

学習内容準拠の単語による発問自動生成手法の有効性について

坂上 公亮[†] 松原 雅文[‡]

岩手県立大学大学院ソフトウェア情報学研究科[†] 岩手県立大学ソフトウェア情報学部[‡]

1. はじめに

学校教育における授業の中で、教員は生徒の学習内容の理解、あるいは思考の促進を図るために生徒に対して「発問」という問いかけを行う。このような授業内の取り組みは学習指導案で計画され、単元内容に沿って行われる。実際の授業で発問を行う場合、使用する用語や文法、学習内容の定着度などの点で、生徒が答えることができる、あるいは、生徒自身の意見や考察を述べられるように配慮しなければならない。

そこで、我々は、任意の用語について、その単語の特徴を活用して発問の自動生成を行う手法を提案している¹⁾。これは、学校教員を対象として、授業内で行われる発問の生成を容易にすることを目的としたものである。発問に使用する任意の1単語からその他の複数の単語を出力することにより、複数通りの単語のパターンを用いて発問文の生成が可能である。

本稿では、学校教育で取り扱われる教科の学習内容に準拠したデータおよび単語を使用して実験を行い、その結果から本手法の有効性について述べる。

2. 関連研究

2.1. 過去問題をリソースとした自動作問システム

選定した単語をもとに問題文を自動で作成するシステムとして、福坂らが提案した手法がある²⁾。知識ベースと出題テンプレートを特定分野の検定問題から構築することで、特定の領域、分野に沿った検定問題を自動で作成することを可能にしている。

この手法では、手動で知識ベースと出題テンプレートの構築を行っており、検定問題に従ったルールベースでの自動作問を可能にしている。しかし、学校教育で取り扱われる知識は国語、地理、日本史、化学などの様々な分野が存在しており、地域、学校ごとに指導範囲が異なる場合がある。そのため、手動での構築および管理は困難であると考えられる。そこで、本手法では教科書や参考書、Web上の電子データなどを利用して、自動で知識ベースを構築する。

2.2. word2vec

Mikolovらは、文法規則などのパターンを分散表現ベクトルに変換し、単語を定量的に利用する手法であ

る word2vec を提案した³⁾。単語ベクトル間のコサイン類似度を活用することにより、単語間類似度の算出、類似単語の取得を可能としている。

本手法では、word2vecの学習モデル作成に学習内容準拠のテキストデータを使用することにより、指導内容に沿った発問の自動生成を行う。

3. 提案手法

3.1. 学習モデル作成

発問文に使用する単語を抽出するための知識ベースにあたる word2vec の学習モデルをコーパスデータから作成する。コーパス作成時、テキストクリーニングによる記号やページ番号などの除去、発問テンプレートに適用不可能な単語の削除を行う。

3.2. 単語出力

入力単語以外で発問に使用可能と考えられる単語の候補を word2vec から取得する。発問文における「生徒が学習済みの単語」を入力単語とし、取得する単語は「解答に導くための補足となる単語」「発問の解答」の2件とする。

3.3. テンプレートの適用

入力単語1件、word2vecから取得した単語2件を発問テンプレートのスロットに適用し、発問文を生成する。テンプレートは発問文と解答からなり、単語が挿入されるスロットが発問文内に2件、解答部分に1件設けられている。発問文内のスロットには「生徒が学習済みの単語」「解答に導くための補足となる単語」、解答部分のスロットには「発問の解答」が挿入される。このテンプレートは挿入される単語の意味、科目指導内容に則った形式で構成する。

以上の手順を行うことにより、発問文に使用する1件の単語から授業で使用可能な発問を複数通り提示する。

4. 実験

4.1. 実験条件

科目指導の内容に沿った単語を生成するために、学習内容準拠のコーパスとして2020年6月17日時点の「世界史の窓¹⁾」穴埋め式ノートを使用する。このコーパスは、高等学校の普通科目「世界史B」に対応しているため、本実験では「世界史B」における発問を取り扱うものとする。発問文に使用可能な単語を取り扱うために、学習させる単語は一般の名詞、固有名詞、自立する動詞のみとする。

¹⁾<https://www.y-history.net/>

Effectiveness of Automatic Question Generation Method using Learning Content based Words

Kosuke SAKAUE[†], Masafumi MATSUHARA[‡]

[†]Graduate School of Software and Information Science, Iwate Prefectural University, [‡]Faculty of Software and Information Science, Iwate Prefectural University

コーパスの学習範囲は「世界史の窓」全範囲とし、形態素解析器には MeCab, 辞書データには mecab-ipadic-NEologd を使用する。また、本実験に使用する学習モデルのオプションを表1に示す。

表 1: 学習オプション

オプション	内容	値
-size	単語ベクトルの次元数	200
-window	学習に使用する前後の単語数	5
-sample	単語を無視する頻度	0.001
-negative	周辺に出現しない単語を類似していない単語とする数	5
-hs	学習に階層的ソフトマックスを使用するか(しない場合は0)	0

学習指導案²で実際に設計された発問文からキーワードを抜粋し、これを発問自動生成の入力単語に使用する。このとき、入力単語に使用する単語は学習モデル中に分散表現ベクトルが存在するものに限る。そこで、入力単語に「フランク王国」を用いて実験を行い、世界史の窓第5章第1節「西ヨーロッパ世界の成立」の単元内容に沿って本実験の評価を行う。

入力単語の分散表現ベクトルとのコサイン類似度が高い順に上位10件まで出力する。これを「解答に導くための補足となる単語」とし、出力単語Aとする。また、出力単語Aそれぞれの分散表現ベクトルとの類似度が高い順に上位10件、計100件を出力する。これを「発問の解答」とし、出力単語Bとする。

4.2. 実験結果と考察

単語出力結果を表2に示す。

表 2: 単語出力結果

入力単語	出力単語 A	出力単語 B (重複なし)
フランク王国	統一	
	中心	建設, 支配, 後出, 地方, 中心,
	破る	現在, 時代, 進出, 征服, 成立,
	時代	貴族, 始まる, 発達, ローマ, 明,
	後出	強まる, 続く, 建国, 侵入, イギリス,
	前出	右, 統一, 中国, 貿易, 皇帝,
	キリスト教	意味, 勢力, 前出, 諸国, ドイツ,
	侵入	フランス, 1世, 国王, 認める,
	続く	広がる, 都, 主張
地方		

「世界史の窓」全範囲を学習しているため、「統一」「支配」「時代」のような科目特有の単語、「前出」「後出」「意味」などの解説テキストで頻出の単語が多く出現した。

学習指導案²で実際に設計されている発問のうち、入力単語と表2に示した出力単語Aのいずれもが含まれる発問を以下に示す。

- 481年に全フランク人を統一し、フランク王国を建国したのは誰か。
- フランク王国がブルグンド王国を滅ぼして全ガリアを統一したのはいつ頃か。

これらの発問には入力単語の「フランク王国」、出力単語Aの「統一」が含まれている。これは、学習指導案における使用教材が「世界史の窓」と同様の教材であったことから、「フランク王国」と「統一」の関連性が高いと学習モデルが解釈した結果であると考えられる。

しかし、解答にあたる出力単語Bでは、実際の発問に対する解答として適切なものは出現しなかった。これは、実際の解答である「クローヴィス」という単語がコーパス内に頻出する単語ではなかったため、word2vecの学習モデル作成時に除外されてしまったことが原因であると考えられる。また、「～はいつ頃か。」のような時期を問う問題に対しては、「711年」「6世紀半ば」のような定量的解答、「〇〇の戦い後」「帝政末期」などの複数単語で構成される解答が想定されるため、word2vecによる単語単位での対応は困難であると考えられる。

5. おわりに

本稿では、発問自動生成手法に学習内容準拠のコーパスを使用して行った実験について述べた。実験の結果、任意の単語から解答の補足となる単語が出力でき、これらが実際の指導案での発問に使用されていることを確認できたことから、本手法の有効性が示された。

今後は、考察で示した「クローヴィス」のような頻出ではない重要な単語が除外されないよう、学習オプションを変更して評価実験を行う予定である。時期を問う発問は単語単位での解答提示が困難であるため、単語の前後関係をもとに複数の単語を組み合わせて提示する方法を検討する。また、指導案に示されている発問文から、単語を挿入することで有効な発問文となるテンプレートの作成も行う予定である。

謝辞

本研究の一部は JSPS 科研費 18K11358 の助成を受けたものである。

参考文献

- 1) 坂上 公亮, 松原 雅文: “学校教員向け発問自動生成手法の提案”, 2019年度電気関係学会東北支部連合大会 (TSJC2019), 2J07 (2019).
- 2) 福坂 祥基, 高木 正則, 山田 敬三, 佐々木 淳: “過去問題をリソースとする知識ベースを活用した問題自動生成システムの開発と評価”, 情報教育シンポジウム (SSS2016), pp.39-46 (2016).
- 3) Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean: “Distributed Representations of Words and Phrases and their Compositionality”, Advances in Neural Information Processing Systems 26 (NIPS 2013), pp.3111-3119 (2013).

²沖縄県立那覇西高等学校 世界史 B 学習指導案
入手先: <<http://www2.open.ed.jp/data/38761/01.doc>>