

LDA を用いた大学シラバスの分析

— 進路指導に役立つ検索システムの提案と評価 —

富田 晟央[†] 小松 昌史^{††} 加納 徹[†] 赤倉 貴子[†]

[†]東京理科大学工学部情報工学科 ^{††}東京理科大学大学院工学研究科経営工学専攻

1. はじめに

大学進学を目指す学生は、学習内容を重要視しているが、事前にその内容を把握することは困難な場合がある。例えば、シラバスの閲覧は学習内容を知る手段の一つであるが、膨大な科目数を横断的・網羅的に閲覧するには時間を要する。そこで本研究では、トピックモデル分析手法の一つである LDA (Latent Dirichlet Allocation) [1] を用いて、シラバスを分析する。そして、LDA の結果から得られるトピックを検索の『キーワード』として用いることで、選択されたキーワードとその関心度に基づき、大学・学部・学科を検索可能な進路指導支援システムを提案する。システムは、閲覧の容易さから、Web アプリケーションとして開発する。

2. 分析手法・結果

国立大学 5 大学から収集したシラバスデータを文系・理系に分類し、授業目標、授業概要、授業計画に該当するテキストを対象に、LDA を用いた分析を行った。ここで、トピック数の決定には Perplexity を使用し、学科種類数とシステムの一望性を加味し、文系 30、理系 35 トピックとした。LDA の結果から命名したトピックの例を表 1 に示す。

表 1. トピックの例

トピック	プログラミング	地球・宇宙	言語・文化	創作・鑑賞
代表単語	C 開発 ソフトウェア アルゴリズム ネットワーク	地震 観測 気象 太陽系 大気	日本語 外国 音韻 歴史 文法	制作 美術 デザイン 鑑賞 芸術

LDA を用いた分析結果は、シラバス単位の各トピックの構成割合を明らかにする。この値を学科単位に集約するため、学科ごとに平均値(平均トピック割合)を算出した。

3. システム開発

システム開発にあたり、インタフェース上で選択されたキーワードとその重みづけを、平均トピック割合と比較する必要がある。そこで、キーワードと関心度の組み合わせに応じた、仮想のトピック割合を生成する。関心度は 0~3 の 4 段階とした。例えば、トピック A の関心度が 2、トピック B の関心度が 3 の場合、仮想トピック割合はそれぞれ 0.4、0.6 となり、それ以外のトピック割合は 0 となる。

利用者がキーワードと関心度の組み合わせを変更する度に、仮想トピック割合も更新される。同時に、システムは平均トピック割合と仮想トピック割合の類似度を計算し、類

似度が高い順に大学・学部・学科をソートする。類似度は、トピック割合をベクトルとみなし、Cosine Similarity を用いて算出した。インタフェースを図 1 に示す。

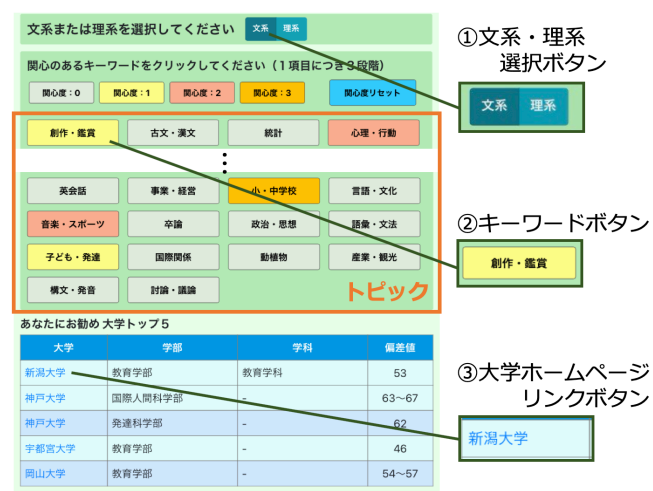


図 1. システムのインタフェース

システムの利用者はまず、文系・理系を選択する。この選択によって表示されるキーワードが異なる。キーワードとその関心度を任意に変更でき、キーワードと関心度の組み合わせに応じて、画面下に大学・学部・学科が 5 つ表示される。大学名はその大学のホームページとリンクしている。

4. 評価実験

本システムを学習塾の進路指導員 4 名に試用してもらい、評価アンケートとインタビューを行った。その結果、「キーワードは学科の専門性を連想できる内容だったか」、「本システムは進路指導に役立つと思うか」という質問項目について、いずれも平均 4.0 (高評価 5) を上回った。また、一見関連性の薄いキーワードの組み合わせであっても、必ず学部学科を提示できる点、キーワードが一般用語に近く、関心と結びつけやすい点などが回答理由として挙げられた。

5. まとめ

LDA によるシラバスの分析結果を用いた進路指導支援システムは、評価実験の結果より、有用であることが示唆された。今後の課題として、大学数の充実、トピックの細分化・統合、および利用履歴に基づく適応的な進路指導支援などが挙げられる。

参考文献

- [1] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning*, vol.3, pp.993-1022, 2003.