

科学研究費助成事業 研究成果報告書



令和 3 年 5 月 10 日現在

機関番号： 2 4 4 0 3
研究種目： 奨励研究
研究期間： 2020 ~ 2020
課題番号： 2 0 H 0 1 1 6 3
研究課題名 学術論文の意味的解析を活用した研究マネジメント支援システムの開発

研究代表者

西田 泰士 (Nishida, Yasushi)

大阪府立大学・研究推進課・リサーチ・アドミニストレーター

交付決定額（研究期間全体）：（直接経費） 480,000 円

研究成果の概要： 本研究では、共クラスタリングとマルチビュー分析を適用することにより、技術的単語間の意味的な関連性を見出すことを目指した。

文書データの分析の際には、各文書での基底単語の出現頻度を用いるBag-of-Words(BoW)の概念がしばしば用いられるが、多次元のBoWベクトルを構築する際には、非常に疎（スパース）な特徴ベクトルとなることも多く、文書間の類似性がうまく把握できない。そこで本研究では、単語間の類似性を考慮したファジィ重み写像を導入するファジィBoW行列の活用を検討した。その結果、従来の手法では見出されなかった知見の発見を実現した。

研究成果の学術的意義や社会的意義

大学等における研究マネジメント人材(URA＝リサーチ・アドミニストレーター)は、論文データ等を用いて研究力を分析し、研究力強化戦略の立案に寄与することが求められている。

本研究では、採択された科研費のキーワードデータを活用した研究マネジメント支援システムを構築した。当該システムは、URAが技術的文章を入力することで、意味的な関連性から研究者をシステムティックに見つけ出すことを実現するものであり、研究マネジメントの高度化が期待できる。

研究分野： ソフトコンピューティング

キーワード： URA リサーチ・アドミニストレーター 推薦システム 共クラスタリング マルチビュー分析

1. 研究の目的

(1) 学術論文や特許文献データに対し、研究相互間の意味的関連性を見出すことを目的とする。共クラスタリングやマルチビュー分析を用いることで、技術的単語間の意味的な関連性を見出し、特にベクトルのスパース性による影響を軽減させる手法を研究する。

(2)(1)による成果をもとに、大学等における研究マネジメント人材(URA＝リサーチ・アドミニストレータ)が利用できるシステムを構築する。

2. 研究成果

(1) 研究相互間の意味的関連性を見出すことを目的に、技術分野に特化した技術要素に着目した知識発見を行うべく、共クラスタリングにより技術分野の分割を行った後に、クラスターごとに SOM による可視化を行った。未活用の潜在的な技術要素対を見出すためには、多様な文献データを分析対象に含めることが望ましいが、技術分野が広すぎる場合には、共クラスタ分割により技術分野に特化した知見を得ることができることが分かった。

また、分析対象の文献データに加えて、異なる観点で収集されたデータを融合的に活用するマルチビュー分析を試みた。特許文献のデータでは、F-term を用いて、クラスター分析に基づく技術分野の分割において、“特許文書×単語”の共起関係に加えて、第2の観点として“特許文書×F-term”を加味したマルチビュークラスタリングを実装した。さらには、ファジィ BoW モデルの構築に第2の観点の F-term による技術分野の類似性の考慮を検討した。“単語×F-term”の関係性データ行列を算出し、単語間の類似性を F-term による技術分野の観点から判定することにより、技術分類の視点に基づくファジィ写像を考慮したファジィ BoW モデルを構築した。

次に、文書×単語のスパースな Bag-of-Words(BoW)行列の拡充法として、単語間の類似性を考慮したファジィ重み写像を導入したファジィ BoW 行列の活用を検討した。通常の BoW ベクトルでは、ある単語が使用されたか否かのハードな写像により頻度を計数していたが、ファジィ拡張であるファジィ BoW モデルではファジィな写像を考え、ある単語が使用された際に、意味がよく似た単語についても類似度に基づくファジィ重みを加算し、スパース性の改善を試みた。

概念的類似性に基づくファジィなマッチングを用いることで、従来の BoW 表現に比しておののの文章はより多くの基底単語と関連付けられ、BoW ベクトルのスパース性を改善できることが分かった。加えて、共通する単語の共起がない文章の間でも、概念的な類似性がある場合には、互いの BoW ベクトル間に相関を表出できることから、埋もれた知識の発見につながると期待できる。

個々の文献データは多様な単語群のごく一部しか使用しておらず、文書間の関連が見出しづらい場面も多いが、ファジィ BoW 行列に拡充することで、スパース性を軽減し、元の BoW 行列では見出されなかった知見を発見できることを確認した(図1)。

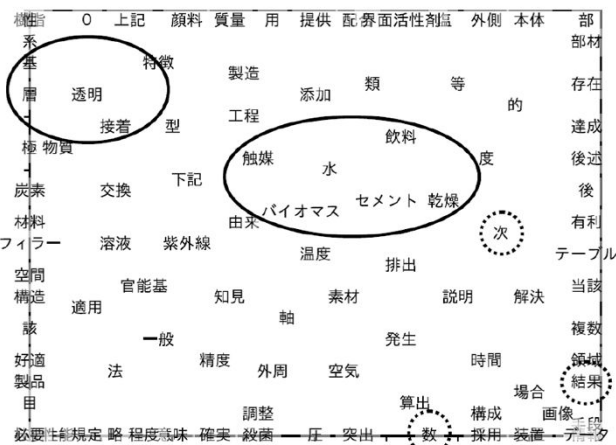


図 1: マルチビュー分析を適用し生成した SOM

(2) 大学等における研究マネジメント人材(URA=リサーチ・アドミニストレーター)は、論文データ等を用いて研究力を分析し、研究力強化戦略の立案に寄与することが求められている。そのため、採択された科研費のキーワードデータを活用した研究マネジメント支援システムを構築した。

文献データに共起出現する単語同士は、例えば構成要素とその動作を示す単語のような、技術分野の観点の下で相互に関連性のある対と考えられるため、共起出現を把握することは既存の技術解決手段における直接的な関連性を捉えるうえで重要である。そこで、まずは単語の出現頻度および共起頻度を反映した特徴ベクトルを生成した。文章中の単語の共起頻度に着目し共起確率を求め、各単語全てに対して算出することで、各単語との共起状態を包含させた共起確率ベクトルとした。これにより、同じような共起の特徴がある単語同士は近いベクトルとなることが期待できる。

意味的関連性への対応としては、word2vec による埋め込みベクトルを作成して技術的単語間の類似性を測るアプローチを導入した。これにより、類義語等の意味的な解釈を取り入れる

ことができた。

当該システムは、研究マネジメント人材が、技術的文章を入力することで、意味的な関連性から研究者をシステムティックに見つけ出すことを実現するものであり、研究マネジメントの高度化が期待できる。

主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1．発表者名 Y. Nishida, K. Honda
2．発表標題 SOM-Based Visualization of Potential Technical Solutions with Fuzzy Bag-of-Words Utilizing Multi-view Information
3．学会等名 Proc. of 8th International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (国際学会)
4．発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

研究組織（研究協力者）

氏名	ローマ字氏名
----	--------