

学生向け論文推薦システムにおける 説明インタフェースの評価

久徳 泰知^{1,a)} 西岡 千文^{2,b)} 緒方 広明^{3,c)}

概要：

本稿は、大学生・大学院生を対象とした論文推薦システムのインタフェースの比較評価を実施する。推薦システムは、講義資料閲覧システム上で、表示されている教材に関連する論文を推薦する。説明インタフェースは、「なぜそのアイテムが推薦されているか」という情報をユーザへ提示することから、ユーザの意思決定プロセスを支援し、推薦システムへの信頼性を向上させる効果があることが知られている。本稿では、論文推薦システムにおける説明インタフェースを、効果・透明性・満足度から評価する。具体的には、No explanation, Keywords, Tag cloud (mono-color), Tag cloud (color) の4つの説明インタフェースを比較する。経済学あるいは情報学の教材を利用して、30名の学生を対象に被験者実験を実施した。結果、各説明インタフェース間の有意差は観察されなかった。被験者の属性に着目すると、修士課程の被験者の満足度について、Tag cloud (mono-color) が最も高かった。推薦されている論文と教材との関連性を色で表現する Tag cloud (color) よりも、関連性を明示しない Tag cloud (mono-color) の評価が高かった理由として、彼らは教材との関連性よりも彼らの関心に基づいて推薦アイテムの適合性を判断していることが挙げられる。対して、学士課程の被験者の満足度については、No explanation と Tag cloud (color) が最も高かった。説明インタフェースに表示されるキーワードを理解できない場合は No explanation の満足度が高く、理解できる場合は Tag cloud (color) が高い。このことから、学士課程の被験者は推薦アイテムの適合性を判断する際に教材との関連性に注目しており、説明インタフェースが関連性の理解を支援しているといえる。

Experimental Comparison of Explanation Interfaces on Research Paper Recommender System for University Students

1. はじめに

京都大学において実施された調査 [13] では、学生は研究に関心をもっていることが明らかとなった。しかし、多くの学部生向けの講義では基礎的知識の習得に焦点を当てており、論文に反映されている最先端の学問・研究を目にする機会が少ない。このようなことから、先行研究 [8] では、学生向けの論文推薦システムが開発された。論文推薦システムは、講義資料閲覧システム BookRoll [10] に組み込まれている。BookRoll は、講義中ならびに予習時・復習時に

スライドなどの教材を表示する Web アプリケーションである。

論文推薦システムは、学生が BookRoll で閲覧している教材に関連する論文を推薦する。学生は推薦された論文によって、「学習している内容が最先端の学問とどのように繋がっているか」を知ることができるため、学習活動が刺激されることが期待される。しかし、先行研究 [8] では、推薦された論文に気が付いた学生は約半数に留まり、そのうち約半数の学生のみが推薦された論文にアクセスした。気が付いたがアクセスしなかった学生は、推薦された論文と教材との関連性を見つけることが難しかったということを知り、アクセスしなかった理由として挙げた。

本稿では、講義資料閲覧システム上の論文推薦システムに対する学生のエンゲージメントを高めることを目的として、論文推薦システムの説明インタフェースの比較評価を

¹ 京都大学大学院 情報学研究所

² 京都大学 附属図書館

³ 京都大学 学術情報メディアセンター

a) kyutoku.yasutomo.86c@st.kyoto-u.ac.jp

b) nishioka.chifumi.2c@kyoto-u.ac.jp

c) ogata.hiroaki.3e@kyoto-u.ac.jp

実施する。推薦システムにおける説明インタフェースは、アイテムが推薦されている理由をユーザに明示する。説明インタフェースは、ユーザの意思決定過程を支援し、推薦システムに対する信頼性を向上させ、ユーザのエンゲージメントを向上させる [11]。本稿では、Gedikli らが実施した映画の推薦システムにおける説明インタフェースの比較評価 [3] で評価が高かった説明インタフェースを採用して、論文推薦システムの説明インタフェースの比較評価を実施する。特に、(1) No explanation, (2) Keywords, (3) Tag cloud (mono-color), (4) Tag cloud (color) の4つの説明インタフェースを比較評価する。本稿では、30名の学生を対象に、経済学あるいは情報学の教材を利用して、被験者実験を実施した。結果、各説明インタフェース間の有意差は観察されなかった。しかし、透明性においては Keywords が、満足度においては Tag cloud (mono-color) が他の説明インタフェースよりもわずかに高かった。特に、修士課程の被験者が Tag cloud (mono-color) に高い満足度を与えたことが観察された。推薦されている論文と教材との関連性を色で表現する Tag cloud (color) よりも、関連性を明示しない Tag cloud (mono-color) の評価が高かった理由として、彼らは教材との関連性よりも彼らの関心に基づいて推薦アイテムの適合性を判断しているということが挙げられる。対して、学士課程の被験者の満足度については、No explanation と Tag cloud (color) が最も高かった。説明インタフェースに表示されるコンセプトを理解できない場合は No explanation の満足度が高く、理解できる場合は Tag cloud (color) が高くなる。このことから、学士課程の被験者は、推薦された論文の適合性を判断する際に教材との関連性に注目しており、Tag cloud (color) が関連性の理解を支援しているといえる。インタビューにおいても、学士課程の被験者からは、「Tag cloud (color) によって、教材と推薦論文の関連性を理解することができた。」といった肯定的な意見を得られた。

本稿の構成を以下に示す。次章では、論文推薦システムと説明インタフェースに関する関連研究を挙げる。3章では、論文推薦システムにおける推薦の計算方法と比較対象となる4つの説明インタフェースについて述べる。4章では、実験方法について述べる。5章では、実験結果と考察を報告する。6章を本稿のむすびとする。

2. 関連研究

過去20年以上にわたり、数多くの論文推薦システムが開発されてきた [2]。これらは学術情報の発見支援を目的として研究者を対象とすることに対して、本稿は学生を対象とする。論文を推薦することで、学生の学術的関心を深め、学習活動を活発化することが目的である。学生向け論文推薦システムを開発した先行研究 [8] では、約半数の学生のみが推薦されている論文に気が付き、そのうち約半数

(全体では約27%)の学生のみが推薦された論文にアクセスした。気が付いたにも関わらずアクセスしなかった学生は、推薦された論文と教材との関連性を理解することが出来なかったことを、アクセスしなかった理由として挙げた。

先行研究 [8] を含めて論文推薦システムのインタフェースの多くは、推薦されている論文の著者やタイトルといった基本的な書誌情報をリストとして表示する。しかし、書誌情報のみの表示では、推薦された論文が実際に関心を惹きつけるものであるか判断することは、難しい。特に、論文の閲読経験が浅い学生にとっては困難であるといえる。

映画といったその他の分野の推薦システムでは、推薦システムにおけるユーザの意思決定を支援するため、説明インタフェースによってアイテムが推薦された理由を明示することの効果について調査されてきた [4], [7]。説明インタフェースは、推薦システムに対するユーザのエンゲージメントを向上させる方法として効果的であると報告されている [11]。Gedikli ら [3] は、105名の被験者を対象に、映画の推薦システムにおいて10つの説明インタフェースの比較評価を実施した。結果として、Tag cloud がユーザの満足度が最も高いことが報告された。

本稿は、講義資料閲覧システム上における論文推薦システムの説明インタフェースの比較評価を実施する。説明インタフェースを導入することで、論文推薦システムに対する学生のエンゲージメントの向上が期待される。実験では、(1) No explanation, (2) Keywords, (3) Tag cloud (mono-color), (4) Tag cloud (color) といった4つの説明インタフェースを比較評価する。比較対象の説明インタフェースについては、3.2にて詳述する。

3. 学生向け論文推薦システムと説明インタフェース

先行研究 [8] 同様、講義資料閲覧システム BookRoll 上に論文推薦システムを開発する。図1が示すように、スクリーン左部に教材が表示されており、右部に推薦される論文が表示される。先行研究と異なり、論文推薦システムは、説明インタフェースを利用して推薦される論文を表示する。

3.1節では本稿における推薦の計算方法について述べ、3.2節では比較対象となる4つの説明インタフェースについて述べる。

3.1 推薦の計算

本稿では、教材の各ページで推薦される論文を、コンテンツベースフィルタリングによって計算する。コンテンツベースフィルタリングで利用するテキストマイニング手法は、HCF-IDF [9] である。HCF-IDF は、テキストから階層構造の知識ベースに保存されているコンセプト(見出し語、キーワード等)を最初に抽出する。その後、伝搬関数 [5]

クラス分類器の発展・流行の歴史 7

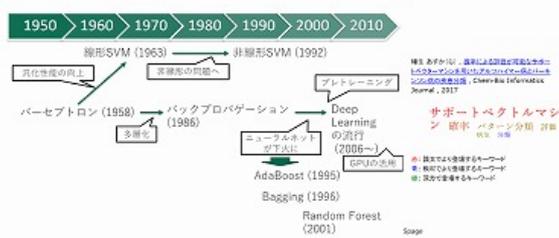


図 1: 講義資料閲覧システムと論文推薦システム

を使用し、知識ベースにおいて抽出されたコンセプトの上位に位置するコンセプトに対しても重み付けを行う。よって、HCF-IDF ではテキストに直接含まれていないが、関連度が高いコンセプトを特徴量として保有できる。また、知識ベースは一つのコンセプトに対し複数のラベルを保持するので、同義語も一つの特徴量として扱うことが可能である。本稿で利用する知識ベースについては、4.2 節で詳細に述べる。なお、本稿では著作権等の懸念により、論文の本文テキストを使用することができないため、タイトルのみを使用して推薦を計算する。タイトルのみを利用した論文推薦システムにおいて HCF-IDF が機能することが報告されている [9] ことから、本稿では HCF-IDF を利用する。

3.2 説明インタフェース

比較評価の対象となる 4 つの説明インタフェースを以下に挙げる。

- (1) **No explanation** 本インタフェースでは、論文の著者、タイトル、学会誌あるいは会議名、出版年のみを表示する。本インタフェースは、論文推薦システムにおいて広く採用されている [1]。よって、本インタフェースはベースラインとなる。
- (2) **Keywords** Keywords は、図 2a で例示されるように、推薦される論文において最も重みが高い 3 個のコンセプトのラベルをリストとして表示する。コンセプトの重みは HCF-IDF により計算される。最も簡素であると考えられるキーワード形式の説明インタフェースは、Gedikli らの実験 [3] では対象となっていなかった。そのため、本稿で評価する。
- (3) **Tag cloud (mono-color)** Tag cloud (mono-color) は、推薦される論文において最も重みが高いコンセプトのラベルを最大 10 個表示する。ラベルのフォントサイズは、コンセプトの重みに応じて決定される。図 2b に Tag cloud (mono-color) の例を示す。Gedikli らの実験 [3] では、Tag cloud が最もユーザの満足度が高いという結果が報告されたので、採用する。
- (4) **Tag cloud (color)** Tag cloud (color) では、Tag cloud (mono-color) と異なり、ラベルのフォントに

色が与えられる。色は推薦される論文と教材の関連性を表現する。教材での重みが比較的高いコンセプトのラベルは青色、論文での重みが比較的高いコンセプトのラベルを赤色、双方で同程度の重みをもつコンセプトのラベルは緑色で表示される。色によって、教材と推薦される論文の関連性を理解することを支援する。Tag cloud (color) の例を図 2c に示す。

4. 実験

学生向け論文システムにおいて、4 つの説明インタフェースの比較評価を行うため、被験者実験を実施する。4.1 節では実験手順を示す。その後、データセット、被験者、評価指標についてそれぞれ述べる。

4.1 手順

実験手順を、実験手順 1 に示す。被験者は実験手順 1 で示されている手順が実装されたウェブアプリケーションを操作し、講義資料閲覧システムにおける論文推薦システムの 4 つの説明インタフェースを評価する。実験を観察し、技術的な質問に回答するため、実験企画者が被験者の側につく。はじめに、実験企画者は被験者に推薦システムの目的を簡単に説明し、映画や EC サイトでの推薦システムの簡単な例を紹介する。また、データの収集と利用について合意を求める。その後、被験者はウェブアプリケーションを操作し、教材として用意された 20 枚のスライドの各ページにて、説明インタフェースへ評価を与えることが求められる。なお、被験者の専攻分野に応じて、経済学あるいは情報学のスライドが利用される。スライドの各ページで推薦される論文は 1 件であり、1 件の論文に対して 1 つの説明インタフェースが表示される。よって、被験者は説明インタフェースを合計 20 回評価する。スライドの各ページで適用される説明インタフェースは、各被験者において無作為に決定される (4 行目)。

被験者は、推薦アイテムについて、適合性 (論文にアクセスしたいと考えるか)、透明性 (論文が推薦されている理由が明らか、論文の内容を推測できるか) を表示されている書誌情報と説明インタフェースのみで評価する (5・6 行目)。その後、被験者は推薦アイテムの内容 (論文の概要・本文) を確認し、内容を考慮して推薦アイテムの適合性を再度評価する (7・8 行目)。適合性・透明性については、4.4 節で詳述する。

ウェブアプリケーション操作による説明インタフェースの評価終了後に、実験企画者は被験者に対してインタビューを実施する。インタビューでは、説明インタフェースに対する意見を収集する。例えば、「特定のアイテムが推薦された理由を理解することは重要であるか」や「説明インタフェースに求める機能・コンセプトは何か」といった質問を尋ねる。

幡生 あすか(ら), [確率による評価が可能なサポートベクターマシンを用いたアルツハイマー病とパーキンソン病の疾患分類](#), Chem-Bio Informatics Journal, 2017

幡生 あすか(ら), [確率による評価が可能なサポートベクターマシンを用いたアルツハイマー病とパーキンソン病の疾患分類](#), Chem-Bio Informatics Journal, 2017

- サポートベクトルマシン
- 分類
- 確率

幡生 あすか(ら), [確率による評価が可能なサポートベクターマシンを用いたアルツハイマー病とパーキンソン病の疾患分類](#), Chem-Bio Informatics Journal, 2017

サポートベクトルマ
シン 分類 確率 パター
ン分類 評価 病気

サポートベクトルマ
シン 確率 パターン分類 評価
病気 分類

赤: 論文でより登場するキーワード
青: 教材でより登場するキーワード
緑: 双方で登場するキーワード

(a) Keywords

(b) Tag cloud (mono-color)

(c) Tag cloud (color)

図 2: 比較対象となる説明インタフェース

実験手順 1

- 1: \mathbf{R} = 推薦の集合
- 2: \mathbf{E} = 説明インタフェースの集合
- 3: for each $r \in \mathbf{R}$ do
- 4: r に対して無作為に e を選択する
- 5: 被験者に対して推薦 r を説明インタフェース $e \in \mathbf{E}$ を使用して提示する
- 6: 被験者に推薦 r を評価させる
- 7: 被験者に対して推薦 r の本文を表示する
- 8: 被験者に推薦 r に対して再度評価させる
- 9: end for

4.2 データセット

本実験で使用する教材, 推薦論文, 知識ベースを以下に示す.

教材 各被験者は自らの専攻に応じて, 経済学あるいは情報学の 20 枚のスライドを教材として使用する. 教材は日本語で記されている. 経済学のスライドタイトルは, 「人工市場を用いた金融市場の制度・規制の議論」^{*1}であり, 情報学のスライドタイトルは「Python による機械学習～SVM から DeepLearning まで～」^{*2}である.

推薦論文 推薦候補となる論文として, 日本語で記述された 707,637 件のオープンアクセスの論文のデータセットを使用する. JaLC メタデータのデータセット^{*3}から, Unpaywall のデータセット^{*4}でオープンアクセスであると判断されている論文のリストを抽出した. そのうち, 日本語の論文タイトルをもつ 707,637 件の論文を利用した. 論文タイトルの言語判定には langdetect^{*5}を利用した.

*1 [http://www.mizutakanobu.com/20180405.pdf](http://www.mizutatakanobu.com/20180405.pdf), 最終アクセス日: 2019 年 2 月 21 日

*2 <https://www.slideshare.net/yasutomo57jp/python-svmdeep-learning>, 最終アクセス日: 2019 年 2 月 21 日

*3 https://japanlinkcenter.org/top/material/material_metadata.html, 最終アクセス日: 2019 年 2 月 21 日

*4 <https://unpaywall.org/products/snapshot>, 最終アクセス日: 2019 年 2 月 21 日

*5 <https://pypi.org/project/langdetect/>, 最終アクセス日: 2019 年 2 月 21 日

知識ベース HCF-IDF に使用する階層構造をもつ知識ベースとして, JST 科学技術用語シソーラス^{*6}を使用する. JST 科学技術用語シソーラスは, 科学技術に関する 37,410 個のコンセプト (見出し語, キーワード等) と, それらコンセプトのラベルが 155,863 個含まれている. 教材のページと論文タイトルからのコンセプトの抽出は, 形態素解析器 MeCab^{*7}に, JST 科学技術用語シソーラスに基づく MeCab 用専門用語辞書^{*8} [14] を利用することで, 実施した.

4.3 被験者

本実験では, 経済学, 情報学, またはそれらに関連する分野を専攻する学生 30 名を被験者として採用した. 彼らは, 大学 2 年生から大学院修士課程 1 年生までのいずれかに属し, 日本語で開講されている講義を受講している. 経済学の教材を利用した被験者は 15 名, 情報学の教材を利用した被験者は 15 名である. また, 20 名が学士課程に所属し, 10 名が修士課程に所属する.

4.4 評価指標

本節は実験で利用する 3 つの評価指標について詳述する. Gedikli ら [3] は, 説明インタフェースの評価において, 効果 (Effectiveness)・効率性 (Efficiency)・説得力 (Persuasiveness)・透明性 (Transparency)・満足度 (Satisfaction) といった評価指標を採用した. 効果・透明性が説明インタフェースの満足度に大きな影響を与えることから, 我々はこの 5 つの評価指標のうち, 効果・透明性・満足度の計 3 つの評価指標を利用する.

効果 (Effectiveness) 効果的な説明インタフェースは, ユーザが推薦アイテムの性質と適合性を正しく判断す

*6 <http://thesaurus-map.jst.go.jp/jisho/fullIF/index.html>, 最終アクセス日: 2019 年 2 月 21 日

*7 <http://taku910.github.io/mecab/>, 最終アクセス日: 2019 年 2 月 22 日

*8 <http://doi.org/10.18908/1sdba.nbd0c2358-001>, 最終アクセス日: 2019 年 2 月 22 日

ることを支援する。このような説明インタフェースは、ユーザの関心に合致しない推薦アイテムを除外し、推薦システムの有用性と利便性を著しく向上させる。説明インタフェースの効果を測定するために、最初に被験者は論文の書誌情報と説明インタフェースのみを考慮して、推薦アイテムの適合性を5段階リッカート尺度（1：最低～5：最高）で評価する。その後、被験者は推薦アイテムの内容（論文の概要・本文）を確認し、それらを考慮して再度推薦アイテムの適合性を5段階リッカート尺度（1：最低～5：最高）で評価する。効果は、書誌情報と説明インタフェースのみを考慮した場合と、推薦アイテムの内容を考慮した場合の2つの適合性の評価値の差によって測定される。

透明性 (Transparency) 特定のアイテムが推薦されている理由をユーザへ明示して推薦システムの透明性を高めることは、システムの信頼性の向上及びユーザの意思決定に貢献する。このことから、推薦システムにおいて、透明性は重要な要因として位置付けられる [12]。被験者は、論文が推薦されている理由の透明性を、5段階リッカート尺度（1：最低～5：最高）で評価する。

満足度 (Satisfaction) 推薦システムに対するユーザの全体的な満足度は、推薦アイテムとそれらの説明インタフェースに強く影響される [12]。実験では、説明インタフェースに対する被験者の満足度を解析することに焦点を当てる。被験者は説明インタフェースの満足度を5段階リッカート尺度（1：最低～5：最高）で評価する。さらに、他の評価指標（効果、透明性）が満足度へ与える影響を分析する。

5. 結果と考察

本章では、被験者実験の結果と考察について述べる。5.1節では、4つの説明インタフェースの各評価指標の結果について報告する。5.2節と5.3節では、被験者属性が評価に与える影響と評価指標間の相関について、それぞれ述べる。5.4節では、被験者へのインタビュー結果を報告する。なお、本稿では有意水準として $\alpha = .05$ を利用する。

5.1 説明インタフェースの比較

表1に、4つの説明インタフェースの効果・透明性・満足度における平均と標準偏差を示す。透明性と満足度は5段階リッカート尺度によって評価されているため、最小評価値は1で最高評価値は5である。対して、効果は5段階リッカート尺度（1：最低～5：最高）で評価された適合性の差であり、差が小さいほど効果的な説明インタフェースであると認識されるため、最小評価値は4で最高評価値は0である。

各評価指標において、4つの説明インタフェースの間に有

意差があるか分析するため、統計的検定を実施する。最初に、Mendozaの多標本球面性検定 [6] により、群間の共分散行列が等質であるという球形仮説が成立するか検定する。球形仮説が棄却された場合 ($p < .05$)、Greenhouse-Geisserの ϵ によって自由度を調整し、一元配置反復測定分散分析 (One-way repeated measures ANOVA) を実施する。

効果については、Mendozaの多標本球面性検定が有意であったため ($p < .05$)、Greenhouse-Geisserによる自由度調整 ($\epsilon = 0.80$) を行い、分散分析を実行した結果、 $F(3, 87) = 0.65$ ($p = .58$) であり、有意差は観察されなかった。透明性については、Mendozaの多標本球面性検定が有意でなかったため ($p = .74$)、自由度調整を行わず分散分析を実行した結果、 $F(3, 87) = 1.01$ ($p = .39$) であり、有意差は観察されなかった。満足度についても、Mendozaの多標本球面性検定が有意でなかったため ($p = .96$)、自由度調整を行わず分散分析を実行した結果、 $F(3, 87) = 0.09$ ($p = .97$) であり、有意差は観察されなかった。よって、いずれの評価指標においても、4つの説明インタフェースの間で、有意差は観察されなかった。これは、被験者が推薦されている論文の適合性や透明性を判断する際に、説明インタフェースよりも論文の書誌情報に注目することが理由として挙げられる。

有意差は観察されなかったものの、効果ではNo explanationとTag cloud (color)の平均がその他2つの説明インタフェースよりも高く、透明性ではKeywordsが、満足度ではTag cloud (mono-color)の平均が高かった。Tag cloud (mono-color)が満足度において他のインタフェースよりも高かった理由として、Tag cloud (mono-color)では表示されるコンセプト（見出し語、キーワード等）の数が多いたことが理由として挙げられる。また、Tag cloud (mono-color)は、推薦されている論文と教材との関連性を色で表現するTag cloud (color)よりも、満足度と透明性において評価が高かった。理由として、各コンセプトに色を与えたことによってユーザの認知努力を増加させて判断をしづらくさせてしまったことが挙げられる。特に、修士課程の被験者が学士課程の被験者と比較すると、Tag cloud (mono-color)に高い満足度を与えることが観察された。この点については、5.2節で詳述する。

5.2 被験者属性が評価に与える影響

被験者の属性が異なると、説明インタフェースに与える評価に差異が発生する可能性があると考えられる。被験者の在籍課程と実験で利用した教材の分野が、説明インタフェースの評価に与える影響を明らかにするため、統計的検定を実施した。5.1節同様、Mendozaの多標本球面性検定 [6] により、群間の共分散行列が等質であるという球形仮説が成立するか検定する。球形仮説が棄却された場合 ($p < .05$)、Greenhouse-Geisserの ϵ によって自由度を調整

表 1: 各説明インタフェースの効果・透明性・満足度の平均 (Mean) と標準偏差 (SD)

	効果		透明性		満足度	
	Mean	SD	Mean	SD	Mean	SD
No explanation	0.91	0.55	3.12	0.89	3.06	0.90
Keywords	1.01	0.66	3.29	0.75	3.07	0.94
Tag cloud (mono-color)	1.02	0.70	3.12	0.81	3.12	0.97
Tag cloud (color)	0.88	0.49	3.04	0.66	3.07	0.78

し、二要因混合計画の分散分析（一被験者内要因，一被験者間要因）（Mixed ANOVA, one within-subjects factor, one between-subjects factor）を実施する。この場合、被験者内要因は、在籍課程（学士課程または修士課程）あるいは利用教材の分野（経済学あるいは情報学）であり、被験者内要因は説明インタフェースである。

在籍課程による影響。 被験者の在籍課程別における各評価指標の平均と分散を表 2 に示す。表 2 から、修士課程の被験者は学士課程の被験者よりも、いずれの評価指標においても高い評価を与えていることがわかる。以下、各評価指標において、二要因混合計画の分散分析を実施し、被験者の在籍課程が説明インタフェースの評価に与える影響を調査する。

効果について、Mendoza の多標本球面性検定が有意であった ($p < .05$)。よって、Greenhouse-Geisser による自由度調整 ($\epsilon = 0.79$) を適用して分散分析を実行した結果、在籍課程 ($F(1, 28) = 0.15$ ($p = .70$)), 説明インタフェース ($F(3, 84) = 0.63$ ($p = .59$)), 在籍課程 \times 説明インタフェース ($F(3, 84) = 0.77$ ($p = .51$)) いずれの要因でも、有意差は観察されなかった。透明性について、Mendoza の多標本球面性検定が有意でなかったため ($p = .17$)、自由度調整を適用せず分散分析を実行した。結果、在籍課程 ($F(1, 28) = 5.21$ ($p < .05$)) で有意差が観察された。具体的には、表 2 に示されているとおり、修士課程の被験者は学士課程の被験者と比較すると、説明インタフェースの透明性を高く評価する。修士課程の被験者が専門分野の知識を比較的多く備えていることが理由として挙げられる。学士課程の被験者と比較すると、彼らにとって推薦される論文の書誌情報や説明インタフェースに表示されるコンセプトは理解しやすいことから、透明性が高くなったと考えられる。説明インタフェース ($F(3, 84) = 0.70$ ($p = .52$)), 在籍課程 \times 説明インタフェース ($F(3, 84) = 1.75$ ($p = .16$)) では、有意差は観察されなかった。満足度について、Mendoza の多標本球面性検定が有意でなかったため ($p = .06$)、自由度調整を行わず分散分析を実行した。在籍課程 ($F(1, 28) = 2.16$ ($p = .15$)), 説明インタフェース ($F(3, 84) = 0.73$ ($p = .54$)) では有意差は観察されなかった。しかし、在籍課程 \times 説明インタフェース ($F(3, 84) = 3.01$ ($p < .05$)) では有意差が観察された。そのため多重比較を実施した結果、修士課程の被験者は、Tag cloud (mono-color) の満足度を、学士課程の被験

者よりも高く評価することが認められた ($F(1, 28) = 6.45$ ($p < .05$)). Tag cloud (mono-color) の満足度を、学士課程の被験者は平均 2.83 (分散: 0.90) の評価を与えているのに対して、修士課程の被験者は平均 3.70 (分散 0.88) の評価を与えている。修士課程の被験者が Tag cloud (mono-color) を最も高く評価した理由として、Keywords と比較すると論文を表現するコンセプトが多く含まれていることが挙げられる。また、その分野の基礎的知識を習得済みである彼らにとって、それらのコンセプトを理解することが可能であったことも理由として挙げられる。修士課程の被験者にとって、Tag cloud (mono-color) は Tag cloud (color) よりも満足度が高かった。理由として、修士課程の被験者は、教材の内容との関連性には気を配らず彼らの関心に基づいて推薦された論文の適合性を判断するため、教材との関連性を表現するコンセプトの色が意味をなさなかったことが挙げられる。対して、学士課程の被験者の満足度については、No explanation と Tag cloud (color) が最も高かった。説明インタフェースに表示されるコンセプトを理解できない場合は No explanation の満足度が高く、理解できる場合は Tag cloud (color) が高くなると考えられる。このことから、学士課程の被験者は、推薦された論文の適合性を判断する際に教材との関連性に注目しており、Tag cloud (color) が関連性の理解を支援しているといえる。

利用教材の分野による影響。 被験者の教材の分野別における各評価指標の平均と分散を表 3 に示す。表 3 から、情報学の教材を利用した被験者は経済学の教材を利用した被験者よりも、いずれの評価指標においても高い評価を与えていることがわかる。以下、各評価指標において、二要因混合計画の分散分析を実施し、利用教材の分野が説明インタフェースの評価に与える影響を調査する。

効果について、Mendoza の多標本球面性検定が有意であった ($p < 0.01$)。よって、Greenhouse-Geisser による自由度調整 ($\epsilon = 0.79$) を適用して分散分析を実行した結果、教材の分野 ($F(1, 28) = 0.00$ ($p = .97$)), 説明インタフェース ($F(3, 84) = 0.64$ ($p = .59$)), 在籍課程 \times 教材の分野 ($F(3, 84) = 0.67$ ($p = .58$)) いずれの要因でも、有意差は観察されなかった。透明性について、Mendoza の多標本球面性検定が有意でなかったため ($p = .77$)、自由度調整を適用せず分散分析を実行した。結果、教材の分野 ($F(1, 28) = 5.65$ ($p < .05$)) で有意差が観察された。具体

表 2: 被験者の課程別における各評価指標の平均 (Mean) と標準偏差 (SD)

	学士課程 (n = 20)		修士課程 (n = 10)	
	Mean	SD	Mean	SD
効果	0.98	0.61	0.91	0.60
透明性	2.98	0.72	3.48	0.78
満足度	2.93	0.85	3.37	0.91

表 3: 被験者の利用教材の分野別における各評価指標の平均 (Mean) と標準偏差 (SD)

	経済学 (n = 15)		情報学 (n = 15)	
	Mean	SD	Mean	SD
効果	0.96	0.60	0.95	0.61
透明性	2.90	0.72	3.39	0.77
満足度	2.87	0.92	3.29	0.80

的には、表 3 に示されているとおり、情報学の教材を利用した被験者は経済学の教材を利用した被験者よりも、説明インタフェースの透明性を高く評価する。説明インタフェース ($F(3, 84) = 1.01$ ($p = .39$)), 在籍課程 × 教材の分野 ($F(3, 84) = 1.05$ ($p = .38$)) では、有意差は観察されなかった。満足度について、Mendoza の多標本球面性検定が有意でなかったため ($p = .31$), 自由度調整を行わず分散分析を実行した。教材の分野 ($F(1, 28) = 2.26$ ($p = .14$)), 説明インタフェース ($F(3, 84) = 0.09$ ($p = .97$)), 教材の分野 × 説明インタフェース ($F(3, 84) = 1.97$ ($p = .13$)) いずれの要因でも、有意差は観察されなかった。

5.3 評価指標間の相関

効果・透明性・満足度、それぞれの評価指標間にある相関を調査するため、ピアソンの積率相関係数 (Pearson's product-moment correlation coefficient) とケンドールの順位相関係数 (Kendall's rank correlation coefficient) を算出した。結果を表 4 に示す。

効果/満足度では、わずかな相関が観察された。これらの p 値は、それぞれ $p = .09$, $p = .08$ であり、有意に近い。相関係数は負であるものの、効果の値は小さいほどよいため、説明インタフェースの効果が高ければ満足度も高くなる傾向があるといえる。効果/透明性では有意な相関は観察されなかった。透明性/満足度ではケンドールの順位相関係数でもピアソンの相関係数でも有意な正の相関が観察された。透明性は、効果と比較すると、満足度の評価へより大きな影響を与えるといえる。

5.4 インタビュー結果

インタビューで被験者より得られた説明インタフェースに対する肯定的な意見として、「Tag cloud (color) によって、教材と推薦論文の関連性を理解することができた。」

表 4: 各評価指標のピアソン (Pearson) の積率相関係数とケンドール (Kendall) の順位相関係数

	Pearson	Kendall
効果/満足度	-0.15 (0.09)	-0.11 (0.08)
効果/透明性	-0.01 (0.96)	0.01 (0.89)
透明性/満足度	0.47 (0.00)	0.36 (0.00)

() 内は p 値

「論文を読むときに、キーワードを参考にすることで教材との関連に注目しながら読むことができる。」といったことが挙げられた。一方、否定的な意見として、「説明インタフェースに表示されるキーワード (コンセプト) が少ない。」、「推薦された論文は学士課程の学生にとっては理解できない。」といったことが挙げられた。修士課程の被験者の推薦システムならびに説明インタフェースに対する反応は、学士課程の被験者と比較するとより好意的であった。推薦システムと説明インタフェースにおける透明性と満足度を向上させるためには、論文の難易度などを考慮する必要があると考えられる。

6. おわりに

本稿は、講義資料閲覧システムに設置された学生向け論文推薦システムの説明インタフェースの比較評価を実施した。説明インタフェースは、特定のアイテムが推薦されている理由をユーザに明示することで、ユーザの推薦システムに対する信頼性を高める。このことから、説明インタフェースを利用することで、学生の論文推薦システムに対するエンゲージメントを向上させることができると考える。被験者実験では、(1) No explanation, (2) Keywords, (3) Tag cloud (mono-color), (4) Tag cloud (color) といった 4 つの説明インタフェースを比較した。各説明インタフェース間の有意差は観察されなかったが、透明性においては Keywords が、満足度においては Tag cloud (mono-color) が他の説明インタフェースよりもわずかに高かった。特に、修士課程の被験者が Tag cloud (mono-color) に高い満足度を与えたことが観察された。推薦されている論文と教材との関連性を色で表現する Tag cloud (color) よりも、関連性を明示しない Tag cloud (mono-color) の評価が高かった理由として、彼らは教材との関連性よりも彼らの関心に基づいて推薦アイテムの適合性を判断しているということが挙げられる。対して、学士課程の被験者にとっては、Tag cloud (color) は Tag cloud (mono-color) よりも評価が高く、インタビューにおいても「Tag cloud (color) によって、教材と推薦論文の関連性を理解することができた。」といった肯定的な意見があった。一方、「学士課程の学生にとって推薦された論文を理解することは困難である。」といった意見もあった。

今後は、学生の在籍課程や学習分野の差異を考慮した上

で、論文の難易度といった他の要因を考慮した推薦システムならびに説明インタフェースを検討していく。

謝辞

本研究の一部は、JSPS 科研費若手研究 18K13235, JSPS 科研費基盤研究 (S)16H06304, 及び NEDO SIP 18102059-0 の助成を受けたものです。

参考文献

- [1] J. Beel, A. Aizawa, C. Breitingner, and B. Gipp. Mr. DLib: recommendations-as-a-service (RaaS) for academia. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 1–2. IEEE, 2017.
- [2] J. Beel, B. Gipp, S. Langer, and C. Breitingner. paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4):305–338, 2016.
- [3] F. Gedikli, D. Jannach, and M. Ge. How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382, 2014.
- [4] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 241–250. ACM, 2000.
- [5] P. Kapanipathi, P. Jain, C. Venkataramani, and A. Sheth. User interests identification on twitter using a hierarchical knowledge base. In *European Semantic Web Conference (ESWC)*, pages 99–113. Springer, 2014.
- [6] J. L. Mendoza. A significance test for multisample sphericity. *Psychometrika*, 45(4):495–498, 1980.
- [7] T. Nava and M. Judith. Effective explanations of recommendations: user-centered design, 2007.
- [8] C. Nishioka and H. Ogata. Research Paper Recommender System for University Students on the E-Book System. In *ACM/IEEE on Joint Conference on Digital Libraries (JCDL)*, pages 369–370. ACM, 2018.
- [9] C. Nishioka and A. Scherp. Profiling vs. time vs. content: What does matter for top-k publication recommendation based on Twitter profiles? In *ACM/IEEE on Joint Conference on Digital Libraries (JCDL)*, pages 171–180. IEEE, 2016.
- [10] H. Ogata, C. Yin, M. Oi, F. Okubo, A. Shimada, K. Kojima, and M. Yamada. E-book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE)*, pages 401–406, 2015.
- [11] P. Pu and L. Chen. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6):542–556, 2007.
- [12] K. Swearingen and R. Sinha. Interaction design for recommender systems. In *Designing Interactive Systems*, volume 6, pages 312–334, 2002.
- [13] 京都大学. 平成 21 年度京都大学学生生活実態調査報告書, 2009. <http://www.kyoto-u.ac.jp/static/ja/education/campus/report/documents/h21/i.pdf>.
- [14] 建石 由佳, 信定 知江, and 高木 利久. JST 科学技術用語シソーラスに基づく MeCab 用専門用語辞書. In *言語処理学会 第 23 回年次大会 発表論文集*, pages 485–488. 言語処理学会, 2017.