

科目推薦のための doc2vec の応用方法の検討

竹森 汰智† 亀井 清華‡

広島大学 大学院工学研究科†‡

1. はじめに

大学では、学生自身が時間割を作成しなければならない。学生が時間割を作成する際には、大学で配布されるシラバスや学生便覧を参照する。しかし、これらに掲載されている科目情報は膨大である。そのため、自分が興味のある学問分野の科目を探し、履修したいと思うような科目を見つけることは容易ではない。特に学部新入生にとって、それは負荷の大きい作業であると言える。

本稿では、学部新入生を対象に教養科目を推薦するためのシステム設計を行う。各科目の特徴を表現するために、doc2vec [1]を用いて科目毎のシラバスを分散表現する。得られたベクトルを科目ベクトルと呼ぶことにする。その科目ベクトルに対して階層クラスタリングを行い、クラスタ毎にキーワードをワードクラウドで表現することで科目集合の性質を可視化する。また、高校のシラバスをベクトル化したものとの距離に着目し、それをレーダーチャートで表現する。学生には、これらのワードクラウドとレーダーチャートを選択肢として提示する。これにより、学生の学びの志向とマッチする科目を選択しやすいうように支援する手法を提案する。

2. 関連研究

履修のための支援システムの既存研究としては、以下のものが挙げられる。

西森ら [2]は、学生の履修履歴を用いた科目推薦システムを提案した。具体的には、学生の履修履歴を元に得意な科目と不得意な科目を推薦する。しかし、この手法では履修履歴を持たない学部新入生には適応できない。

関ら [3]は、階層クラスタリングを用いたシラバス分析手法を提案した。具体的には、シラバスをその単語の TF-IDF 法でベクトル表現し、階層クラスタリング手法の各クラスタ間距離尺度の比較を行った。そして、ウォード法の有効性を示した。

3. 提案システム

提案システムでは、シラバスから抽出した科

目情報に基づき、科目に対して階層クラスタリングを行う。そして、樹形図の根の子から順に、予め決めた選択肢の数 k 個のクラスタをユーザーに提示する。ユーザーが選択肢から1つ選ぶと、そのクラスタから樹形図を葉に向けて辿り、次の k 個のクラスタを提示する。ユーザーが選択を繰り返すことで、候補科目を絞りこんでいき、最終的には一つの科目を選択できるものとなっている。図 1 に、本システムの構成を示す。

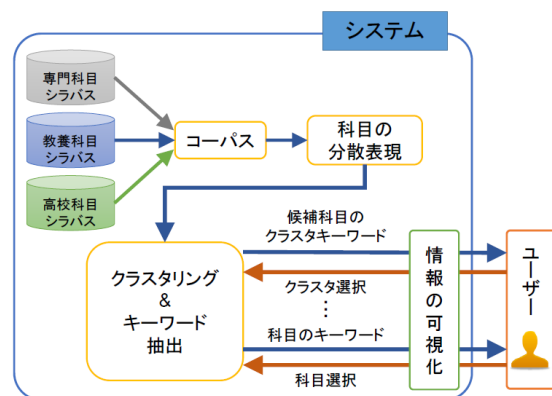


図 1 : システム構成

まず、シラバスの科目情報を分散表現する方法について説明する。科目情報は、シラバス内の「科目名」や「授業目標」、「授業計画」など科目の特徴を表すと思われる要素のみに限定する。各科目情報から名詞のみを抽出したものを、科目特徴語と呼ぶことにする。各科目の科目特徴語をもとに、doc2vecを用いてその科目の科目ベクトルを獲得する。また、doc2vecの学習時には、推薦対象である教養科目に加えて、学部の専門科目と高校科目のシラバスを利用する。これにより、学問体系を網羅した空間における教養科目の科目ベクトルが得られる。その後、科目ベクトル集合に対して、ウォード法を用いてクラスタリングを行う。

次に、クラスタの提示方法について説明する。各クラスタに属する全ての教養科目の特徴語を、クラスタの特徴語とする。本システムでは、各クラスタの可視化にその特徴語のワードクラウドと、高校の科目との距離を可視化したレーダーチャートを用いる。

ワードクラウドの作成のためには、各クラスタの特徴語に対して、TF-IDF 法で重みづけを行う。TF-IDF 値が大きい単語ほど、そのクラスタ

Consideration of application method of doc2vec for course recommendation

† Daichi Takemori, Hiroshima University

‡ Sayaka Kamei, Hiroshima University

の特徴を表していると言える。重みの大きい特徴語を大きく、小さい特徴語を小さく表示することで、クラスタの特徴を一目で印象付けることを期待する。

レーダーチャートの作成のためには、各クラスタに属する教養科目と高校の各主要科目との類似度を以下の手順で計算する。高校の主要科目には、国語、数学、理科、社会、英語の5科目がある。

- ① 高校の主要科目毎に科目ベクトルの平均を算出する
- ② 全教養科目に対し、①のベクトルとの類似度を算出する
- ③ 各クラスタに属する教養科目がもつ②の平均値を算出する
- ④ 全教養科目がもつ③の最小値を 0，最大値を 5 として正規化する

図 2 に、提示するワードクラウドとレーダーチャートの組の例を示す。

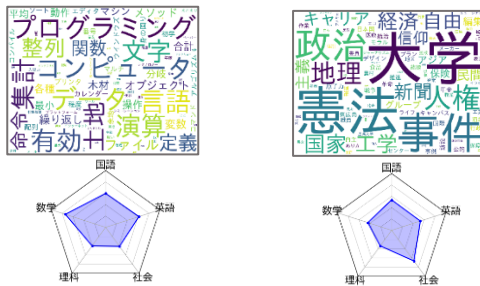


図 2 : システムが提示するワードクラウドとレーダーチャート

4. 実験評価

本実験では、システムがユーザーに提示するワードクラウドによって、科目の特徴を表現できているかを確認するために、アンケートによる実験を行った。シラバスとしては、広島大学シラバス¹と第一学習社の教科書シラバス案²の科目情報を利用した。推薦対象の科目は、科目特徴語を 50 語以上もつ教養科目 146 科目とした。doc2vec の学習には、総合科学部、文学部、法学部、経済学部、理学部、工学部、生物生産学部の講義形式の専門科目 1391 科目と、高校の主要科目に該当する 22 科目のシラバスを利用した。

システムが提示する選択肢として、 $k = 5$ とし、1 回目の選択肢に該当する 5 つのクラスタのワードクラウドとレーダーチャートを作成した。各ワードクラウドは、クラスタの特徴語の TF-IDF 値の上位 100 語から構成した。

そして、情報工学を専攻する学部生・大学院

生 20 名を被験者とし、以下の実験を行った。

[実験 1] : 教養科目 15 科目のシラバスと、前述の 5 つのワードクラウドを見て、各科目がどのクラスタに属するか回答してもらった。対象とした 15 科目は、各被験者に対してランダムに選出した。全被験者から得られた回答の正解率は、70.48%であった。

[実験 2] : 興味がありそうなワードクラウドと履修したい科目の一致度を確認した。まず、各被験者に、5 つのワードクラウドから興味がありそうなものを一つ選択してもらった。次に、5 科目のシラバスから履修したいと思える科目を一つ選択してもらった。5 科目の選択方法は以下の 2 通りを試した。

リスト A : 各クラスタから科目特徴語の数が最大のもの

リスト B : 各クラスタから以下の点数が最大のもの。点数は、ワードクラウド中の単語を TF-IDF 値の大きい順にランキングした上で、各科目特徴語に対して、1 位の単語である場合は 100 点、2 位の単語は 99 点、…として加算する。

表 1 に、結果を示す。科目とクラスタが一致しなかった理由を調査すると、ワードクラウドに大きく出現する単語が被験者に与える印象が大きすぎたことが分かった。

表 1 : 提示した科目とワードクラウドの一致率

リスト A	35.00 %
リスト B	65.00 %

[実験 3] : ワードクラウドにレーダーチャートを付加することで、科目群のイメージが分かりやすくなったかどうか回答してもらった。被験者の 90 %が、レーダーチャートがあった方がイメージしやすいと回答した。

5. おわりに

科目クラスタリングによって候補科目を絞り込むことで、履修支援を行うシステム設計と、そのためのインターフェースの検討を行った。今後の課題として、ユーザーへ提示する選択肢の改善や提案手法を用いたシステムの構築、そして、学部新入生に実際にシステムを使ってもらい、実証的な評価を行うことが挙げられる。

参考文献

- [1] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. In International Conference on Machine Learning, Vol. 14, pp. 1188–1196, 2014.
- [2] 西森友省, 堀幸雄, 今井慈郎. 履修履歴を用いた科目推薦システム, 情報処理学会第 75 回全国大会, 2013
- [3] 関陽一. 科目の関係性に基づくシラバス分析手法に関する研究, 筑波大学修士学位論文, 2014

¹ <https://momiji.hiroshima-u.ac.jp/syllabusHtml>

² www.daiichi-g.co.jp/kyoka