

インターネット学習における学習支援ガイド生成に向けた理解支援関係の自動抽出

高橋健太郎[†] 伊藤 寛祥^{††} 松原 正樹^{††} 森嶋 厚行^{††}

[†] 筑波大学情報学群メディア創成学類 〒305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学図書館情報メディア系 〒305-8550 茨城県つくば市春日 1-2

E-mail: [†]takahashi.kentaro@mlab.info, ^{††}{ito,masaki,mori}@slis.tsukuba.ac.jp

あらまし 教科書は学習において、学習の順番や事項の重要性を示す重要なガイドの働きをするため、初学者の学習には重要である。しかし、現在発展途上国などにおいてインターネットへのアクセスはあるものの必要な教科書が買えない学生が多数存在するという状況が報告されている。このような状況に置かれた学習者によりよい学習支援を行うことは重要な社会問題である。また、次々と現れる新たなテーマについては、インターネットに関連情報はたくさんあっても、教科書が存在しない分野が多数存在する。このような背景のもと、本論文では、インターネット上にある様々な教材コンテンツを手掛かりに、初学者のために学習項目に関連付けた理解支援関係グラフを自動構築する問題に取り組む。具体的には、クラウドソーシングを用いて収集された Web コンテンツから知識間の理解支援関係を抽出する手法を複数構築し、専門家が作成した学習ロードマップが存在する領域について、それぞれ理解支援関係を抽出した結果を示す。

キーワード 学習支援, クラウドソーシング, Knowledge Graph, 理解支援関係の抽出

1 序 論

発展途上国などでは十分な学習コンテンツに恵まれず、インターネットを用いて学習をしなければならないケースが多数ある。Zell [1] によると、アフリカの大学において、特に輸入された教科書が非常に高価であることや、国内での高等教育用教科書の生産が少ないことから、多くの学生は必要な教科書を買えない状況がある。例えばナイジェリアのある大学で 300 人を対象にしたインターネット利用に関する研究 [2] では、毎日インターネットを使う人が 76%、利用目的の 75.6% が学習目的と答えた。アフリカには 2020 年において大学生が約 1 億 6 千万人¹ おり、その大部分が同様の状況にあることが予想される。教科書が入手困難で、インターネットへのアクセスがある人たちにによりよい学習支援を行うことは、アフリカをはじめとする困難な学習環境に置かれた学習者により良い学習機会を提供できるという観点から、非常に重要な問題である。

教科書が提供する重要な機能は、どのトピックがどのトピックよりも先に勉強すればよいかを示すことである。この情報が与えられれば、先に述べた問題に貢献することができると考える。“Composition”, “Generation”, “Curriculum”, “Roadmap”, “Educational”, “Knowledge Graph”, “Learning Material”, “from web resources”などの単語を組み合わせた検索キーワードにより 200 件以上の既存研究を調査したが、任意の領域について学ぶための Web 学習教材を収集するという研究 [3] [4] は存在するものの、「どのトピックをどのトピックより

も先に学べば良いか」を自動抽出する研究は我々の知る限り存在しない。

本論文では、学習するトピックの集合とこれに関連する Web リソース (文書集合) が与えられたときに、学習という視点からの理解支援関係を求める問題に取り組む。理解支援関係とは、あるトピックの知識 v_1, v_2 を学習する際に、これらの間に「 v_1 を理解するためには v_2 を理解しておいた方がよい」ということが成立するときの関係である。例えば、トピック「線形代数学」に関する知識「固有値」と「行列の掛け算」について、「固有値よりも行列の掛け算を先に学習したほうがよい」という関係が成り立つ。

また、「りんご」と「果物」という知識について、「りんご」のことを知っていれば「果物」についても理解しやすく、同時に「果物」のことを知っていれば「りんご」についても理解しやすいことが予想される。理解支援関係はこのようなケースにも対応する関係である。したがって理解支援関係は順序ではないことに注意したい。

このような理解支援関係を得ることによって、すでに理解している知識をもとに、次に学習すべき知識を推薦することや、理解が十分ではない知識をもとに、先に学習すべき知識を推薦することなどが可能になる。もし順序関係によって知識間の関係が定義されるのであれば、「りんご」と「果物」の間の関係は、二項関係 \prec を用いて「りんご」 \prec 「果物」または「果物」 \prec 「りんご」に限定される。ここで、 $v_1 \prec v_2$ は「 v_2 を理解するためには v_1 を先に理解しなければならない」という関係とする。前者の場合、「りんご」を学習した後に「果物」を学習するように条件づけられるため、「果物」を先に学習すれば「りんご」を理

1: <http://data.uis.unesco.org/>

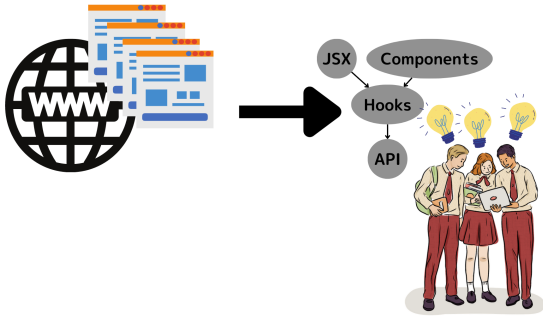


図 1 研究の概要 本研究では、web コンテンツ中に存在する知識間の理解支援関係を、知識に関連する文書間の関係性から自動的に抽出する手法を提案する。

解できる可能性があるのにもかかわらず、「りんご」を知らないユーザーに「果物」を推薦することができない。後者の場合も同様である。しかしながら、理解支援関係が順序関係ではないことによって、「りんご」を知らないユーザーに「果物」を推薦することも、「果物」を知らないユーザーに「りんご」を推薦することも可能になる。したがって、理解支援関係は、学習者に適切な学習事項を推薦できる可能性を持つ関係である。

Web リソースから学習支援コンテンツを作ることで、そのコンテンツをインターネット上で自由にアクセスできるようになるが、それ自体は新しいものではない。しかし、特定のテーマやトピックについて、まだ初学者向けのコンテンツが存在しない場合は、そのような学習支援コンテンツを作成することで、そのテーマやトピックについて学習するための新しいリソースが提供されることになるため、特に、前述した困難な学習環境に置かれた学習者により良い学習機会を提供することが可能である。また、Web リソースを用いることにより、ほとんどのテーマやトピックについて学習リソースを提供することができる。例えば、学習者が特定の学科や分野について学習することを目的としている場合は、その学科や分野に関するテキスト、ビデオ、オンラインのインタラクティブな学習アクティビティ、試験や問題を含む学習リソースを作成することが技術的に可能である。また、特定のスキルやタスクを習得することを目的としている場合は、そのスキルやタスクを学習するための段階的なガイドやチュートリアル、実践的な練習問題を含む学習リソースを作成することが可能である。さらに、文化や歴史、自然科学や人文科学などのテーマについても、同様の学習リソースを作成することが可能である。

文書集合を入力とした場合の理解支援関係の自動抽出については様々なアプローチが考えられるが、第一歩として本論文では Term Subsumption [5] に基づく関係抽出が、理解支援関係についてどの程度有効であるかを調査する。Term Subsumption [5] とは、2つの単語 w_1, w_2 が与えられたとき、それぞれを含む文書集合 $D(w_1)$ と $D(w_2)$ が $D(w_1) \subset D(w_2)$ の時、 w_2 をより一般的な単語と見なす考え方である。本論文では、これが理解支援関係の抽出にも有効であるとの仮説をたて、これに基づく理解支援関係抽出手法を2つ検討し、行った実験の結果を示す。

実験では、専門家が作成した「React.js²」に関する学習ロードマップを正解理解支援関係とし、与えられたトピック集合に対して直接 Term Subsumption 抽出（単純 TS 法）をした場合と、一時的にカテゴリを表すノードを導入し、エッジを中間ノードによって抽出した場合（NMF 中間ノード法）の比較をおこなった。その結果、あるトピックについては、後者が理解支援関係に関する Precision の減少を抑えつつ、Recall を大きく増加させることがわかった。

本論文の貢献は次の通りである。

- (1) Web に存在する解説文書集合からの理解支援関係関係を抽出する問題について、Term Subsumption [5] に基づく手法の比較を行ったこと。
- (2) カテゴリ中間ノードを導入した方法が、単純 TS 法と比較して、約 30% の Precision で約 80% の Recall を示した。これは本手法で提案した関係をクラウドソーシングなどを用いてフィルタリングするということが現実的なアプローチとして利用できることを示している。

2 関連研究

本節では、学習コンテンツの自動作成に関する研究を紹介する。

Novak らによると、Knowledge Graph などの可視化ツールは知識の記憶や整理、上位思考スキルの発達を促進すると考えられている [6] ため、クラウドソーシングを用いて収集された Web コンテンツから知識間の理解支援関係を抽出する手法を開発することができれば、本問題に貢献できると考える。

授業計画の作成に関する研究 [3] では、全順序として定義される学習順序を出力するが、各々の知識が複雑な依存関係を持つような領域では全順序が適切とは言えない。そのためより柔軟に知識間の関係を表現できる Knowledge Graph が適切である。また、学習コンテンツと学習順序の作成に関する研究 [4] は、エキスパートによって学習順序が作成されるが、複雑な構造を表現する場合のことを考えると、自動で抽出できる方が好ましい。なお、自然言語で記述されたテキストから Ontology (=Entity とそれらの関係) を抽出する Ontology Learning は本論文と類似した試みであるものの、本論文は Entity とそれらの関係を、理解支援関係で表現することに相違点がある。

2.1 授業計画の作成

CollectiveTeach [3] は、生徒のニーズに合わせた高品質でインタラクティブなLESSンプランを作成するための簡単で効果的なツールを教師に提供するために、特定のトピックに関するコンテンツのリストを自動的に生成し、それらのリソースをアノテーション付けすることで、従来であれば時間と労力を要する手作業での授業計画の作成を自動で行うことを目的としている。具体的には、特定のトピックに有用である可能性が高いリ

2: <https://reactjs.org/>

ソースを推薦する検索エンジンと、検索されたりソースを一貫性を保ちつつ論理的な順序で並べて授業計画のアウトラインを生成するシーケンスフレームワークによって実現する。システムは次の3つの観点から評価された。

- (1) 情報検索システムのパフォーマンス
- (2) 作成された授業計画の適切さ、品質、カバレッジ
- (3) 検索された文書のシーケンス化の妥当性

Ranawat らは、これら3つの観点において、システムのパフォーマンスは高いと結論付けた。しかしながらこの手法によって作成される学習順序は連続した1次元の順序であるため、各々の知識が複雑に依存しているような領域には適さない。そのため本論文では、より柔軟に知識間の関係を記述できる Knowledge Graph を用いて理解支援関係を表現する。また、CollectiveTeach [3] のターゲット利用者が教師であるのに対し、本論文で提案する手法によって作成する学習支援コンテンツのターゲット利用者は学習者本人である。

学習順序の決定 Ranawat らは、一つのコンセプトについて簡潔に述べられている単一の Web ページ内の教材、PDF ファイル、YouTube などのマルチメディアコンテンツ、または CollectiveTeach システム内のユーザー投稿コンテンツは ALU (Atomic Learning Unit) と定義した。そして、Agrawal ら [7] によって導入された "comprehension burden" を最小化するような ALU の順序 (授業計画) $L_s = (A_1, A_2, \dots, A_n)$ を学習順序とした。

2.2 学習コンテンツと学習順序の作成

Li ら [4] は、学生個人の学習ニーズや目標に合わせたカリキュラムをアルゴリズムを使用することによって動的に作成する手法について議論した。著者らは、学生の個々の学習ニーズや目標に対応し、彼らの特定のニーズに合わせたカリキュラムを作成することができるシステムを提案した。具体的には、Learning Object [8] (LO とする) を用いてパーソナライズされたカリキュラムを作成する。学習者は学習したい事柄に関するキーワードを検索クエリとしてシステムに入力すると、システムがユーザーのプロフィールの情報 (学習状況) などを用いて検索クエリにアノテーションを行う。このクエリによって膨大な LO の中から学習者の学習ニーズにマッチするものを選び、教育理論に基づいて決定されたそれぞれの LO の順序を学習順序としてユーザーに提供する。これによりユーザーの興味・関心および学習状況を考慮した個別最適なカリキュラムが提供される。著者らはまた、このようなシステムを実装する上での課題も議論しており、学生の能力を正確に評価する必要性や、アルゴリズムを訓練するための大量のデータセットの必要性などが挙げられている。

この手法では、本論文と同様に Knowledge Graph によって学習コンテンツのカリキュラムを構成するが、学習順序をシステムのユーザーが作成するため、やはり知識間に複雑な依存関係を持つ領域では限界があると考えられる。そのため本論文で

は Term Subsumption [5] などのアルゴリズムによって理解支援関係を抽出する。

学習順序の決定 学習するトピックがノードとして、それらがエッジ (e.g. is-a 関係, part-of 関係, Sequential, ...etc) によって意味論的に相互に関係しているグラフ (シラバスとする) のそれぞれのトピックについて一つ以上の LO を紐づける。この際に LO を選定及び順序付けるために次の教育学的規則が用いられる。

- (1) 概要説明のための LO は演習のための LO よりも先行する。
- (2) 同じトピックに関して、初学者には単純な導入と例を提供する一方で、学習が進んだ者にはより詳細で深い情報を提供する。

これに加えて、LO のメタデータに記載されている自身の品質の説明もより品質の高い LO の選定に用いられる。

例を挙げると、ある初学者が「線形代数学」というクエリを検索システムに入力したとすると、「線形代数学」にマッチするトピックノードを起点のノードとして選択する。学習者はこのトピックについてほとんど知識がない (システムに登録された個人情報や背景知識、興味をもとに推論) ため、生成されるカリキュラムには基本的な内容と必要な背景知識が含まれている必要がある。すなわち、シラバスにおいて「線形代数学」ノードに対して、親的関係のエッジ (e.g. Subtypeof, Sequential) で結びついている別のトピックノードを深さ N (事前に定義) に到達するまで選択していく。選択されたトピックノードに関連する LO を親的トピックから子トピック順に並べつつ、同じトピックの LO は導入的内容から演習の内容になるよう並べることでカリキュラムの作成とする。

2.3 Ontology Learning

Ontology Learning は、自然言語で記述された文書集合から、あるドメイン内の概念の集合とそれらの概念間の関係を形式的に表現することである。Ontology は知識を構造化された形式で表現することができ、機械がより効率的に情報を理解し、推論し、処理することを可能にする。Ontology Learning は、自然言語処理、情報検索、Semantic Web の開発などの様々なアプリケーションで使用される。具体的な Ontology Learning のタスクを次に記述する。

- (1) Ontology に表現する必要がある概念と関係を特定する。
- (2) 必要な情報を抽出するために、関連するデータを収集して処理する。
- (3) 概念と関係を、階層やグラフなどの形式的な構造で表現する。
- (4) Ontology を検証し、ドメインを正確に表現するように洗練する。

本論文では、同領域の Term Subsumption [5] と呼ばれる手法から本論文の着想を得ているものの、本論文で導くノード間の

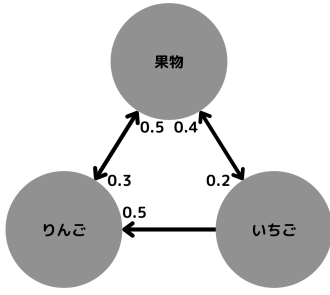


図 2 理解支援関係グラフの例

関係は理解支援関係であり，これは Ontology Learning で導くノード間の関係との相違点である．

3 問題定義

理解支援関係．本論文では，あるトピックに関する知識をどの知識と合わせて学べば良いかを発見する問題を扱う．ここで，トピックに関する知識とは，そのトピックを理解するための学習事項である．例えば「対角行列」は「線形代数学」というトピックに関する知識の例である．

あるトピックの知識 v_1, v_2 が与えられたとする．もし v_2 を理解するためには v_1 を理解しておいた方がよいとき， $v_1 \prec v_2$ と書く．例えば，トピック「線形代数学」に関する知識「固有値」と「行列の掛け算」について，「固有値よりも行列の掛け算を先に学習したほうが良い」という関係が成り立つ．

本論文で扱う問題．本論文では，入力として (1) トピック T に関する解説の文書集合，(2) T に関連する知識の集合をとり，理解支援関係関係を表すエッジ集合 \mathcal{E} を推定することで，そのトピックの知識を学習する順番を意味する重み付き理解支援関係グラフを生成する問題を扱う．図 2 に理解支援関係グラフの例を示す．

より具体的には，各文書 $d_i \subseteq \mathcal{V}$ は知識の集合 $\mathcal{V} = \{n_1, n_2, \dots, n_{d_i}\}$ の部分集合で構成されるとし，知識間の有向エッジ $e_{i,j} = (n_i, n_j)$ の集合 $\mathcal{E} = \{e_{i,j}\}$ を推定することを目的とする．この推定は重み関数 $W : \mathcal{E} \rightarrow [0, 1]$ として表現される．

本論文では，知識の集合 \mathcal{V} は事前に与えられるものとする．ここで，知識集合 \mathcal{V} と知識間の有向エッジ集合 \mathcal{E} から構成されるグラフ $G = (\mathcal{V}, \mathcal{E})$ を理解支援関係グラフと定義し， \mathcal{E} の推定により学習支援ガイドの生成とする．

4 関係導出手法

本論文では，「より一般的な知識をより先に学習する」ような関係が理解支援関係として適切であるという考えのもと，Term Subsumption [5] を用いた理解支援関係抽出手法を 2 つ検討する．

単純 TS 法 知識の集合 \mathcal{V} の任意の 2 つの知識と少なくともどちらか 1 つに一致する文字列を含む文書集合に対して

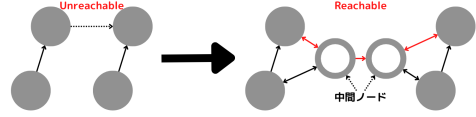


図 3 NMF 中間ノード TS 法

Term Subsumption [5] を用いてエッジを張る．

NMF 中間ノード TS 法 本手法は，入力として与えられた文書集合と知識の集合から，NMF（非負値行列因子分解）を用いて求められたクラスターを各知識に紐づく中間ノード（次に記述するカテゴリを表す）として生成し，エッジの集約を行う．

以下の項で各要素の詳細を記述する．

4.1 カテゴリ

各知識が所属する分類のことである．例えば，「因数分解」や「積分」といった知識は「数学」というカテゴリに所属する．

4.2 単純 TS 法

2 つの知識に関する文書集合があった時，他方を内包している集合の方が一般的な知識であるという考え方に基いて知識間の階層的関係を決定する手法．本論文では，一般的な集合を学習した後他方の集合に移動するという考え方で Knowledge Graph を作成するほか，内包している度合いを示す値 δ (次に記述) を導入する．

ここで，ノード対に対する重み $W : \mathcal{E} \rightarrow [0..1]$ を以下の様に定義する．

$$W(n_i, n_j) = \frac{|\{d_k \in \mathcal{D} \mid n_i \in d_k \wedge n_j \in d_k\}|}{|\{d_k \in \mathcal{D} \mid n_j \in d_k\}|} \quad (1)$$

この重みを利用し，以下のようにエッジ集合を生成する．

$$\mathcal{E} = \{e_{i,j} \mid W(n_i, n_j) \geq \delta\} \quad (2)$$

ここで δ はパラメータである．

4.3 NMF 中間ノード TS 法

単純 TS 手法では，実際には関係のある知識であるのに，それぞれに関する文書において，もう片方の知識を示す単語が出現しない場合に対応できなかった．そこで，各知識が所属するカテゴリを表すノード（中間ノード）を NMF によって生成し，因子行列を用いてエッジを中間ノードに集約し，中間ノード間のエッジを Term Subsumption [5] に基づいて生成することで，単純 TS 手法では到達不可能であったノードに到達できるようになる可能性がある（図 3）．このような考えのもと，NMF を用いて中間ノードを生成する．次に詳細を記述する．

中間ノードの集合を $\mathcal{V}_{nmf-mid} = \{n_1^*, \dots, n_N^*\}$ とし、ノード集合を $\mathcal{V} \cup \mathcal{V}_{nmf-mid}$ とする理解支援関係グラフを新たに生成する。

はじめに 4.2 項によって生成した理解支援関係グラフのノード集合 \mathcal{V} を、NMF によって得られた因子行列を用いて k 個のクラスタに分割する。ここで、各中間ノード n_k^* は各クラスタのインデックス k に 1 対 1 対応しているとする。また、4.2 項によって得られた理解支援関係グラフの隣接行列 M は、NMF によって得られた非負値の $m \times k$ 行列 W と非負値の $k \times m$ 行列 H^T を用いて以下に近似できる。

$$M \approx WH^T \quad (3)$$

ここで、因子行列はそれぞれ $W = [w_{ij}]$, $H = [h_{ij}]$ とする。ここで、 W , H^T は一意に決まらないことに注意する。解を一意にするために、Wei ら [9] の手法に基づき行列 W の各列のユークリッド距離を 1 にする。以下によって W の正規化を行う。

$$w_{ij} \leftarrow \frac{w_{ij}}{\sqrt{\sum_i w_{ij}^2}} \quad (4)$$

$$h_{ij} \leftarrow h_{ij} \sqrt{\sum_i w_{ij}^2} \quad (5)$$

次に各クラスタに属するすべてのノードに対して、中間ノードとの重み 1 のエッジを生成する。その集合を

$$\begin{aligned} \mathcal{E}_k^{(nmf-in)} = & \{(n_i, n_k^*) \mid n_i \in \mathcal{C}_k \wedge k = \underset{j}{\operatorname{argmax}} h_{ij}\} \\ & \cup \{(n_k^*, n_i) \mid n_i \in \mathcal{C}_k \wedge k = \underset{j}{\operatorname{argmax}} h_{ij}\} \end{aligned} \quad (6)$$

とする。ここで、 \mathcal{C}_k は、以下のように定義される。

$$\mathcal{C}_k = \{n_i \mid k = \underset{j}{\operatorname{argmax}} h_{ij}\} \quad (7)$$

続いて、中間ノード間のエッジを知識間の共起関係に基づいて生成する。より具体的には以下のようにエッジの重みを生成する。

$$W(n_i^*, n_j^*) = \frac{|\{d_k \in \mathcal{D} \mid n_l \in \mathcal{C}_i \wedge n_m \in \mathcal{C}_j \wedge n_l, n_m \in d_k\}|}{|\{d_k \in \mathcal{D} \mid n_l \in \mathcal{C}_i \wedge n_l \in d_k\}|} \quad (8)$$

この重みを利用し、以下の様に中間ノード間のエッジ集合 $\mathcal{E}^* = \{e_{i,j}^*\}$ を生成する。

$$\mathcal{E}^* = \{e_{i,j}^* \mid W(n_i^*, n_j^*) \geq \delta\} \quad (9)$$

ここで δ はパラメータである。

5 実験

本節では、提案手法の有用性を検証する。より具体的には以下の観点について検証を行う。

5.1 実験設定

本論文では、「提案手法によって作成したある領域に関する理解支援関係グラフが、専門家が作成した同領域に関する理解支援関係に類似していれば正しい理解支援関係を抽出できている」という考え方にに基づき、roadmap.sh³に掲載されている学習ロードマップを正解の理解支援関係グラフとし、このグラフにおける任意の 2 つの知識 (= ノード) 間の到達可能性と、生成した理解支援関係グラフの対応する知識間の到達可能性を比較することで評価を行う。

トピック. 最も利用者の多いプログラミング言語”JavaScript⁴”において、16,085 人を対象に行われた調査で約 84% が利用している⁵ライブラリである”React.js⁶”をトピックとして、同領域に関する理解支援関係グラフを作成する。

文書集合. クラウドソーシングで web リンクを収集し、収集したリンクから対象の文書及びサイト全体の文書を合計で 6279 件収集した。

知識の集合. 正解の理解支援関係グラフに現れる学習事項の集合を知識の集合とした。

生成理解支援関係グラフ 提案手法を用いて生成した理解支援関係グラフを生成理解支援関係グラフとする。

正解理解支援関係グラフ 2022 年 10 月 31 日時点で roadmap.sh³ にて公開されている”React.js⁶”に関する学習ロードマップ (図 4) を用いて正解理解支援関係グラフを構築した。

まず、図 4 を説明する。この図には 2 種類の色 (黄色とベージュ) のノードがあるのが見える。黄色のノード (”Fundamental Topics”, ”Advanced Topics”, そして”Ecosystem”) はカテゴリノードであり、具体的に学習する知識を表しているわけではない。他のベージュ色で記されているノードが、具体的に学習する知識ノードに対応する。また、上から下に向かって学習するという流れになっている。ただし、知識ノードの中でより細かく分割されているノード (例えば、components に対する functional components と class components や、表現は違うが、Common Hooks に対する useCallback 等) は、その全体像 (components や Common hooks) を学習してからそれぞれのサブ知識を学ぶ必要がある。

したがって、本実験では、次のように正解の理解支援関係を定義した。

$v_1 \prec v_2$ iff v_1 と v_2 がベージュノード (知識) $\wedge v_1$ がサブ知識を持たない知識 $\wedge v_2$ がサブ知識ではない知識。

評価 正解理解支援関係グラフにおけるすべてのノード集合 (= 知識の集合) を $\mathcal{V}_{ans} = \{n_i^{ans}\}$, 正解理解支援関係グラフにおけるすべてのエッジ集合を $\mathcal{E}_{ans} = \{e_{i,j}^{ans}\}$ としたとき、 $n_i = n_i^{ans}$, $n_j = n_j^{ans}$ なる生成理解支援関係グラフの任意のノード対 n_i, n_j について、それらの到達可能性と n_i^{ans}, n_j^{ans} の到達可能性を比較することで真偽を評価する。次に詳細を記述

3 : <https://roadmap.sh>

4 : <https://insights.stackoverflow.com/survey/2021#most-popular-technologies-language>

5 : <https://2021.stateofjs.com/en-US/libraries/front-end-frameworks/>

6 : <https://reactjs.org/>

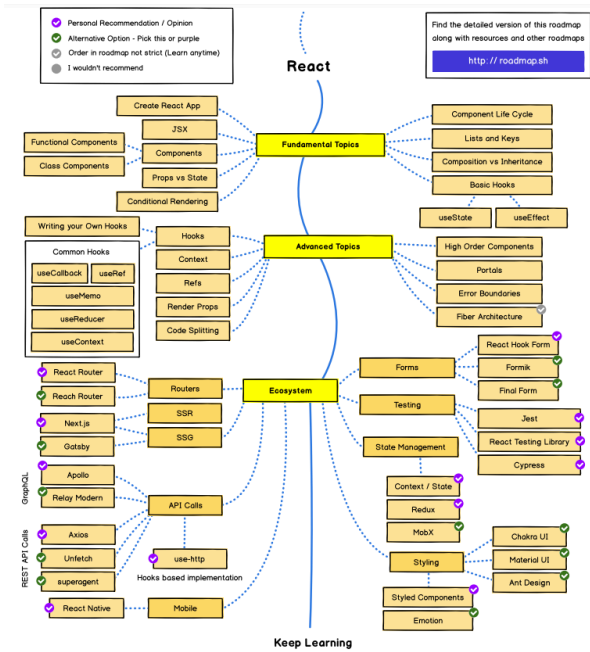


図 4 roadmap.sh³に掲載されている”React.js”のロードマップ (2022 年 10 月 31 日時点)

出典:<https://roadmap.sh/react/>

する。

$$\begin{aligned} \text{真 iff } (n_i \rightsquigarrow n_j \text{ where } n_i, n_j \in \mathcal{V}_{ans} \wedge n_i, n_j \in \mathcal{V}) \\ \vee (n_i \rightsquigarrow n_j \text{ where } n_i, n_j \notin \mathcal{V}_{ans} \wedge n_i, n_j \notin \mathcal{V}) \end{aligned} \quad (10)$$

$$\begin{aligned} \text{偽 iff } (n_i \rightsquigarrow n_j \text{ where } n_i, n_j \in \mathcal{V}_{ans} \wedge n_i, n_j \notin \mathcal{V}) \\ \vee (n_i \rightsquigarrow n_j \text{ where } n_i, n_j \notin \mathcal{V}_{ans} \wedge n_i, n_j \in \mathcal{V}) \end{aligned} \quad (11)$$

ここで、 $u \rightsquigarrow v$ は、グラフ中でノード u から v に到達可能であることを意味する。

本論文では、学習コンテンツに関連する Web リンクをクラウドソーシングにより取得し、提案手法により理解支援関係グラフを作成する。クラウドソーシングプラットフォームは Lancers⁷であり、本実験で理解支援関係グラフを作成する領域である”React.js”⁶を用いた業務に 3 年以上従事した経験があるワーカーにタスクを依頼した。ここで、知識は抽出済みであると仮定し、正解理解支援関係グラフ中の知識を流用する。

5.2 クラウドソーシングによるリンクの収集

”React.js”⁶に関する初歩的な内容から発展的な内容の文献を収集するために、また収集するリンクの重複を避けるために、クラウドソーシングプラットフォーム Lancers⁷にて同領域に 3 年以上の業務経験があるエキスパート 1 人に対してタスクを依頼し、200 件の Web リンクを収集した。リンクに加え、それぞれのリンクには付加情報として 2 つのラベル (「そのリンクの Web ページのみを対象とする」、「そのリンクのサイト全体の Web ページを対象とする」) で分類され、前者のラベルで

分類された Web リンクについては、該当の Web ページを、後者のラベルで分類された Web リンクについては、そのサイト中の Web ページを幅優先探索で可能な限り収集した。結果として 6279 件の文書を収集した。

5.3 実験結果と考察

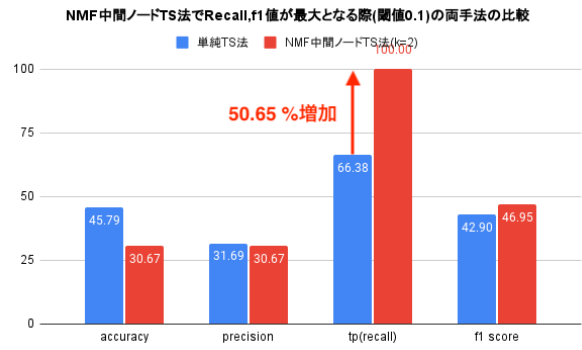


図 5 NMF 中間ノード TS 法で Recall, f_1 Score が最大となる際の両手法のパフォーマンス比較 (閾値 0.1)

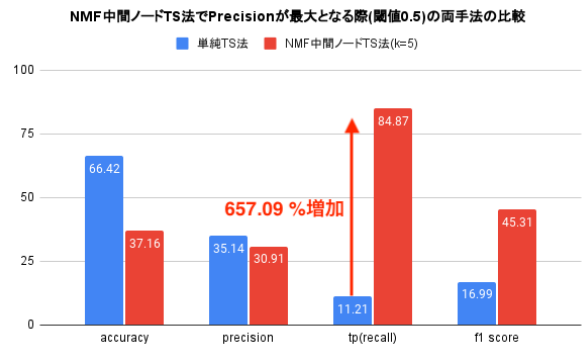


図 6 NMF 中間ノード TS 法で Precision が最大となる際の両手法のパフォーマンス比較 (閾値 0.5)

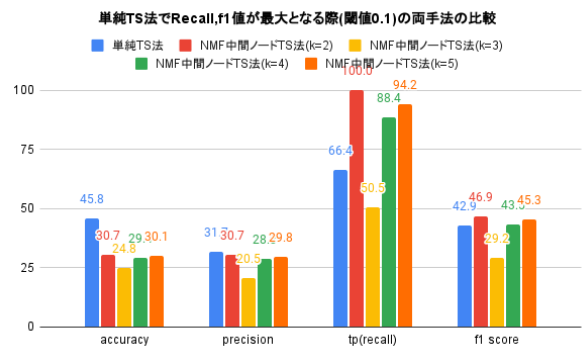


図 7 単純 TS 法で Recall, f_1 Score が最大となる際の両手法のパフォーマンス比較 (閾値 0.1)

NMF 中間ノード TS 法において Recall, f_1 Score, Precision が最大になった閾値 (式 9 における δ) とクラスタ数の組み合わせでの対応する閾値での単純 TS 法のパフォーマンスを比較

⁷ : <https://www.lancers.jp/>

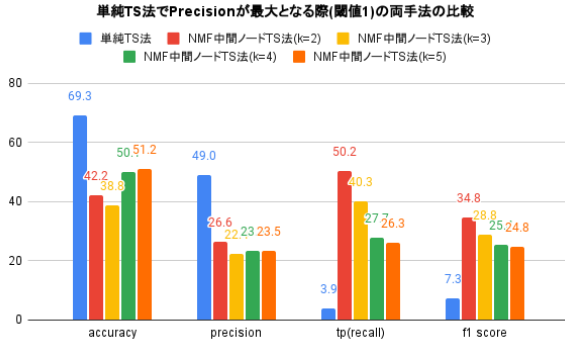


図 8 単純 TS 法で Precision が最大となる際の両手法のパフォーマンス比較 (閾値 1)

した結果をそれぞれ図 5 及び図 6 に示す。両方とも Precision の減少を抑えつつ Recall を大きく増加させている。また、単純 TS 法において Recall, f_1 Score, Precision が最大になった閾値 (式 9 における δ) とクラスタ数の組み合わせでの対応する閾値での単純 TS 法のパフォーマンスを比較した結果をそれぞれ図 7 及び図 8 に示す。図 7 でも同様に Precision の減少を抑えつつ Recall を大きく増加させている。図 8 では他と比較して Precision の減少が大きい、以前として Recall は大きく増加させる結果となっている。

表 1 は式 2 における δ (閾値とする) を 0.1 から 1.0 まで 0.1 刻みで変化させた際の単純 TS 法のパフォーマンスである。これに対し、表 2 は、0.1 から 1.0 まで 0.1 刻みで変化させた閾値と、2 から 5 まで変化させたクラスタ数の組み合わせで理解支援関係グラフを作成した際に、Precision, Recall, f_1 Score がそれぞれ最大の値を取った際のパフォーマンスである。上から順に Precision が最大となった組み合わせ、Recall と f_1 Score が最大となった組み合わせを表している。また、対応する閾値での単純 TS 法によるパフォーマンスとの差が最大になる組み合わせを表 3 に示した。上から Precision が最大となった組み合わせ、Recall が最大となった組み合わせ、 f_1 Score が最大となった組み合わせになっている。表 3 より、記述されている 3 つの組み合わせにおいて、Precision の減少を小さく抑えつつ、Recall を大きく増加させる結果となった。

中間ノードによるエッジの集約により、単純 TS 法では到達できなかったノードに中間ノードを介して到達できるようになったことが今回の結果に繋がったと考えられる。

6 結 論

本論文では、インターネット上にある様々な教材コンテンツを手掛かりに、初学者のために学習項目を関連付けた理解支援関係グラフを自動構築する問題に取り組んだ。具体的には、クラウドソーシングを用いて収集された Web 上の文書集合から、学習トピックに関する知識間の理解支援関係を抽出する Term Subsumption [5] ベースの手法を複数構築し、専門家が作成した学習ロードマップが存在する領域について、それぞれ理解支援関係を抽出した結果を示した。今後の課題としては、本論文

表 1 各 δ (閾値) における単純 TS 法のパフォーマンス

手法	閾値	Accuracy	Precision	Recall	f_1 Score
単純 TS 法	0.1	45.79	31.69	66.38	42.90
	0.2	47.5	30.09	53.76	38.58
	0.3	59.66	32.43	29.08	30.66
	0.4	64.40	35.58	19.83	25.47
	0.5	66.42	35.14	11.20	17.00
	0.6	67.38	35.27	7.60	12.51
	0.7	68.10	38.03	6.35	10.88
	0.8	68.70	42.61	5.89	10.33
	0.9	69.18	47.89	5.33	9.59
	1.0	69.27	49.02	3.92	7.26

表 2 NMF 中間ノード TS 法において Precision, Recall, f_1 Score がそれぞれ最大となった際のパフォーマンス

クラスタ数	閾値	Accuracy	Precision	Recall	f_1 Score
5	0.5	37.16	30.91	84.87	45.31
2	0.1	30.67	30.67	100.00	46.95

表 3 NMF 中間ノード法によって変化したパフォーマンス

クラスタ数	閾値	Accuracy	Precision	Recall	f_1 score
2	0.2	-16.83	+0.59	+46.24	+8.36
4	0.5	-37.00	-6.33	+77.19	+26.46
5	0.5	-29.25	-4.23	+73.67	+28.32

では行わなかった知識の抽出や、機械学習を含めた他の理解支援関係抽出アプローチの評価、自動生成したグラフと教材コンテンツとの連動などがあげられる。

謝 辞

本研究の一部は JSPS 科研費 (JP22H00508, JP21H03552) と JST CREST の支援 (Grant Number JPMJCR22M) の支援を受けたものである。ここに謝意を示す。

文 献

- [1] Hans M. Zell. Digital vs print resources at african university institutions: A discussion document. *African Research & Documentation*, Vol. 138, p. 3–16, 2020.
- [2] Olateju Ajanaku. Utilization of the internet by undergraduate students of the university of ibadan, nigeria. Vol. 10, pp. 30–36, 07 2019.
- [3] Rishabh Ranawat, Ashwin Venkataraman, and Lakshminarayanan Subramanian. Collectiveteach: A system to generate and sequence web-annotated lesson plans. In *ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '21, p. 1–13, New York, NY, USA, 2021. Association for Computing Machinery.
- [4] Yanyan Li and Ronghuai Huang. Dynamic composition of curriculum for personalized e-learning. pp. 569–576, 06 2006.
- [5] Muhammad Nabeel Asim, Muhammad Wasim, Muhammad Usman Ghani Khan, Waqar Mahmood, and Hafiza Mahnoor Abbasi. A survey of ontology learning techniques and applications. *Database*, Vol. 2018, , 10 2018. bay101.
- [6] Joseph D. Novak, D. Bob Gowin, and Gerard T. Johansen.

The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education*, Vol. 67, No. 5, pp. 625–645, 1983.

- [7] Rakesh Agrawal, Sunandan Chakraborty, Sreenivas Gollapudi, Anitha Kannan, and Krishnaram Kenthapadi. Empowering authors to diagnose comprehension burden in textbooks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, p. 967–975, New York, NY, USA, 2012. Association for Computing Machinery.
- [8] Ieee standard for learning object metadata. *IEEE Std 1484.12.1-2020*, pp. 1–50, 2020.
- [9] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, p. 267–273, New York, NY, USA, 2003. Association for Computing Machinery.