

例示型英文書作成支援ツールにおける例文のレコメンドシステム

Example Recommendation System for Example-Based Web Application to Support Writing Technical Documents in English

池田 善博^{*1}, 宮崎 佳典^{*2}, 田中 省作^{*3}
 Yoshihiro IKEDA^{*1}, Yoshinori MIYAZAKI^{*2}, Shosaku TANAKA^{*3}

^{*1} 静岡大学情報学部
^{*1}Faculty of Informatics, Shizuoka University
^{*2} 静岡大学大学院情報学領域
^{*2}College of Informatics, Shizuoka University
^{*3} 立命館大学文学部
^{*3}College of Letters, Ritsumeikan University
 Email: cs14006@s.inf.shizuoka.ac.jp

あらまし：技術英文書を作成する際に、作成者は適切な語や構文を選択する目的で、専門家によって書かれた技術英文を参考にするという方法をしばしば取る。我々は、技術英文書作成を支援する Web アプリケーションの開発を行っている。ユーザが英作文の参考にした例文は、より理解が容易で参考になる表現を多く含んでおり、これらをユーザ同士で共有することでより効率的な英文作成が可能になると考えられる。本研究では、参考例文の共有を目的として、トピック分析でユーザの検索トピックを推定し、それに合致する他者の参考例文を提示するレコメンド機能を実装した。本稿では、その実験概要と結果を報告する。
キーワード：英作文支援、レコメンド、トピック分析、Web アプリケーション

1. はじめに

技術英文書を作成する際に、作成者は適切な語や構文を選択する目的で、専門家によって書かれた技術英文を参考にするという方法をしばしば取る。

先行研究において、宮崎ら⁽¹⁾は英語技術文書の例文検索システム（以後、本システム）を開発した。本システムには約 29 万件の例文が技術文献から登録されており、ユーザは英文で検索して例文を取得する。参考にした例文をデータベースに登録することが可能である（以後、これらを参考例文と呼ぶ）。

ユーザが数ある例文から選択した参考例文は、より有益だと考え、他者の参考例文を共有することをコンセプトとして本システムにレコメンド機能を実装した。レコメンド機能の関連研究には、質の低い知識の提示を防ぐため、知識の質の定量的評価指標を作成した知識共有型レコメンドシステム⁽²⁾がある。

2. システムの概要

本システムは、類似文検索とフレーズ検索の 2 種の検索機能から成る。類似文検索では、ユーザの入力した英文およびコーパス内の技術英文の n-gram, 品詞, 同義語を考慮したうえでコサイン類似度を算出する（既存の類似度算出方法）。そして、コサイン類似度が高い上位 50 件の技術英文をユーザに提示する。フレーズ検索では、ユーザの入力したフレーズが含まれる技術英文を提示する。入力するフレーズは品詞や任意の一単語を指定でき、例えば”a common [N]” ([N]は名詞を表すシンボル) と入力することで [N] が任意の名詞の単語に置換され”a common ancestor”や”a common feature”を含む技術英

文を検索することができる。

新たに追加した例文評価機能は、参考例文登録時に評価をつける機能である。評価値は 1~3 の 3 段階あり、ユーザが例文を見て、(1:少し, 2:まあまあ, 3:とても)参考になった、と感じた参考度合いに対応した数値が評価値として保存される。同じく新規機能のレコメンド機能では一度に 3 件のレコメンド文が表示され、各例文に対する評価人数や平均評価値などが確認できる

3. レコメンド手法

レコメンド機能におけるアルゴリズムの概要図を図 1 に示す。

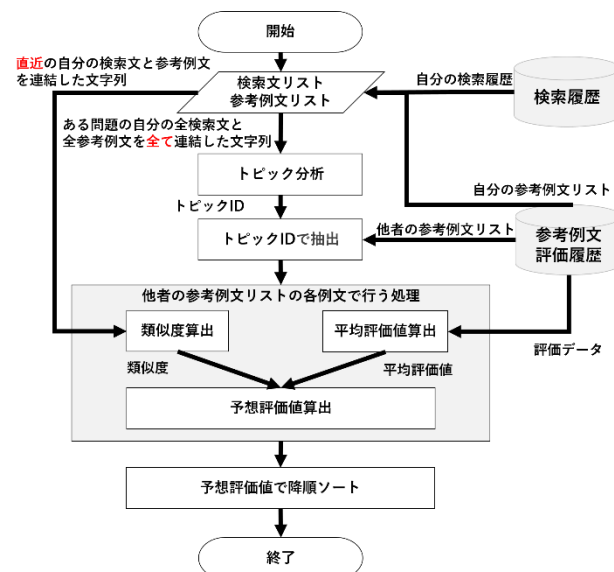


図 1 レコメンドアルゴリズム

3.1 トピック分析

文章から潜在的なトピック（単語の出現頻度を変動させる要因となるもの）を統計的モデルによって推定するためのモデルをトピックモデルという。本手法ではトピックモデルの Hierarchical Dirichlet Process (HDP, 階層ディリクレ過程) を用いた。本システムの技術文献コーパスを学習させた HDP モデルに、ある文章を入力すると、その文章のトピックと対応するトピック ID が得られる。本手法では、ユーザの検索トピックを得るため、「自分の参考例文と検索文を連結させた文字列」を入力し、それと同一のトピック ID を持つ他者の参考例文リストを抽出した。

3.2 類似度算出

他者の参考例文リストのある例文を S_{other} 、「自身の直近の検索英文と自身の直近の参考例文を連結させた文字列」を S_{user} とする。英語版 Wikipedia を学習コーパスとして用いた word2vec モデルから、それぞれの文章中の単語のベクトル表現（200 次元のベクトル）を取得する。本研究では、文章に含まれる全単語のベクトルの平均を、その文章のベクトルとし、文章のベクトル \vec{s}_{other} および \vec{s}_{user} を得る。類似度 Sim は 2 つの文章のベクトル間のコサイン類似度とした。類似度は以下のように表される。

$$\text{Sim}(\vec{s}_{other}, \vec{s}_{user}) = \frac{\vec{s}_{other} \cdot \vec{s}_{user}}{|\vec{s}_{other}| |\vec{s}_{user}|}$$

3.3 予想評価値

レコメンドには「ユーザの直近の検索文に類似しており、かつ評価が高い例文を上位に表示する」のが良いと考え、その指標となる値を「予想評価値」と呼ぶことにした。ある例文 S の平均評価値が $\text{AvgRating}(S)$ （なお、値域は 1 から 3 の実数値）で表されるとき、3.2 節の S_{other} と S_{user} から求める予想評価値は類似度と平均評価値の積で求め、以下のよう表される。

$$\text{pred}(S_{other}, S_{user}) = \text{Sim}(\vec{s}_{other}, \vec{s}_{user}) \cdot \text{AvgRating}(S_{other})$$

レコメンド文には他者の参考例文を予想評価値が高いものを順に表示させた。

4. 現行ツールを用いた実験

4.1 実験の概要

レコメンド機能の評価を得る目的で、某大学の大学生 8 名を対象に「部分訂正問題」および「自由英訳問題」を各 8 問ずつ、計 16 問の問題に本システムを用いて解答してもらった。本問題は前実験⁽³⁾と同じ問題を用いた。それぞれの問題カテゴリの内容を表 1 に示す。解答の際に参考にした例文には例文評価機能を用いて評価することを指示した。実験の終了後にレコメンド機能に関するアンケートを実施し、レコメンド機能に関する所感と改善点を調査した。

表 1 実験で用いた問題の詳細

カテゴリ	問題の詳細
部分訂正	日本語とその英訳（要訂正箇所あり）が与えられ、より自然な英文に修正する
自由英訳	問題となる日本語が与えられるので、それを英訳する

4.2 A/B テスト

本レコメンド手法の有効性を検証するため、問題解答の途中で、異なる抽出法で提示される 2 例文のどちらかを選択するテストを実施した。本実験では問題解答の参考になる方を選択するよう指示した。各手法の選択回数を Fisher の正確確率検定で分析し、本手法と他の例文提示手法との比較を行った。

4.3 実験結果

本実験で登録された参考例文の総数は 276 件であり、うち 56 件がレコメンド機能から登録された。

本レコメンド手法（手法 A）とトピック分析を、その問題で登録された参考例文の抽出に置換した手法（手法 B）の A/B テスト結果を表 2 に示す。

表 2 レコメンド文と検索例文の選択回数

手法 A（トピック）	手法 B（問題番号）	p 値
19	14	0.325

アンケートでは、「前の問題で登録された参考例文がレコメンド文に多く表示された」という意見が見受けられた。ユーザの検索文にマッチするレコメンド文をあまり提示できていなかったからだと考えられる。今後、検索機能で用いられている既存の類似度算出方法を用いるよう改善する方針である。

その他の実験結果やアンケート結果については、発表当日に分析および考察を行う。

5. まとめ

本研究では、技術文書検索システムに登録された他者の参考例文をレコメンドする機能を追加した。全参考例文の 20% がレコメンド機能から登録され、ユーザ間で知識の共有が確認された。

参考文献

- (1) Y. Miyazaki, S. Tanaka, Y. Koyama: Development of a Corpus-Based Web Application to Support Writing Technical Documents in English, E-Learn 2014, Vol. 2014, No. 1, pp. 1371-1380 (2014)
- (2) 鳥羽美奈子, 森靖英, 田代大輔, 知識共有型レコメンドシステム “Knowledge Recommender” の提案とビル省エネ管理事業への適用, 情報処理学会論文誌, 53(1), pp. 149-162 (2012)
- (3) 天野翼, 宮崎佳典, 田中省作, 長谷川由美, 簡略化を用いた例示型英文書作成支援ツールの開発と検証, 統計数理研究所共同研究リポート 397, pp. 13-24 (2018)