で増えている。日本で、Web 上のオンライン授業としてよく使われているものとしては、school<sup>(注4)</sup>や Moodle<sup>(注5)</sup>などが挙げられる。一方、MOOC とは異なり、大学で行う講義は取扱わないが YouTube ベースの教育コンテンツ提供サービスとして世界に広がり続けている Khan Academy<sup>(注6)</sup>というものもある。現在のオンライン教育において、一番の難点は、生徒のモチベーションを保つことである。授業が無償化されたことにより、視聴ユーザが価値を見出さなくなってしまうという問題がある。2013年11月イギリスのファイナンシャルタイムズ紙の11月25日の記事によると、Coursera では、授業を最後まで受講したのは、数十万人のみで、修了率は7~9%に留まったと示されている。本研究で、生徒のモチベーション維持という問題解決手段の一つとして、受講ユーザ間のコミュニケーション促進を目的とするシステムを提案する。それが、ユーザ特性を活かした自動被リンクシステムである。システム構成に関して、第3章で述べる。

### 2.2 関連研究

ユーザの知識支援として、Sakusaら[3]、武吉ら[4]とZhouら[5]は、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た。本研究では、Webではなく、SNS上のユーザの投稿情報を分析することで、投稿情報とユーザごとに異なるリンク生成を行う点が異なる。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発した。本研究では、各ユーザの投稿情報とユーザ間の関係性に基づきユーザ特性を分析している点が異なり、またそれに基づいた会話促進を目的としている。

リンク自動生成に関連する研究として、田中ら[7]と原田ら[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た。松

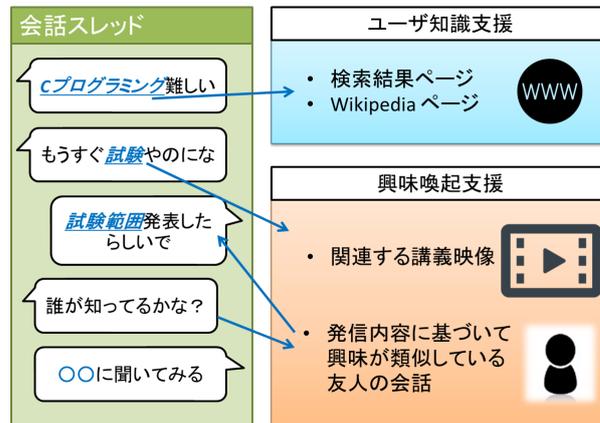


図1 システム構成

尾ら[9]とMladenec[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た。本研究のリンク生成においても、オンライン教育におけるユーザの投稿履歴からユーザの意図や興味を抽出する。

## 3. システム構成

提案するシステムはオンライン教育における各ユーザの投稿情報から各ユーザの特性を分析し、各ユーザ特性によって語彙を選択し、その語彙にリンクを生成する。まず、各ユーザごとに、投稿情報から語彙の重要度を算出する。次に、ユーザ間の関係性に基づき、重要度の高いあるいは低い語彙をリンク対象とする。リンク先は1) 知識支援リンク情報と2) 興味喚起支援被リンク情報に分けられる(図1)。

### 3.1 ユーザ特性分析

まず、Yahoo!WebAPIのキーワード抽出<sup>(注7)</sup>を用いてユーザの投稿情報から重要度付き単語を特徴語として抽出する。次に、各特徴語の重要度を用いて特徴語を含むユーザの投稿件数により特徴語の重要度平均値を算出し、TF-IDF手法に基づき、各投稿情報を1つのドキュメントとしてユーザごとに下記の式より、各特徴語に重みを付与する。

$$TF = \frac{\text{単語 } i \text{ の重要度}}{\text{単語 } i \text{ が出現した投稿数}}$$

$$IDF = \frac{\text{総投稿数}}{\text{単語 } i \text{ が出現した投稿数}}$$

また、SNSによっては、「いいね」や「シェア」ボタンが用意されており、ユーザが投稿した情報に対して、他ユーザが気に入れば「いいね」、他ユーザにとってもっと広めたいと感じるのであれば「シェア」をする仕組みとなっている。この場合において、算出された特徴語の重みに「いいね」数と「シェア」数を考慮した算出方法を次章で検証する。

### 3.2 知識支援および興味喚起による会話促進支援

ユーザ特性を用いて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(図1)。知識支援被リンク情報は、受信者の不明な

(注1) : <https://www.udacity.com/>

(注2) : <https://www.coursera.org/>

(注3) : <https://www.edx.org/>

(注4) : <https://school.jp/guest>

(注5) : <https://moodle.org/>

(注6) : <https://school.jp/guest>

(注7) : <http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html>

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 (1) 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。図 1 では、受信者は「C プログラミング」に関して知識不足と判定され、リンクが生成されている。

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

$\vec{x}$  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、 $|\mathcal{V}|$  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 (1) 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。図 1 では、「試験」に関して、共通の友人が投稿した内容にリンクが生成され、そのリンク先の内容を参照して返答している。

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 (1) 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 へ提供しなくなる情報は、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. 検 証

本稿では、実際のデータを用いてオンライン教育におけるユーザ特性抽出を検証する。ここでは、Facebook API<sup>(注8)</sup>を用いて、以下の Facebook の投稿情報を取得した。

- A: 小学生からはじめるわくわくプログラミングの投稿直近 50 件
- B: アプリ開発で学ぶオンラインプログラミング学習サービスの投稿直近 50 件

A はプログラミング初級学習者向けの Facebook 公開ページ、

表 1 A から抽出した特徴語上位 15 件と下位 15 件

ページ	特徴語 (上位) 興味喚起支援	(下位) 知識提供
A (1)	茂木健一郎, 日経ソフトウェア, 討論会, 東京大学大学院情報学環, 倉本大資, 先進性, 自己表現, 産業競争力会議, ロボット・プログラミング教材, プログラミング必修化, ニュースウィーク日本語版, トライリンガル, スタディーノ, すがやみつる, アカウント一つひとつ	プログラミング, ご紹介, 記事, プログラミング教育, 著者, 10 冊, 3 月 21, 5 歳, 人, 3 種類, Why, 以下, 気, 土, 写真
A (2)	Hour of Code Japan, 東京大学大学院情報学環, プログラミング必修化, スクラッチ デイ, 橋本良明教授, パソコン離れ, 数行, スタディーノ, Show & Tell, Touch & Try, Code.org, ワールドビジネスサテライト, 自己表現, Minecraft EDU, ロボットプログラミング	プログラミング, ご紹介, 記事, プログラミング教育, 小学生, 5 歳, 著者, 人, 3 種類, 10 冊, 3 月 21, 時代, 背景, 世代, 向上
A (3)	スクラッチ デイ, 東京大学大学院情報学環, Show & Tell, Touch & Try, 橋本良明教授, Hour of Code Japan, パソコン離れ, スタディーノ, ロボットプログラミング, Code.org, ワールドビジネスサテライト, プログラミング必修化, 日経スタイル, 前者, 協力者	プログラミング, ご紹介, 記事, プログラミング教育, 5 歳, 時代, 背景, 世代, 向上, 2020 年度, 人, 3 種類, 連日, 受講, ビジネス
A (4)	今日この頃, 教育現場, スクラッチ デイ, 教員側, 東京大学大学院情報学環, Show & Tell, Touch & Try, 義務化, 高校生, 日経 BP ブース, 課題, 橋元良明教授, 最大, ケース, Hour or Code Japan	ご紹介, プログラミング教育, 5 歳, 時代, 背景, 世代, 向上, 2020 年度, 人, 3 種類, 連日, 受講, ビジネス, NHK ニュース, 職種

B はプログラミング上級学習者向けの Facebook ページである。本稿では、A, B それぞれのページから投稿情報を抽出し、以下の 4 つの方法を用いて、ユーザ特性を算出した。

- (1) 単語  $i$  の  $TF-IDF$
- (2) (1) × 単語  $i$  が出現した投稿のいいね数
- (3) (2) × 単語  $i$  が出現した投稿のシェア数
- (4) (3) + 各コメントに対する (1) × 各コメントのいいね数

ここでは、それぞれのページにおいていいね数、シェア数を 0 から 1 に収まるように正規化した。上記のように、(1) ユーザが投稿した文章のみから算出した  $TF-IDF$ , (2) その  $TF-IDF$  に正規化したいいね数を重みとして積算した値, (3) さらに (2) に正規化したシェア数を重みとして積算した値とした。さらに、投稿した文章に対する他ユーザのコメントも考慮し、(4) (3) にその投稿にされた全コメントとそのコメントに対するいいね数を積算した値を加えた値より抽出したユーザ特性 (特徴語) を検証した。A, B から抽出された上下位 15 件の特徴語をそれぞれ表 1 および 2 に示す。特に、そのユーザ特性を表していると思われる語彙を太字で表に示している。

表 1 および 2 より、A と B ともに上位特徴語において、「東京大学大学院情報学環」や「jQuery UI」といった固有名詞が多く見られた。また、方法 (1) ~ (4) において、特徴語の順位が異なっていることがわかる。たとえば、A の「茂木健一郎」という語彙に注目してみると、方法 (1) では最大の重要値であるにも関わらず、方法 (2) ~ (4) よりいいね数、シェア数、

(注8) : <https://developers.facebook.com/>

表 2 B から抽出した特徴語上位 15 件と下位 15 件

ページ	特徴語 (上位) 興味喚起支援	(下位) 知識提供
B (1)	CSS3, EdTech JAPAN Pitch Festival vol.4, go to jappan, Higher or Lower, IE KMD Venture Day Tokyo, jQuery UI, Tech academy, u-note, パララックス, 工, good, SF JAPAN NIGHT セミファイナル進出チーム, 学習, Now we're hiring a great web designer, SF JapanNight, 学習	CODEPREP, プログラミング学習サービス CODEPREP, アプリ開発, ブック, 皆さん, プログラミング, 会員, 皆さま, 3 位, 6 チーム, 新しいブック, 公式 facebook ページ, 数時間, 3 点, 3 つ
B (2)	CSS3, jQuery UI, Thanks for Five Thousand Fans, 学習, u-note, Higher or Lower, フィードバック, SF JAPAN NIGHT 準決勝進出決定, インターン, we'll launch a radical web wervice which, Trello, Pyhonista, Now we're hiring a great web designer, SF Japan Night, This new service has already decided	CODEPREP, プログラミング学習サービス CODEPREP, アプリ開発, ブック, 皆さん, 00-3, 会員, 新しいブック, 42408, 公式 facebook ページ, 皆様, 2 冊, プログラミング, 公式 Twitter アカウント, 点
B (3)	まつもとゆきひろ, 学習, オブジェクト指向スクリプト言語, jQuery UI, サーバー再度言語解禁, タブ, SF JAPAN NIGHT 準決勝進出決定, Higher or Lower, 三段組レイアウト, 学習済, お問い合わせ, お声, 学習状況, Mats, CSS3	CODEPREP, プログラミング学習サービス CODEPREP, アプリ開発, ブック, 00-3, オンライン, 多く, 作り, swift, アプリ大会, ヤフー, 本日 2, 検討ちゆ, I want, コードブレップ
B (4)	ベリー, すー, まつもとゆきひろ, 学習, Koushou Kawasoe, オブジェクト指向スクリプト言語, jQuery UI, サーバサイド言語解禁, タブ, SF JAPAN NIGHT 準決勝進出決定, Higher or Lower, 三段組レイアウト, 学習済み, お問い合わせ, お声	CODEPREP, プログラミング学習サービス CODEPREP, アプリ開発, ブック, 00-3, オンライン, 多く, 作り, swift, アプリ大会, ヤフー, 本日 2, 検討ちゆ, I want, コードブレップ

他ユーザからのコメントを含めて重要度を計算した場合、上位 15 件にも入っていない。このことから、発信ユーザの情報のみから算出した上位特徴語と、他ユーザの発信情報も含めた上位特徴語では、抽出される語彙が異なることがわかった。本システムでは、重要度の高い語彙は、他ユーザとの会話における会話促進を目的とする興味喚起として使用する。そのため、語彙の中でも他ユーザが注目した語彙を用いたユーザ特性ベクトルを生成した方が受信ユーザだけでなく、発信ユーザにとっても有用な情報になると考えられる。以上より、表 1, 2 は、いいね数, シェア数, 他ユーザからのコメントを考慮した特徴語が興味喚起支援として最も有効な手法といえる。

A と B の下位特徴語において、方法 (1) ~ (4) を比較した場合、あまり大きく変動していないことがわかる。たとえば、B の「CODEPREP」という単語に注目すると、方法 (1) ~ (4) のすべての算出方法を用いても下位に属している。他にも B の「プログラミング学習サービス CODEPREP」にしても、すべての算出方法において下位に属しており、変化がないことがわかる。重要度の低い語彙に関しては、知識支援のためのリンクを生成する。このため、知識支援を必要とするような固有名詞が多くなく、逆に、「皆さま」や「会員」というような、広く一般的に使われる語彙が含まれているため、あまり有用とはいえない結果になった。今後、一般的に使われる単語を特徴語

から取り除く算出方法を加え、検証していく必要がある。

## 5. まとめ

本稿では、オンライン教育における会話促進として、受信側ユーザの知識支援および相手ユーザへ周知させたい興味喚起による知識提供支援となるリンク生成に関して述べた。また、Facebook の投稿情報を用いてユーザ特性の算出の有用性を比較した。上位特徴語の算出には、単語  $i$  の  $TF-IDF$  に単語  $i$  が出現した投稿のいいね数, シェア数, そしてコメントの語彙重要度平均を考慮した方法により、ユーザの発言以外の語彙をユーザ特性として新たに抽出でき、会話促進に有用な手法となり得るといえる。また、下位特徴語の算出には、方法 (1) ~ (4) を比較した結果、それぞれの方法での有用性があまり変わらないため、より多くのユーザを対象とすることで一般性の高い特徴語を検出し取り除くことや、上位特徴語の併用などの改善が必要である。

今後、発信ユーザ, 受信ユーザの双方にとって興味喚起となる語彙抽出の定性的評価を行う予定である。また、ユーザ特性を活かしたリンク先の選定の検証を行い、本システムの有用性を検証、さらにユーザ間の関係性を利用した関係マトリックスを構築し、ユーザ間の関係性を考慮したユーザ特性抽出方法を検討する。

## 謝 辞

本研究の一部は、JSPS 科研費 26280042 および 15K00162 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, 小野野弘, 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.