

階層化された論文情報とユーザ発話の埋め込みに基づく論文推薦システム

奥田由佳¹ 須藤 克仁^{1,2} 品川政太郎^{1,2} 中村 哲^{1,2}

¹ 奈良先端科学技術大学院大学

² 理化学研究所革新知能統合研究センター

{okuda.yuka.ou@, sudoh, sei.shinagawa, s-nakamura}@is.naist.jp

概要

初学者はユーザの検索意図が曖昧かつ専門知識を持たないため、既存の検索・推薦システムには適さない。そこで、階層化された論文情報とユーザ発話の2つをベクトル化することにより、効率的にユーザの検索意図を明確化し論文を推薦するシステムを提案する。将来的には対話システムの構築を目標とし、本研究はその土台となる。

1 はじめに

初学者を対象に論文を推薦する場合、広く流通している検索・推薦システムでの推薦は難しい。なぜなら、既存のシステムではユーザの検索履歴を必要とすることが多いが、初学者は検索履歴を持つことが困難だからである [1][2]。この問題へのアプローチとして、ユーザの情報を都度取得するインタラクティブなシステムを用いる手法 [3][4][5] がある。映画の推薦タスク [6][7] では、対話データで機械学習モデルを訓練し自然な会話の中でアイテムの推薦を行う手法 (Conversational Recommender System ; CRS) が提案されている。これらの研究は対話データを必要とする研究だが、論文の推薦においては公開済の対話データが存在しない。また、論文の推薦にはスキルが必要とされるため、対話データの構築もコストが高くこの手法は現状適用が難しい。

対話データを必要としない研究は、推薦アイテム同士の関係性に基づきインタラクティブに検索意図を絞り込む手法 [8][9] がある。しかし、この手法は、複数の興味を取得することを考慮していないため「興味の取りこぼし問題」が発生する恐れがある。これは、ユーザが検索意図が曖昧な状態で複数の分野に興味を持っていた場合、ユーザの興味をシステムが無視してしまうことを指す。例を図 1 に示

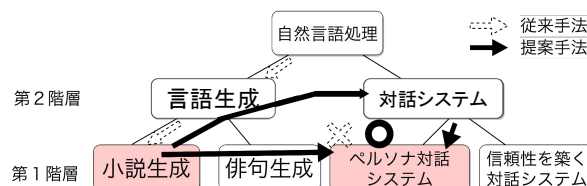


図1 提案手法を用いたノード間の移動

す。第1階層の各セルは論文に対応するノードを表し、第2階層の各セルは第1階層の論文に紐づく分野を表す。点線の矢印は、アイテム同士の関係性に基づきインタラクティブに検索意図を絞り込む従来の手法を示し、黒い矢印は提案手法を示す。例えば、ユーザの興味が図 1 中の赤色が付いた「小説生成」と「ペルソナ対話システム」だった場合、最初に「言語生成」を選択すると「小説生成」か「俳句生成」しか推薦できず「ペルソナ対話システム」の興味を無視してしまう。全ノードを確認することでこの問題は解決できるが、ユーザに多大な時間を使うことを強いる。そこで、アイテムの階層的知識をベクトル化し階層を跨いだノード間の移動を可能にすることで、効率的なノードの移動を可能にし興味の取りこぼしを起こしにくい手法を提案する。

2 関連研究

推薦システムの代表的な手法に協調フィルタリング [1][10] がある。しかし、協調フィルタリングはユーザの検索履歴を必要とし、検索履歴をもたない初学者には適さない。対して検索履歴を必要とせずユーザの情報をインタラクティブに都度取得する研究は、CRS [6][7]、Conversational Search [5]、Interactive Information retrieval (IIR) [3][4] などがある。

CRS に関する研究では、Chen らがアイテムの情報やアイテム同士の関係を考慮したシステムを提案している [6]。この研究では、アイテムの関係性と

対話の戦略を相互利用できるように教師あり学習によってモデルを構築する。しかし、階層的関係性や検索意図の明確化に着目したものではなく、これらに関する知見は明らかになっていない。また、学習に対話データを必要とし、対話データの取得が困難である論文を対象にした推薦システムには現状適応できない。

検索意図の明確化に関する研究には鈴木らの研究 [8] がある。この研究ではアブストラクトに含まれる単語の共起性を用いたクラスタリングと選択型の質問を用いて、各論文に対するユーザの興味度であるスコアを計算している。しかし、複数分野に跨る興味の取得を目的にしたモデルにはなっておらず、興味の取りこぼしを起す恐れがある。また、論文をユーザに複数提示しその選択結果を使用するが、初学者には論文は理解しにくく、選択の判断が難しいという懸念点がある。

3 提案手法

3.1 提案システムの概要

「興味の取りこぼし問題」を解決するための提案手法の仮説を述べる。提案手法では、構造的知識において階層を親から子へと下がっていくことを基本としながら、子から親ノードへの上方向への移動や違う根を持つノードへの移動など、あるノードから全ノードへの効率的な移動を可能にすることで「興味の取りこぼし問題」が解決できると仮説を立てた。例えば図 1 の場合、黒矢印のように、「小説生成」から「対話システム」や「ペルソナ対話システム」へ移動が可能ならば取りこぼしは解決できる。あるノードから全ノードへの移動を可能にするために、ユーザからのフィードバックと、階層的知識の各ノードの表現の仕方に工夫を加えた。ユーザのフィードバックについて、既存の研究では鈴木ら [8] のように提示した選択肢に対するユーザの選択結果を使用する研究が多い。対して、提案手法では、ユーザのフィードバックを発話の形で得ることで、より広いユーザの興味を取得可能にする。そして、ユーザ発話と階層的知識の各ノードをベクトル化することで、階層を跨いだノード間の移動を可能にし、興味の取りこぼしを起しにくくする。提案手法を用いた具体的な対話例を図 2 に示し、また具体的な手順を以下に記す。

(システム) 以下のキーワードから気になるものを選択してください
1. 小説生成 2. 俳句生成

(ユーザー) 1

(システム) 「小説生成」選んだ理由はなんですか？

(ユーザー) 魅力的なキャラクターのセリフを再現して小説生成がやりたいんだ！

(システム) あなたにオススメの論文はこれです。
・プロットと背景知識を用いた短編小説の自動生成
・模倣と置換に基づく超短編小説の自動生成
・人物ベースの Seq2Seq モデルを用いた対話システム Deep EVE における小説中の登場人物らしい応答文の生成

(システム) 以下のキーワードから気になるものを選択してください
1. 小説生成 2. ペルソナ対話システム

図 2 提案手法を用いた具体的な対話例

1. 前処理として、与えられた階層的知識の各ノードをベクトルで表現する。
2. ユーザに対して複数キーワードを提示し、ユーザの選択したキーワードを得る。
3. キーワードを選択した理由を問い、ユーザ発話取得する。このとき、取得した発話もベクトル化する。
4. ベクトル化したユーザ発話と、ベクトル化した各ノードで構成された階層的知識を照らし合せて、各ノードのスコアの計算を行う。
5. スコアの計算結果をもとに、ユーザに最適な論文を推薦する。
6. 手順 2 から 5 で 1 ターンとみなし、繰り返すことでユーザの意図を明確化する。

以降では、前処理の階層的知識の作成と提案手法の詳細を述べる。

3.2 前処理：階層的知識の作成

階層的知識の作成について説明する。まず、事前に図 3 のような階層的な知識が与えられるとする。第 2 階層のノードを図 3 の「言語生成」のような研究分野に対応させ、第 1 階層を「深層学習を用いた俳句の自動生成」のような論文のタイトルに対応させる。この事前に与えられた階層的知識の各ノードをそれぞれベクトル化する。各ノードのベクトル化には、ニューラルネットワークで単語のベクトルを学習する word2vec [11] を用いる。まず、第 1 階層の各ノードをベクトル化する手法を述べる。各論文のアブストラクトから、動詞・名詞・形容詞を特徴語として抽出し、学習済みの word2vec を用いて論文ごとに特徴語のベクトルの平均を取る。これを、第

1 階層のベクトルとする。第 2 階層のノードについては、対象のノードに紐づく第 1 階層のノードのベクトルに対する平均をとる。例えば、図 3 で表せられるように、第 2 階層の「言語生成」のノードのベクトルは、第 1 階層の「小説生成」と「俳句生成」のベクトルの平均になる。

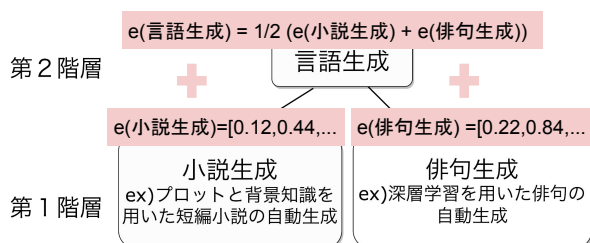


図 3 階層的知識のベクトル化

3.3 論文の推薦機構

どの論文を推薦すべきか判断する基準であるスコアの計算方法と、スコアを元にしたキーワード生成と、スコアの更新について述べる。

3.3.1 スコアの計算および更新

ユーザ発話とベクトル化した階層構造を用いてスコアの計算を行う。まず、ユーザ発話のベクトルを得る。ユーザ発話の埋め込みは、3.2 節に示したアブストラクトから論文の埋め込みを得る方法と同一の処理によって得る。得られた発話ベクトルと階層的知識の各ノードとの cosine をとり、その値をスコアとした。スコアを算出する式 1 に示す。

$$score_x = \frac{x \cdot y}{\|x\| \|y\|} \quad (1)$$

このとき、 $score_x$ はノード x のスコアを示し、 x は各ノードを、 y はユーザ発話をベクトル化したものを示す。最終的に、ノードごとに 1 つのスコアが割り当てられる。また、スコアの更新は各ノードで加算し、前ターンの情報を引き継ぐ。

3.3.2 キーワード生成

キーワード生成について説明する。3.3.1 節で算出したスコアを用いて、スコアが高い上位 3 つのノードを選択した後、各ノードに関連するキーワードをユーザに提示する。このとき、初学者が理解できるようなキーワードを提示するのが望ましいが、今回は階層を跨ぐノード間の移動が有益かどうかの確認が目的であるため主検討事項としなかった。

第 2 階層のノードが選択された場合はセッション名を、第 1 階層から選ばれた場合は論文のタイトルをキーワードとしてユーザに提示した。また、1 ターン目のキーワードは階層的知識の最上階層のノードに対応する。今回のモデルでは最上階層は第 3 階層であり、第 2 階層の研究分野をまとめたさらに抽象的な分野を表す。例えば図 1 ならば、1 ターン目のキーワードは「自然言語処理」のみとなりキーワードの選択が一意に決定し、システムの発話は「あなたが自然言語処理に興味がある理由を教えてください」になる。システムはキーワードの提示からユーザ発話の取得までを複数ターン繰り返すことで、曖昧な検索意図を明確化する。

4 実験

4.1 実験設定

提案システムの階層を跨いでノード移動を行う手法の有効性を検証するため、シミュレーション実験を行った。

データセット： ACL の 2020 年度の年次カンファレンスのプログラムである“Annual Meeting of the Association for Computational Linguistics (2020)(ACL2020)”を用いた。プログラムに記載されているセッション名を第 2 階層の各ノードに対応させ、論文のタイトルを第 1 階層の各ノードに対応させた。第 2 階層のノード数は 23、第 1 階層のノード数は 68 となった。また、スコアの計算に利用する word2vec では、事前学習済みモデルである GoogleNews-vectors-negative30¹⁾を利用した。

仮想ユーザと比較モデル： 実験のシナリオとして潜在的に 3 つの論文に興味を持つユーザ（仮想ユーザ）を事前に 3 人分設定した。このとき 3 つの論文はランダムに決定した。各仮想ユーザは初学者であると想定し専門的な用語を使用せず、各ユーザに設定された興味に従って著者が図 2 のようなユーザ発話を入れてシステムと 5 ターン対話した。

提案システムの階層を跨いでノード移動を行う手法の有効性を測るため、比較するシステム（ベースライン）として階層を跨ぐ機能を無効にしたシステムを用意した。ベースラインは図 2 の点線矢印のあるノードから属する子ノード

1) <https://code.google.com/archive/p/word2vec>

にしか移動できない挙動をする。スコアの計算は提案手法と同一であり、スコアの計算を行うノードの範囲はシステムが遷移可能なノードのみに対して行った。よって、次ノードの選択のみにスコアの計算を使用するため、スコアの更新は行なっていない。属する子ノードが無くなった地点で、一番初めの分岐点に戻り最初に選択したノードの次にスコアの高いノードを選択する。キーワード生成としては、属するノードが3個以下の場合はそのままの個数でユーザに提示し、3個を超える場合はランダムに3つ選択して提示した。

評価指標： 事前に仮想ユーザに設定した3つの論文に対し、システムがどのような順位づけをしたか把握するために評価指標を導入する。評価指標として、Average Precision (AP) や RR (Reciprocal Rank) も検討したが、今回は正解アイテムを推薦できたかの精度評価が目的ではないため、独自の以下評価スコアを導入した。

$$\text{評価スコア} = \frac{\text{論文の昇順順位}}{\text{論文の総数}} \quad (2)$$

評価スコアは、論文ごとに算出されたスコアに基づく順位を総数で割った数であり、対象とする論文が上位何%にあたるかを表す。スコアが低いほど、ユーザが他の論文に比べてその論文に興味を持っていることを表す。論文の総数については全ターンの中で遷移可能だった第1階層のノードの総数を指す。提案手法はすべてのノードに遷移可能なため、一貫して第1階層のノードの総数の68となる。論文の昇順順位もこの論文の総数の中で決定する。

4.2 実験結果・考察

評価指標による実験結果を表1に示す。表1中の論文1, 論文2, 論文3はランダムに設定した仮想ユーザが興味を持つ3つの論文を指し、横線が表示された箇所はシステムがスコアを取得できず、興味を取りこぼしを起こしていることを示す。

4.2.1 スコアの計算の妥当性

表1の結果から、提案手法の仮想ユーザAの論文1や論文3など設定した3つの論文に対してユーザの興味が強いつことを取得できている箇所が散見されることから、ユーザ発話によるスコアの計算の妥当性が示唆された。また、今回の実験では初学者を想

表1 評価指標による比較

仮想ユーザ	論文1	論文2	論文3
A (ベースライン)	0.0833	-	-
A (提案)	0.0882	0.985	0.0735
B (ベースライン)	0.438	-	-
B (提案)	0.529	0.868	0.176
C (ベースライン)	1.0	-	0.167
C (提案)	0.50	0.691	0.279

定しユーザ発話に専門的用語などを含めていないため、ユーザが専門知識を持たなくともシステムは有効なスコアの計算を行えることが示唆された。

4.2.2 取りこぼし問題が解決できたか

表1の結果から、前提としていた「取りこぼし問題」がベースラインでは発生し、提案システムでは発生せずスコアを取得できていることから提案システムの有効性が示唆された。また、ユーザAとユーザBの論文3に着目すると、ベースラインはユーザの興味を取得できていないが、提案システムではユーザの強い興味を取れている。このことから、提案システムではユーザの大きな興味の取りこぼしを防げることが示唆される。しかし、仮想ユーザに興味があると仮定した論文すべてにおいて提案システムが強い興味を把握できるとは限らないことがわかった。また、ユーザAの論文1やユーザCの論文3のスコアに着目すると、ベースラインの方がユーザの興味が強いつことを判断できていることがわかる。このことから、ベースラインの方が興味の取りこぼしは起こしやすいつが、確実に一つの分野に関する興味が把握できると考えられる。以上の考察から、複数の興味をより効率的に取得するため、階層を跨ぐ挙動に加えノード移動の仕方により工夫が必要であると考えられる。

5 おわりに

本研究では、「興味の取りこぼし問題」に対して、階層的知識をベクトル化し階層を跨いだノード間の移動を可能にすることで解決する手法を提案し、その有用性を検討した。結果として、期待した通りに提案手法は曖昧な検索意図を明確化しつつ取りこぼし問題に対して有効であることが示唆された。今後の展開としては、本研究で得られた知見をもとに論文推薦の対話システムの構築を行い、最終的には主観評価などを行う予定である。

参考文献

- [1] Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breiter. Research-paper recommender systems: a literature survey. **International Journal on Digital Libraries**, Vol. 17, No. 4, pp. 305–338, 2016.
- [2] Jyotirmoy Gope and Sanjay Kumar Jain. A survey on solving cold start problem in recommender systems. In **2017 International Conference on Computing, Communication and Automation**, 2017.
- [3] Hung-Yi Lee, Pei-Hung Chung, Yen-Chen Wu, Tzu-Hsiang Lin, and Tsung-Hsien Wen. Interactive spoken content retrieval by deep reinforcement learning. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, Vol. 26, No. 12, pp. 2447–2459, 2018.
- [4] David Robins. Interactive information retrieval: Context and basic notions. **Informing Science The International Journal of an Emerging Transdiscipline**, Vol. 3, No. 2, Jan 2000.
- [5] Corby Rosset, Chenyan Xiong, Xia Song, Daniel Campos, Nick Craswell, Saurabh Tiwary, and Paul Bennett. Leading conversational search by suggesting useful questions. In **The Web Conference 2020**, 2020.
- [6] Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. Towards knowledge-based recommender dialog system. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing**, 2019.
- [7] Raymond Li, Samira Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. Towards deep conversational recommendations. In **Proceedings of the 32nd International Conference on Neural Information Processing Systems**, 2018.
- [8] 鈴木亮詞, 工藤聖広, 辻野友孝, 清水堅, 白松俊, 大園忠親, 新谷虎松. 論文リポジトリに基づく研究支援のための対話的ユーザモデル構築手法の提案. 全国大会講演論文集, 第 72 回, 人工知能と認知科学, pp. 741–742, Mar 2010.
- [9] Yi-Cheng Pan, Hung-Yi Lee, and Lin-Shan Lee. Interactive spoken document retrieval with suggested key terms ranked by a markov decision process. **IEEE Transactions on Audio, Speech, and Language Processing**, Vol. 20, No. 2, pp. 632–645, Feb 2012.
- [10] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In **Proceedings of the 26th International Conference on World Wide Web**, 2017.
- [11] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, **1st International Conference on Learning Representations, Workshop Track Proceedings**, 2013.

付録

A 仮想ユーザに事前に設定した3つの論文

4節で行ったシミュレーション実験の、仮想ユーザに事前にランダムに設定した3つの論文を表Aに示す。

表A 仮想ユーザが興味を持っていると仮定した3つの論文

仮想ユーザ	興味のある論文のタイトル
A	1.You Don't Have Time to Read This: An Exploration of Document Reading Time Prediction 2.Multiscale Collaborative Deep Models for Neural Machine Translation 3.Predicting the Topical Stance and Political Leaning of Media using Tweets
B	1.SyntaxGym: An Online Platform for Targeted Evaluation of Language Models 2.Text Classification with Negative Supervision 3.Revisiting the Context Window for Cross-lingual Word Embeddings
C	1.Semantic Scaffolds for Pseudocode-to-Code Generation 2.You Impress Me: Dialogue Generation via Mutual Persona Perception 3.Reverse Engineering Configurations of Neural Text Generation Models,

B 対話履歴の例

4節で行なったシミュレーション実験の対話履歴を例として図Bに示す。pinakesがシステムを指し、userがユーザを示す。例として取り上げたのは4節の仮想ユーザAに対する提案手法での対話例である。

(🤖 pinakes:) Please tell me why you are interested in research on natural language processing.

(user) Because I have an interest to relationship between language and emotion.

(🤖 pinakes:) Please select the keyword you are interested.

1: Media Bias the Social Sciences and {NLP}: Automating Frame Analyses to Identify Bias by Word Choice and Labeling,

2: Discourse and Pragmatics,

3: What determines the order of adjectives in {E}nglish? Comparing efficiency-based theories using dependency treebanks

(user) 1

(🤖 pinakes:) Please tell me the reason you chose.

(user) I'm curious about how people's everyday writing changes with or without emotional bias.

(🤖 pinakes:) Please select the keyword you are interested.

1: Media Bias the Social Sciences and {NLP}: Automating Frame Analyses to Identify Bias by Word Choice and Labeling,

2: What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context,

3: Cognitive Modeling and Psycholinguistics

(user) 2

(🤖 pinakes:) Please tell me the reason you chose

(user) I am interested in sentiment analysis contained in SNS. Also, I would like to analyze emotions not only from SNS but also from texts of books such as literature.

(🤖 pinakes:) Please select the keyword you are interested in from the bottom

1: Media Bias the Social Sciences and {NLP}: Automating Frame Analyses to Identify Bias by Word Choice and Labeling

2: What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context

3: Cognitive Modeling and Psycholinguistics

(user) 3

(🤖 pinakes:) Please tell me the reason you chose

(user) By the way, I'm also interested in psychology.

(🤖 pinakes:) Please select the keyword you are interested in from the bottom

1: What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context,

2: Media Bias the Social Sciences and {NLP}: Automating Frame Analyses to Identify Bias by Word Choice and Labeling,

3: Cognitive Modeling and Psycholinguistics

(user) 3

(🤖 pinakes:) Please tell me the reason you chose

(user) Well, more specifically, I'm interested in emotions, psychology and language. I especially like books, and I'm interested in analyzing emotions from the text of books and how they affect the human mind.

図B 仮想ユーザAの提案手法での対話例