

技能継承を支援することを目的とした対話的な文章可視化ツールの 提案

* 郷原浩之⁽¹⁾ , 西原陽子⁽¹⁾ , 大澤幸生⁽¹⁾

(1) 東京大学大学院工学系研究科システム創成学専攻

東京都文京区本郷 7-3-1

E-mail : tt096331@mail.ecc.u-tokyo.ac.jp

Abstract: 本研究では on the job training(OJT)によってなされる技能継承にかかる時間を減らすことを目的として、対話性を有した文章可視化ツールを提案する。分析対象のテキストは、技能保有者によっていくつかの性質に分類される。その分類を教師として Polynomial Semantic Indexing(PSI)を用いて、文章中の全単語について他の単語との類似度を計算する。そして、単語間の類似度を用いて Kamada-Kawai 法によって単語を平面上に配置する。単語と各分類との類似度によって単語の色を定義し、平面上に色付けを行う。

1. はじめに

業務を遂行するために必要な技能を世代ごとに継承させていくことは組織の経営において重要な課題の一つである。しかしながらこのようなノウハウは個々人の経験に基づくものが大きいために、他者に伝わるようにその内容を説明することは難しい。

このような技能継承の問題は構成員が入れ替わる限りあらゆる組織で起こりうる。卑近な例だと、私が所属している東京大学工学部広報誌 **T time!**の編集を行う広報アシスタントでも技能継承は問題となっていた。広報アシスタントは学部3年生から修士課程2年までの学生10名程度で構成される。記事の編集経験が豊富な修士課程2年の学生が書く記事は、他の記者の記事よりも明確に優れている。学部3年生と経験豊富な学生を取材に同行させ、記事の推敲を全員で行うなどして技能を継承する機会を設けているが、現状では効果が見られない。そのため、彼らが大学院を修了した後で、同等の記事の質を確保するためにより長い編集時間を要している。私はこの点に問題意識を感じている。

2. 目的・背景

業務を通じて獲得した経験や勘を継承するためには on the job training(OJT)が有効である。しかしながら、OJTによる技能継承は時間がかかる。本研究では OJTにかかる時間を減らすことを目的として、OJTを補完するシステムを実現する。

技能保有者と学習者の2者が齟齬なくコミュニケーションを取れば、OJTにかかる時間を短縮できると考える。本研究では技能保有者のインタビュー記録を対象として、文書中の単語がもつ意味を技能保有者の視点を学習者に理解させるための可視化システムを提案する。

一つの単語を見たときに技能保有者と学習者で違った意味、複数の意味を捉える可能性がある。本ツールは、文章可視化結果を見ながら議論することで互いに相手の個々の単語の捉え方を理解す

るための支援ツールと位置づけられる。

3. 提案手法

本研究では文章集合を以下の手順で可視化する。

① 文章集合の各文章を性質別に分類する。

技能保有者のインタビュー記録は複数の話題に及ぶ。インタビューを受けた技能保有者はインタビュー記録を読み、インタビュー中の話題を列挙する。その後、インタビュー記録を行ごとに区切り、その行がどの話題に該当しているのかを分類する。分類は2種類または3種類を想定している。

② 与えられた文章集合に出現する全単語について、他の単語との類似度を計算する。

類似度の計算として Polynomial Semantic Indexing(PSI)[1]を用いた。PSI は教師あり学習であり、単語×単語行列を低ランクで近似できる点が特徴である。本研究では教師あり学習である点を活かしたく採用した。

③ 2単語間の類似度をもとに、各単語を2次元平面に配置する。

Kamada-Kawai 法[2]を用いて、各単語の x, y 座標を決定した。本研究では可視化平面を 800×800 の格子としている。そのために x, y は小数点以下を切り捨てて整数値とし、格子と対応させた。

④ 各単語について①で分類した性質との類似度を計算する。

PSI を用いて各単語と文書の類似度が計算できる。その類似度は実数値として表されるため、各性質の全文章と単語の類似度を計算した総和によって、各単語と各性質の類似度が計算できる。

⑤ ①で分類した性質に色を定義し、各性質との類似度を用いて各単語に色付けをする。

①で分類した性質が2つであった場合は各性質と2色を対応させる。また、①で分類した性質が3つであった場合は各性質と光の3原色を対応させる。単語と各性質の類似度に応じて重みをつけることで単語の色が決定する。

⑥ 単語が存在しない格子にも周囲から色を補完し、色付けを行う。

単語が存在しない座標の色は、最も近くにある単語を3つ選ぶ。その色を各単語までの距離に応じて重み付けを行い、色付けをする。

本手法可視化は大規模タグクラウドを地形図として可視化した研究[3]や KeyGraph[4]にポテンシャル図の概念を持ち込んだ研究[5]に似ているが、文章を事前に人手によって分類している点で異なる。技能保有者の糸を事前情報として与えたいために教師あり学習にこだわっている。解析対象の文章量は小さいため PSI の教師データに限らず、高品質な教師を作成することが可能である。

4. 結果

本手法を用いて文章集合を可視化した一例を示す。図1は美浜原発の事故に言及した文章を本ツールによって可視化した例である。文章は、原子力発電所の設備を担当するメーカーの主張が含まれるものと、原子力発電に反対する人々の主張の2つに分けた。この例では技能継承

の例から離れるが、単語と文章の分類性質との関連という点で同一である。各単語がメーカー側の主張に近ければ赤で色付けし、原子力発電に反対する人々の主張に近ければ緑で色付けをしている。どちらの主張にも同程度近いものは黄色となる。

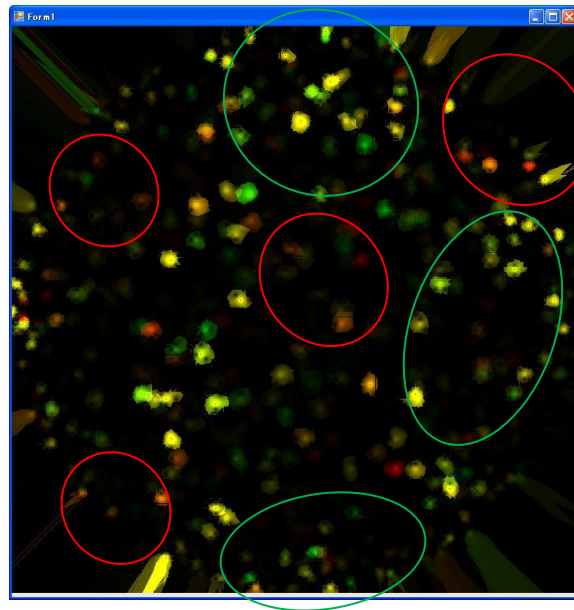


図1：メーカー側の主張（赤）と原子力発電に反対する人々の主張（緑）の勢力図

5. 今後の展望

本研究では対話性の向上が緊急の課題である。今回は平面上の座標をマウスで指定すると近くの単語がわかる仕掛けになっている。今後は例えばユーザーが可視化結果を見て得た感想を書き込むと、その書き込んだテキストを追加したうえで、再び可視化結果が得られるようなフィードバック機構を入れたい。また PSI は大規模なデータを対象とした手法であるため、PSI 以外の方法についても検討を要する。

6. 参考文献

- [1] Bing Bai, Jason Weston, David Grangier, Ronan Collobert, Kunihiro Sadamasu, Yanjun Qi, Corinna Cortes, and Mehryar Mohri. Polynomial Semantic Indexing. In Advances in Neural Information Processing Systems 22, pages 64-72, 2009.
- [2] T. Kamada and S. Kawai, “An Algorithm for Drawing General Undirected Graphs”, Information Processing Letters 31 (1989), pp. 7-15.
- [3] 藤村孝、藤村滋、松林達史、山田武士、奥田英範 大規模タグクラウドを地形図として可視化する Topigraphy 2008 年電子情報通信学会全国大会
- [4] 大澤幸生 チャンス発見のデータ分析 東京電機大学出版局(2006)
- [5] 鶴岡修平 「ポテンシャルモデルを用いたチャンス発見支援のためのデータ可視化」卒業論文 (2005)