

論文の内容の要旨

論文題目	LDAを用いたレポート推薦システムの開発
学 位 申 請 者	加藤 嘉浩

近年、学習理論の主流は、Vygotsky に代表される社会的構成主義に移行しつつある。Vygotsky は、人の知識構築は単なる知識の伝達ではなく、対象の理解の仕方への支援としてモデル化している。初心者、熟達者に問題解決や対象理解を支援してもらうことにより、最初は表層的ではあるが、徐々に、単なる知識のみでなく、理解の仕方、注意・焦点化、内省、態度、動機、情熱などの対象に関する高次の心的スキルを獲得できると主張している。このモデルに従えば、教師は学習対象の面白さや情熱、見方や価値観、倫理、その背景、文化を伴って支援するので、教師の対象の見方そのものを獲得できる。また、初心者は熟達者から一方的に支援されるのではなく、意識的に他者から学ぼうとしており、観察や模倣、他者との比較などが行われる。学習者の発達に伴い、熟達者の支援がなくても自律的に他者からの学びが行われると考えられる。

植野(2008)は、初心者は熟達者から支援されることが主であるが、徐々に発達して学習者自身からの観察・模倣といった自律的な他者からの学びができるように変化すると述べている。そして、この変化が発達の本質であると述べている。本論文では、このモデルに従い、レポートライティングにおける「他者からの学び」を支援するシステムを提案する。具体的には、過去の熟達者のレポートを学習者に適応的に推薦し、レポートライティングにおける「他者からの学び」を支援する。

これまでに、レポートライティングを支援するシステムが多数開発されている。しかし、これらは「導入、背景、目的、方法、結論」などの論文の文章構造の構築を形式的に支援するものである。本提案では、他者のレポートを学習者に推薦し、自分のレポートと比較することにより、レポートの内容を深く推敲する機会を多く作るだけでなく、他者のレポートライティングの方法を学ぶことができると考える。この場合、どのように学習者にレポートを推薦するかが問題である。

これまでに教育分野では、多くの推薦システムが開発されており、いずれも学習者データと類似性が高いコンテンツや人、メッセージを推薦している。しかし、このような従来手法をレポート推薦に適用する場合、内容・表現が類似のレポートばかりが推薦されてしまい、効果的な学習が期待できない。

そこで本論文では、できるかぎり主題は似ているが内容（用いられる単語分布）が異なるレポートを推薦する手法を提案する。これにより、主題は同じでも様々な構成や表現、オリジナリティのレポートが推薦できると期待できる。ただし、ここでいう「構成」とは「導入、背景、目的、方法、結論」などといった形式的な構成ではなく、レポートの主張点の論理構成や文章の流れを意味する。技術的には、Latent Dirichlet Allocation(LDA)を用いて、学習者と他者のレポート間のトピック分布距離を計算することで類似の主題を持つレポートを同定する。LDAを用いて推定されるトピックは、意味が同じで異なる単語も同一のトピックとして推定できる。すなわち、LDAはトピック分布と表層的な単語分布を分離して扱うことができることが特徴ともいえる。本論文では、同じ主題であれば、単語分布がレポートの内容を反映していると仮定する。そして、他者からの学びは、単一の他者のみからよりも多様な他者からの学びの方が効果的であることが知られており、提案手法により、より効率的な学習ができると期待できる。被験者実験により提案手法の有効性を示した。

この手法では、あらかじめLDAのトピック数を決めておく必要があり、専門家によるレポートの分類結果に即したトピック数を用いた。しかし、データが大量になった場合や新たにデータを追加する際に人手による分類を作成しなおす必要があり、システムを利用する上で現実的ではない。また、人手による分類に即したトピック数が、モデルの学習・推定精度を高くする保証はない。そこで本論文では、自動的にトピック数を決定する手法として、LDAのハイパーパラメータを1としたときの周辺尤度を最大化することにより、LDAのトピック数を推定する。トピック数の推定値は、ハイパーパラメータの値に敏感である。ハイパーパラメータを大きくするとトピック数は過小評価され、ハイパーパラメータを小さくするとトピック数は過大評価される。また、ハイパーパラメータを1としたときトピック数を最も正確に推定できることを漸近解析、シミュレーションにより示した。本システムに組み込むことで、その有効性を示した。

論文審査の結果の要旨

学位申請者氏名 加藤 嘉浩

審査委員主査 植野 真臣

委員 大須賀 昭彦

委員 栗原 聡

委員 広田 光一

委員 田原 康之

委員 川野 秀一

近年，高等教育におけるライティング教育の重要性が指摘されている。しかし，初心者には，独力でレポートを書き上げることは難しい。レポートや論文執筆においてその論理構成や論証をサポートする研究は多くなされている。

加藤嘉浩氏は，局所的なライティングスキルを支援するだけでなく，他者のレポートから能動的に学習者が学ぶことが重要であると考え，徒弟的アプローチに基づき，過去の優秀なレポートを適応的に推薦することにより，初心者のレポートライティングを支援する手法を提案している。

従来からもレポート推薦システムは提案されている。しかし，それらはユーザのレポートとのTF-IDF法によるコサイン類似性の高い文書をデータベースから検索してくるというものである。学習のための他者レポート推薦という状況を考えてみると，自身のレポートに類似のレポートが多く推薦された場合，類似のレポートばかりが推薦されてしまう可能性がある。加藤嘉浩氏は，なるべく類似のトピックを扱ったレポートであるが，表現や論理構成，オリジナリティがなるべく多様になるように他者レポートを推薦するシステムのほうが，類似のレポートを推薦するシステムよりもより学習効果が高いと考えた。

具体的には，Latent Dirichlet Allocation (LDA) を用いて，LDAのトピック分布と単語分布を分離できる特徴を用いている。対象学習者のレポートとトピック分布のカルバックライブラー距離が小さく，単語分布のカルバックライブラー距離が互いに大きい評価の高いレポートを推薦するシステムを提案している。これにより，学習者は自分と同じ主題を扱う多様な過去の優秀なレポートから，レポートライティングにおける多様なスキルを学べると期待できる。

第1章では，本論の背景，目標について書いている。

第2章では，本推薦システムで用いる学習者のレポートデータを蓄積しているLMS“Samurai”，レポート推薦システムの関連研究を紹介する。

推薦システムの関連研究を、レポートライティング支援システムと教育分野における推薦システムに大別し紹介している。多くのレポートライティング支援システムは、「導入、背景、目的、方法、結論」といった形式的な構成を解析し、学習者の論文構成を可視化や指摘するシステムが多い。教育分野における推薦システムは、機械学習手法や時系列モデル、オントロジー手法を用い、学習者の学力や興味に応じたコンテンツを推薦するシステムである。このような従来の推薦システムは、いずれも学習者データと類似性が高いコンテンツや人、メッセージを推薦している。また、本推薦システムで用いる文書分類手法の Latent Dirichlet Allocation (LDA) を紹介し、その推定手法と LDA を用いる上で必要となるトピック数の決定に関する関連研究を紹介している。

第 3 章では、LDA を用いたレポート推薦システムを提案している。使用したレポートは、植野(2004)が開発した LMS "Samurai" に蓄積されているレポートデータを用いている。本提案システムは、レポートライティングにおける「他者からの学び」を支援することを目的としている。そのため、従来のレポートライティング支援システムのような学習者のレポートの形式的な構成を解析し、学習者が着目すべき箇所を指摘する手法ではなく、学習者に他者のレポートそのものを推薦する。その際、どのようなレポートを推薦することで、学習者に有用であるかが問題となる。本章では、できるかぎり主題は似ているが内容(用いられる単語分布)が異なるレポートを推薦する手法を提案する。これにより、主題は同じでも様々な構成や表現、オリジナリティのレポートが推薦できると期待できる。ただし、ここでいう「構成」とは「導入、背景、目的、方法、結論」などといった形式的な構成ではなく、レポートの主張点の論理構成や文章の流れを意味する。また、実際の理工系大学生を対象に評価実験を行い、本提案の有効性を示した。

第 4 章では、LDA を用いる際に、予め決定しておく必要があるトピック数の決定手法について述べている。第 3 章において、本提案システムでは、専門家による評価データを用いた分類精度から決定した。しかし、この手法では、データが大量になった場合や新たにデータを追加する際に人手による分類を作成しなおす必要があり、システムを利用する上で現実的ではない。また、人手による分類に即したトピック数が、モデルの学習・推定精度を高くする保証はない。そこで、本章では、トピック数を変え、LDA の周辺尤度を計算し、周辺尤度の値が最も高くなる時のトピック数をモデルの真のトピック数として採用する。周辺尤度を計算する際、LDA のハイパーパラメータが結果に大きく影響することをシミュレーションにより示した。結果として、LDA のハイパーパラメータを 1 としたときに、LDA のトピック数を推定できることをシミュレーションにより示した。この結果を本推薦システムに組み込むことで、その有効性を示した。

最後に第 5 章では、本研究で得られた主な研究成果を統括し、本論文をまとめるとともに本研究の課題について述べている。

加藤嘉浩氏のレポート推薦システムは、これまで提案されてきたレポート支援システムと相補的な関係にあり、十分実用的である。また、これまで難しいとされていた LDA におけるトピック数推定のためにその周辺尤度とハイパーパラメータの関係を明らかにし、最適な設定法を導き、機械学習、統計学の分野にも貢献している。

よって、本論文は博士(工学)の学位請求論文として十分な価値を有するものと認める。