

時系列データの探索的分析を支援する可視化システム: 記事と時系列データのアラインメント方式の提案

Visualization System for Exploratory Analysis of Time-series Data: Alignment Method between Articles and Time-series Data

内藤 峻^{1*} 古田 遼樹² 松下 光範³
Shun Naito¹ Ryoki Furuta² Mitsunori Matsushita³

¹ 関西大学大学院 総合情報学研究科

¹ Graduate School of Informatics, Kansai University

² 数研出版株式会社

² Suken Shuppan

³ 関西大学 総合情報学部

³ Faculty of Informatics, Kansai University

Abstract: The goal of our study is to support a user's analysis of time-series data in an exploratory manner. Such exploratory analysis requires repeated access to various types of information related to the user's interests such as texts and numerical data. To support such the user's analysis, we have proposed a system that visualizes temporal changes in time-series data and presents the causes of those changes with the data. In this paper, we improve the system by adding an alignment function between news articles and time-series data. By using this function, the user can find articles that relates to the time-series data easily.

1 はじめに

時系列データとは、熊本地震の被害者数や気温など時間の経過に伴って変化するデータである。このような時系列データは意思決定や問題解決の場面で役立てられている [1]。意思決定や問題解決の場面では、時系列データの値の変化やその変化の要因を分析することで、有益な情報や新たな知見を得ることが重要である。しかし、このような時系列データの分析は仮説の生成や検証を探索的に繰り返す負荷の高い作業であるため、ユーザがこのような探索行為を円滑に行うことが難しいという問題がある。そこで本研究では、ユーザの興味や関心に応じて様々なモダリティの情報へのアクセスを繰り返しつつ時系列データを分析するための支援システムの実現を目指している。その端緒として、著者らはこれまでに、新聞記事と地図、統計データを対象に、ユーザが時系列データの経時的変化とその変化の要因を把握できるようにする可視化インタフェースを提案してきた [2][3]。本稿では、そのインタフェースに組み込む機能の1つとして、新聞記事と時系列デー

タのアラインメント方式を提案する。この方式をシステムに組み込むことにより、効率的に時系列データと文章を対応付けることができる。さらに、新たに実装したインタフェースの機能について述べる。

2 システムの全体像とこれまでの取り組み

図1に本システムの目指す構成を示す。現状のシステムは新聞記事DB、統計DB、地図DBを人手で作成している。人手での作成は、効率性や網羅性、リアルタイム性の点で問題がある。そのため、システムに用いるデータはWEBから抽出することを考えている。新聞記事DBには、クローラを用いてあるトピックに関する記事を収集し、スクレイピング技術を用いて本文や見出しを抽出する。また、抽出された見出しや本文から自然言語処理技術を用いて日付や国名、出来事に関する文を抽出することを考えている。統計DBや地図DBは、オープンデータとの連携を検討している。

*連絡先： 関西大学大学院総合情報学研究科知識情報学専攻
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

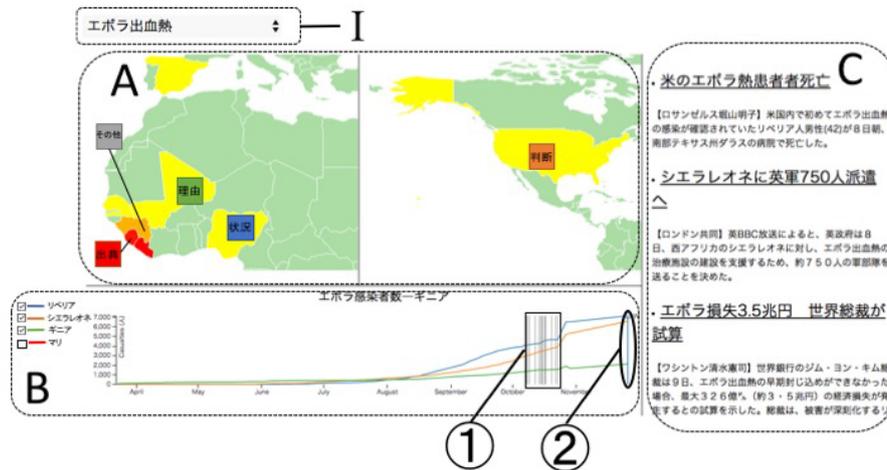


図 2: システムの全体像

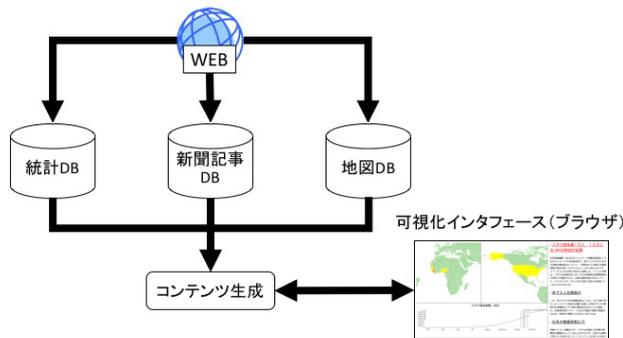


図 1: システムの構成

2.1 システムの全体像

著者らは、ユーザが効率的に時系列データを分析することができる可視化システムについて研究している。図2に理想のシステムの全体像を示す。システムは、トピックを選択する選択ボックス(図2-I)、地図を表示する地図ペイン(図2-A)とグラフを表示するグラフペイン(図2-B)、記事を表示する記事ペイン(図2-C)で構成されている。選択ボックスには、「エボラ出血熱」や「台風」といったトピックがプルダウン形式で表示される。ユーザがトピックを選択すると、そのトピックに関連する統計量がグラフとしてグラフペインに描画され、同時に地図ペインには関連する地図を、記事ペインには関連する記事が表示される。また、地図(図2-A)に記事の有無を表すアノテーションとしてアイコンが付与されている。地図上のアイコンは文献[4]を参考に、5種類(理由、背景、状況、出典、その他)を考えている。ユーザがアイコンをクリックすると、その

アイコンに対応する記事が表示される。さらに、グラフペイン(図2-B)には、複数の統計グラフと凡例が提示されるようになっている。ユーザは凡例の左にあるチェックボックスに比較したい国を選択することでグラフペインに複数の統計グラフを描画することができる。これにより、ユーザは興味を持った国同士の統計量の変化を比較することで他国からの影響や規模を詳しく知ることができる。

現状のシステムには、トピックを選択する機能や地図上のアイコン、複数の統計グラフを描画する機能は実装されていないが、将来的にはこれらの機能を全て実装するつもりである。

2.2 これまでの取り組み

図3に、本研究で対象とするデータの関連性を示す。システムの機能は、図3に示される情報アクセス行為を行えるように実装された。グラフペインには、ユーザによって選択された日付を表示する青い線(図2-①)と記事の有無を表すアノテーション(図2-②)が表示されている。ユーザは青い線を左右にドラッグして日付を選択することができる。日付を選択すると、その時点の統計量が地図にマッピングされる。これによって、ユーザはグラフを見て興味を持った時点の統計量の変化が地理的な影響を受けているのか、周辺の国に影響を与えているのか、把握することができる(図3-①)。また、青い線をアノテーションに重ねると、その時点で起こった出来事について記述された記事がハイライトされる。これによって、ユーザはグラフを見て興味を持った時点の統計量の変化の理由や背景を知ることができる(図3-②)。さらに、地図にマッピングされた国はクリックすることができる。国をクリックすると、

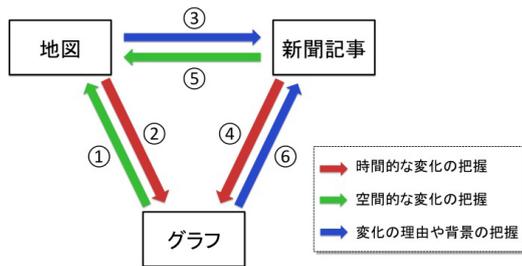


図 3: 対象とするデータの関係性

その国の統計量がグラフとしてグラフペインに描画される。これによって、ユーザは興味を持った国の統計量の変化を詳しく見ることができる(図3-②)。加えて、記事をクリックすると、記事に含まれる日付がグラフ上に表示される。これによって、ユーザは記事を見て興味を持った出来事がグラフのどの時点で起こったのか把握することができる(図3-④)。

残りの新聞記事から地図へのアクセスと地図から新聞記事へのアクセスを行うユーザの振る舞いと機能(図3-③と図3-⑤)については、3節で述べる。

現状のシステムでは、記事の有無を表すアノテーション(図2-②)は、人手で記事の本文を見て日付を抽出し、グラフの日付と対応付けられている。人手での対応付けは効率性や網羅性、リアルタイム性に欠けるという問題がある。4節では、この問題を解決するために、記事の有無を表すアノテーションの付与の自動化について検討した方式について報告する。

3 新聞記事と地図との間のインタラクション

3.1 デザイン指針

本節では、先行研究で検討した新聞記事と統計データ、地図の特徴とデザイン指針について述べる[3]。

新聞記事は、ある時期における出来事やその出来事が起こった原因、場所、統計量、その統計量の具体的な値への言及や予測、記者の意見などが書かれている。そのため、出来事が起こった理由や背景を理解する上で有用である。しかし、新聞記事に書かれている統計量の値は近似値が用いられているため正確でなかったり、記者の観点で纏められていたりするため、客観性に欠ける。

統計データは、ある観測された場所において、ある時点の事象について測定された値である。例えば、人口統計や外国為替相場の推移などがこれに当たる。これらの統計データの値は、政府や国際連合の専門機関などが実施している厳密な環境において観測されてい

たり、センサを用いて取得される。そのため、これらのデータは正確である。

地図は地球や地表、架空の世界の全部もしくは一部を平面上に縮尺表現したものである。例えば、地球全体もしくは大部分を表現している世界地図や統計データを地図上に表した統計地図などがこれに当たる。地図は空間的な位置関係や方向、距離、面積、形、高さを知る上で有用である。また、ある時点の統計量を地図にマッピングすることで、出来事の規模や地理的な広がりを把握できるといった特徴を持っている。

これらの情報はそれぞれ単体でも用いることができるが、ユーザの興味となる要素をトリガとしてインタラクティブに情報を提示することで、円滑な情報アクセスが可能なインタフェースを実現できると考えている。

図3で示したデータの関係性から、ユーザは、統計地図を見て出来事の地理的な影響や規模を把握しつつ、新たに統計値がマッピングされた国や統計量の特徴的な変化に興味を持ち、その理由や背景を知るために新聞記事を参照する(図3-③)。また、新聞記事で言及されている出来事に興味を持ち、その出来事の地理的な影響や規模を把握するために統計地図を参照する(図3-⑤)。本研究では、このような情報アクセスを行えるように、システムの機能をデザインした。これによって、ユーザの探索行為を円滑にする。

3.2 実装

本システムでは、対象データとして2014年の西アフリカエボラ出血熱流行に関する統計データ¹とそれに言及している新聞記事データを用いた²。統計データは、2014年3月22日から2014年11月26日までの日付と累計患者数、各国の感染者数をcsv形式のデータとして纏めたものである。新聞記事データには、2014年10月9日から2014年10月30日までの毎日新聞の記事(計18記事)を用いた。次に、これらの記事からエボラ出血熱について書かれた記事を選び、その記事から出来事に関する文のみを抜き出した。さらに、その文から出来事が起こった日付と場所を示す国名を抜き出し、抜き出した出来事に関する文章の統計量名、を付与したものをcsv形式のデータとして用意した。日付は記事に含まれている「数字+日」を抽出し、年と月を加え「年/月/日」とデータを正規化した。出来事に関する文は、「米国内で初めてエボラ出血熱の感染が確認されていたリベリア人男性(42)が8日朝、南部テキサス州ダラスの病院で死亡した」といった文の一段落分を抽出した。国名は、本文中の「米国」や「リベ

¹<http://ja.wikipedia.org/wiki/2014年の西アフリカエボラ出血熱流行>

²現在は、統計量名の種類としてエボラ感染者数のみしか対応していない

リア」等の名詞を抽出し、「米国」や「米」、「アメリカ」など同じ国のことを指し示している表現は「アメリカ」といったように1つの単語に統一した。統計量名は、本文の「死亡した」や「死者」といった語句から「エボラ死者数」、「感染が確認された」や「感染者数」といった語句から「エボラ感染者数」を人手で判断し、統計量の名称を付与した。また、統計量名がない場合は「非統計情報」とした。記事ページの各記事のスニペットには、これらの情報がタグ付けられている。

3.1節で述べたデザイン指針に基づき実装した機能について述べる。ユーザが国をクリックすると、その国名が含まれる記事がハイライトされる機能を実装した。システムは、国をクリックされたことを判断すると、選択された国名の記事モジュールへと引き渡す。記事モジュールは引き渡された国がタグ付けられている記事をハイライトする。これにより、ユーザは地図を見て興味を持った国に関する統計量の変化の理由や背景を把握することができる。

記事をクリックすると、その記事に含まれる国がハイライトされる。システムは、記事がクリックされたことを判断すると、データがタグ付けられた記事のスニペットから記事中に含まれる国を地図モジュールへと引き渡す。地図モジュールは引き渡された国をハイライトする。これにより、ユーザは興味を持った記事の出来事の影響や規模を把握することができる。

4 新聞記事と時系列データのアラインメント方式

4.1 対象とする課題

現状のシステムでは、時系列データと文書との対応付けは人手に委ねられており、効率性や網羅性の点で問題がある。この問題を解決するため、本研究では新聞記事やブログなど時系列データの変化の理由や背景が記述された文書が、時系列データのどの期間に対応しているのかを自動的に推定し、それらを紐付けて提示する手法の実現を目指す。

時系列データと文書を対応付けるための研究は多数行われている。例えば、小林らはグラフで示された数値情報を自然言語テキストで説明する手法を提案している [5]。提案手法は選択体系機能言語理論を用いてグラフの特徴と言語表現の関係を分析し、それに基づきテキストを生成している。また、日経平均株価のグラフとその動向を説明するテキストを用いてグラフとテキストが協調的に提示される手法を提案している [6]。この手法では、長期的な動向には、グラフの表示状態に合わせてテキストを要約し、提示している。短期的

な動向には、人間が視覚的に捉えるグラフの挙動を説明するのに適切なテキストが生成される。

また、AhmadらやBoydは時系列データにWavelet解析を行い、グラフの特徴を特定し、自然言語テキストを生成する手法 [7][8] を、馬野らは全体的傾向と局所的特徴を組み合わせて時系列データ全体を言葉で表現する手法を各々提案している [9]。

これらの手法を用いることで、時系列データと記事のアラインメント行うことも考えられるが、日付や変動の曖昧な表現についてはほとんど考慮されていない。

本研究は、自然言語表現の中でも期間と変動の曖昧な表現に着目し、グラフの各始点と終点との一致度を算出することでアラインメントを試みる点が先行研究と異なる。

時系列データと文書を組み合わせるには、文書に含まれる時系列データに言及した文章から、該当する時系列データの箇所を特定する必要がある。例えば「中国経済の先行き不透明感から日経平均株価は1月7日に18000円を割り込んだ」という文章の場合、日付に関する表現 (i.e., 「1月7日」) と値に関する表現 (i.e., 「18000円」) から対応する時系列データ (i.e., 日経平均株価データ) の該当する箇所を特定し、時系列データと文書の紐付け処理 (以下、アラインメントと記す) を行う。しかし、文書中の表現は必ずしも明確な数値で記述されるわけではなく、「中国経済の先行き不透明感から、日経平均株価は1月初頭に大きく下落した」のように、下線部のような曖昧な表現を用いて期間や値の変動が記述される場合も多い。このような文章からアラインメントを行うには、これらの表現を解釈し、該当する箇所を特定する必要がある。本稿では、このような文章に含まれる曖昧表現として日付に関する曖昧表現 (e.g., 「中旬」「初頭」) と値の変動に関する曖昧表現 (e.g., 「上昇」「下落」) のふたつに着目し、これらから時系列データの該当箇所を特定する方式について検討する。

4.2 提案手法

図4に、グラフで表現した時系列データとそれに言及したふたつの文章を示す。図4中の (α) と (β) の文書では変動に関してどちらも「下落した」という表現が用いられているが、(α) の文章が紐付けられるべきは図4-Aの領域であり、(β) の文章が紐付けられるべきは図4-Bである。このようなアラインメントを行う場合、文書に含まれる変動に関する表現と日付に関する表現の両方を満たす箇所を時系列データから特定することになる。ここで、「下落した」という表現を解釈する場合、(β) のような場合には時系列データの値は必ずしも単調減少するわけではなく、その途中で一旦上

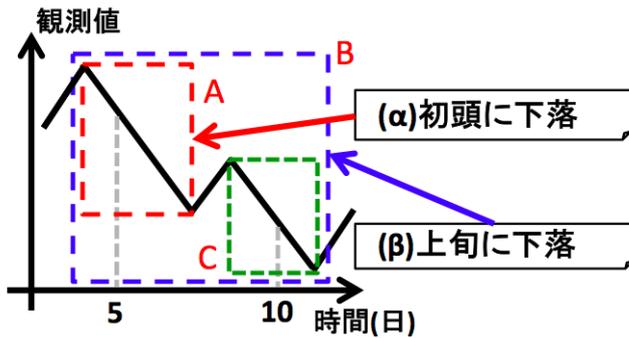


図 4: 時系列データと文書の対応

昇に転じている場合も該当するため、単調減少している箇所のみをアラインメントの候補にするのではなく、(1) 領域内で単調減少している割合が高い、(2) 領域内で始点の値が最も高く終点の値が最も低い、というふたつの条件を満たす領域を選択する必要がある。本研究では、時系列データを極値で分割し、任意の二つの極値の間の時系列データを候補としてアラインメント対象の文章に照らして「変動に関する一致度」と「日付に関する一致度」のふたつを算出し、それらの値の積が最も高くなる期間をアラインメントを行う最尤期間とする。

この方式では、任意のふたつの日付 $d_A, d_B (d_A < d_B)$ で挟まれた期間 $[d_A, d_B]$ の時系列データを対象として、変動に関する一致度 C_f と期間に関する一致度 C_d を求め、それらの積を文書との一致度 C とする。 C_f は $[d_A, d_B]$ で変動に対する曖昧表現を満たす区間の割合である。日付 d の時系列データの値を $f(d)$ とすると、例えば「下落した」の場合は、 $[d_A, d_B]$ において $f(d_A)$ が最大かつ $f(d_B)$ が最小という条件の下で、 $f(d) < f(d+1)$ (但し $d_A \leq d < d_B$) となる部分区間の割合を C_f とする。また、期間に関する一致度 C_d は、日付に関する曖昧表現を日付の始点と終点を表すファジィ集合を各々 $M_s(d), M_e(d)$ として (図5参照)、 $C_d = M_s(d_A) \times M_e(d_B)$ で算出する。これらから、 $C = C_f \times C_d$ を算出し、それが最も高い値となる $[d_A, d_B]$ を最尤区間とする。

4.3 評価と考察

提案手法を用いたシミュレーションとして、時系列データとしてマネースクウェア・ジャパンのHP (<http://www.m2j.co.jp/market/historical.php>) で公開されている米ドル/円の為替レートのうち 2014年 10月 1日から 11月 28日までのデータを、それに言及する文章として WEB 上から取得した (A) 「11月上旬は上昇した」 (B) 「10月末は下落した」の2つの文章を

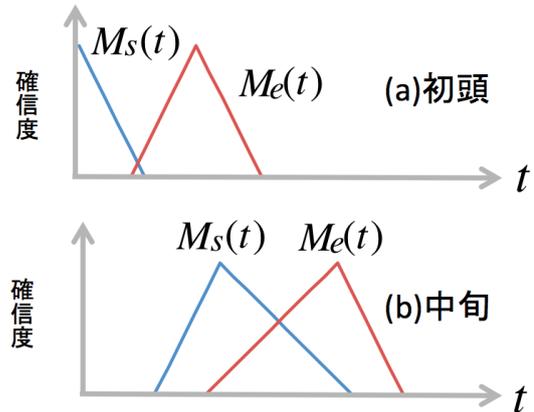


図 5: 期間のファジィ集合

表 1: 文章 (A) に対するアラインメント候補

| 順位 | 期間 | 一致度 |
|----|--------------|------|
| 1位 | 11/1 - 11/11 | 0.80 |
| 2位 | 11/2 - 11/11 | 0.70 |
| 3位 | 11/4 - 11/11 | 0.60 |
| 4位 | 11/1 - 11/14 | 0.54 |
| 5位 | 11/2 - 11/18 | 0.47 |

用いて、これらのアラインメントを試みた。文章 (A) に対するアラインメント候補上位 5 件の期間及び一致度 C を表1に、各々の文章の最尤区間をグラフで示したものを図6に示す。図6の結果から目視ではあるが、それぞれの文章が各時系列データの該当する箇所に対応付けられていることがわかる。しかし、一致度の算出するにあたり、「末」や「初頭」など短い期間に関する一致度が全体的に低くなるとともに、候補数が極端に少なる傾向が見受けられた。これは、ファジィ集合に設定する日付を変更すると確信度が大きく変化するからである。今後は、より多様な時系列データと文章を対象としてアラインメントを行い、提案手法の精緻化を目指す。加えて、システムによって提示されたアラインメント候補の妥当性を検証するために被験者実験を行う必要があると考えている。これにより、人間が言語を介して行っている知的情報処理により近いアラインメント候補を求めることができる。

5 おわりに

本稿では、時系列データの探索的分析を支援するシステムの実現に向けて時系列データと新聞記事をアラインメントする方式を提案した。この方式をシステムに組み込むことにより、効率的に時系列データと文章を対応付けることができる。また、これまでの取り組

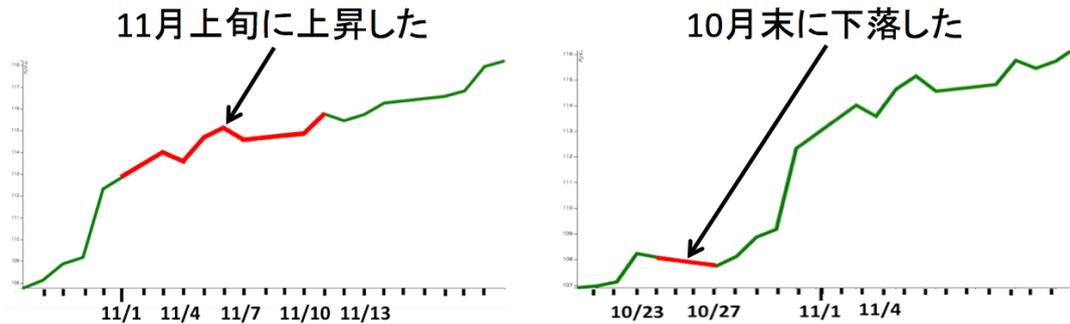


図 6: 推定された最尤アラインメント期間

みとこれから目指すシステムの全体像について述べた。さらに、インタフェースに地図と記事に相互にアクセスできる機能を実装した。これにより、興味を持った国に関する統計量の変化の理由や背景を把握したり、記事の出来事の影響や規模を把握したりできる。現在のプロトタイプシステムは、「エボラ出血熱」というトピックのみしか対応していないが、同じ要素を持つ「熊本地震」や「台風の被害」などのトピックにも応用することができるため、他のトピックも分析できるようにするつもりである。今後の展望としては、システムに必要な新聞記事のデータを Web から自動で収集する仕組みを検討している。具体的には、クローラとスクレイピング技術を用いてニュースサイトから記事の本文を抽出することを考えている。これにより、熊本地震のようなトピックを分析する人がリアルタイムな情報を見て分析を行うことができたり、システムの自動化に繋がったりする。さらに、グラフペインに複数の統計グラフを描画する機能を実装しようと考えている。時系列データの分析をする場面では、1つのデータを見るだけでなく複数のデータを比較し、相関や違いを見ることが重要である。この機能により、ユーザは複数の統計データを比較することが可能になる。

謝辞

本研究の遂行にあたり、文部科学省科学研究費(課題番号:15H02780)の助成を受けた。記して謝意を表す。

参考文献

- [1] 藤本和則, 木村 陽一, 松下 光範, 庄司 裕子: 意思決定支援とネットビジネス, オーム社 (2005)
- [2] Naito, S., Matsushita, M.: Supporting Consecutive Data Exploration by Visualizing Spatio-temporal Trend Information, in *Proceedings of*

the 2015 Conference on Technologies and Applications of Artificial Intelligence, pp. 227–231 (2015)

- [3] 内藤峻, 松下光範: 時空間動向情報を対象とした探索的データ分析のための可視化インタフェースの提案, ARG 第6回 Web インテリジェンスとインタラクション研究会, No.6, pp. 31–36 (2015)
- [4] 松下光範, 加藤恒昭: 言語情報と数値情報の相補的利用を目指した可視化手法, 第21回人工知能学会全国大会, 3H8-3 (2007)
- [5] 小林一郎: グラフ情報の自然言語処理に関する研究 日本ファジィ学会誌, Vol.12, No.3, pp.406–416 (2000)
- [6] 小林 一郎, 渡邊 千明, 奥村 奈穂子: グラフとテキストの協調による知的な情報提示手法: 日経平均株価テキストとグラフの提示を例にして, 情報処理学会論文誌, Vol. 48, No. 3, pp.1058–1070 (2007)
- [7] Saif Ahmad, Paulo C F de Oliveira, Khurshid Ahmad: Summarization of Multimodal Information, *Proc. 4th International conference on Language Resources and Evaluation*, pp.1049–1052 (2004)
- [8] Sarah Boyd: TREND: A System for Generating Intelligent Descriptions of Time-Series Data, *Proc. IEEE International Conference on Intelligent Processing Systems* (1998)
- [9] 馬野 元秀, 小泉 尚之, 篠原 貴之, 瀬田 和久: 全体的傾向と局所の特徴に基づく時系列データの言葉による表現, 第22回ファジィ システム シンポジウム 講演論文集, pp.343–346 (2006)

レビュー閲覧履歴からの価値観モデリングを用いた情報推薦システム

Recommender System Using Personal Values-based User Modeling from Browsing History of Customer Reviews

高間 康史^{1*} 清水 涼人¹
Yasufumi Takama¹ Suzuto Shimizu¹

¹ 首都大学東京システムデザイン研究科

¹ Graduate School of System Design, Tokyo Metropolitan University

Abstract: This paper proposes a method for generating a user model reflecting user's personal values from user's browsing histories of customer reviews. This paper also proposes a recommendation method using the personal values-based user model. Existing recommendation methods such as collaborative filtering and content-based filtering tend to be less accurate for new users and items due to the lack of information about them. The personal values-based recommender system is expected to realize more precise recommendations for new users. As a customer review contains reviewer's evaluation of an item and its attributes, the proposed method estimates attributes on which a target user put high priority when evaluating items from customer reviews the user refers to for his/her decision making. This paper examines the effectiveness of the proposed method with user experiments.

1 はじめに

本稿では、レビューを評価した履歴に基づきユーザの価値観をモデル化する手法を提案し、情報推薦に適用してその有効性を示す。情報推薦は、ユーザが好む情報やアイテムなどをシステムから提示するための技術であり、膨大な情報にアクセス可能な現代社会にとって必要不可欠な基盤技術となっている。Web 上には多様なオンラインショッピングサイトやレビューサイトが存在し、満足度ランキングやベストセラーなどといった非個人化推薦も含めれば、多種多様な情報推薦技術がすでに実用化されている。しかし、個人に適応した推薦を行うためには対象ユーザに関する多くの情報が必要であり、新規に利用を開始したユーザに対して適切な推薦を行えないという問題が指摘されている [11]。また、十分な推薦精度を得るためには、一般に多くの情報を必要とする。これらの問題を解決し、情報推薦システムの適用範囲を広げるために、より少ない情報から推薦を行う必要性が高まっている [5]。

より少数の情報から安定したユーザモデルを獲得するために、本稿では価値観に着目する。価値観は個性や嗜好などとともに、意思決定に影響を与える要因の

一つであり、情報推薦に応用する手法が近年研究されている [2, 7]。価値観とは物事を評価する際に個人あるいは社会などが採用する優先順序や重みづけの体系であり、こだわりとみなすこともできる。従って、ユーザの価値観を適切にモデル化することができれば、単なる人気アイテムではなく、ロングテールでもユーザが好むアイテムの推薦など、情報推薦が本来必要とされる場合に有効性が期待できる。また、個人の価値観が意思決定に安定して反映されるのであれば、少数の情報からユーザモデルを構築できる可能性もある。この点に関して、服部らが提案している評価一致率を用いた価値観モデル [2] では、少ない嗜好情報やインタラクションからモデル構築が可能であることが報告されており、本稿でもこれを採用する。

ユーザモデル構築に用いる情報として、先行研究 [3] ではユーザが投稿したレビューを用いている。しかしこの手法では、レビューを投稿していない新規ユーザに対しては価値観モデルを作成できないため、cold-start問題には対応できない。また、そのアイテムに対する総合評価と各属性に対する評価をユーザから明示的に取得し、価値観モデルを構築するアプローチも存在するが [2]、各属性に対して評価を行うことはユーザにとって負荷の高い作業であり、モデルの作成のコストが増大する。

*連絡先：首都大学東京システムデザイン研究科
〒191-0065 東京都日野市旭が丘 6-6
E-mail: ytakama@tmu.ac.jp

そこで、本稿ではレビューをユーザが閲覧するという行為から、ユーザの価値観をモデル化する手法を提案する。ユーザはレビュー文章や属性別評価などの情報を閲覧し、そのレビューおよび言及されているアイテムに対し評価や選択などを行う。これらの情報から、ユーザの評価とレビュー内で示されている評価を好評・不評に分類し、評価一致率を用いてユーザのアイテム選択における価値観をモデル化する。提案手法では、レビューを書いている新規ユーザに対しても価値観モデルを作成可能であり、またレビュー閲覧はオンラインレビューサイトなどで一般的な行為であるため、コスト面での効率化も期待できる。先行研究 [12] では、ユーザに提示するレビューを事前に決定していたが、提案手法では、ユーザの評価情報から次に提示するレビューを動的に決定することで、システム構築コストの削減とユーザ価値観の素早い推定を目指す。レビューの提示方法に関して、動的提示とランダム提示の2種類でユーザモデルを作成し、アンケート回答との比較や評価件数の観点などから比較を行う。

アイテム推薦に関する評価実験では、提案手法を用いて作成したユーザモデルに基づく推薦、ランダム提示を用いて作成したユーザモデルに基づく推薦、満足度ランキングに基づく推薦の3種類を実施し、その結果を比較して提案手法の有効性を示す。

2 関連研究

2.1 情報推薦

現在では多くの Web サイトで情報推薦技術が使用されている。例を挙げると、Amazon や楽天ショッピングサイトなどでは TOP ページにユーザに対するお勧めアイテムを掲載しており、情報推薦の使用頻度は多くなってきている。

情報推薦の代表的な手法として、協調フィルタリング [9] と内容ベースフィルタリング [8] が挙げられる。協調フィルタリングは、同一サービスを利用する多数ユーザの行動履歴に基づいてユーザ間の類似度を計算する。この計算結果を利用して、「類似ユーザのアイテムに対する嗜好は類似する」という仮定に基づき推薦対象ユーザが未評価のアイテムに対する評価を予測する。

内容ベースフィルタリングは、アイテムのコンテンツに関する情報を利用した推薦手法である。コンテンツに関する情報としては、映画についてはジャンルや監督、主演など、画像などに対しては付与されたタグ、文書であれば含まれる単語などが利用される。ユーザが好む情報を推定し、これをコンテンツとして持つアイテムを推薦する。協調フィルタリングではコンテンツに関する情報が不要という利点がある一方、新規ユーザ

に対する推薦が困難という cold-start [11] 問題や、評価情報の不足による sparsity 問題などが指摘されている。内容ベースフィルタリングは新規ユーザであっても嗜好情報が取得できれば推薦可能である一方、コンテンツに関する情報が場合によっては取得困難であるといった欠点がある。

2.2 価値観モデリングと情報推薦

通常の情報推薦システムではユーザの嗜好を推定し、利用するのに対し、個性や価値観といった、意思決定に影響を与える他の要因に着目した研究もおこなわれている。心理学などの分野において、価値観を構成する要因である Rokeach Value Survey [10] や、個性を構成する要因である Big Five [1] などが知られており、これらに基づきユーザの価値観や個性を分析するアプローチが有名であるが、情報推薦への応用を考えた場合、対象アイテムの評価により直結した要素に基づきユーザの価値観や個性をモデル化の方が好ましい。このようなアプローチとして、伊藤らは主成分分析を用いて評価傾向が類似するユーザを求め、推薦に利用する手法を提案している [4]。また、アイテムの評価に関連する属性に基づき価値観や個性を捉えるアプローチも存在する。Wu らは重要な属性について多様なアイテムを推薦する手法を提案しており、多様性の度合いは、ユーザの個性と多様性に対する要求との関係に基づいて決定している [13]。

服部らは、アイテムの持つ属性が、ユーザの意思決定において重視される度合いを表す指標として評価一致率 (rate matching rate, RMRate) を提案している [2]。この手法では、ユーザのアイテムに対する評価極性 (好評または不評) とアイテムの各属性に対する評価極性を用いて、属性ごとの評価一致率を計算しユーザモデルを作成する。ユーザ u_i のアイテム x_j への評価極性を p_{ij} 、アイテムの属性 a_k への評価極性を p_{ij}^k とすると、 u_i の a_k における評価一致率 RMR_{ik} は式 1 で算出される。 I_i は u_i が評価したアイテム集合、 $\delta(x, y)$ は x と y の値が等しい時 1、異なる時に 0 を返す関数とする。

$$RMR_{ik} = \frac{\sum_{x_j \in I_i} \delta(p_{ij}, p_{ij}^k)}{|I_i|} \quad (1)$$

価値観モデルの情報推薦における活用方法として、内容ベースの推薦手法や協調フィルタリングに適用した手法が提案されている。内容ベースの推薦手法では、外部にある映画データベースを利用した情報推薦システムにおいて、推薦対象ユーザの評価一致率が高い属性について、ユーザが過去に評価した映画の属性値を持つ映画を検索し、推薦アイテムを取得している [2]。

協調フィルタリングに価値観モデルを適用したシステム(価値観モデルベースCF) [6]では、推薦対象ユーザと他のユーザの価値観に基づくユーザモデルの相関係数を計算し、これをユーザ間の類似度として近傍ユーザを求める。通常のユーザベースCFとは、相関係数の求め方が異なるだけであり、予測評価値の計算などは通常のユーザベース協調フィルタリングと同様である。価値観モデルベースCFを用いることで、低い評価値が付与されている場合の予測精度向上などの結果が得られている [6]。また、ユーザモデルの次元数はアイテムの評価に関する属性数と同じであり、これは一般的なレビューサイトでは数種類と少ない。従って、ユーザ間類似度に利用する行列はユーザ・アイテム行列よりも密な行列となるため、類似度を計算可能なユーザ数が増加するという利点もある。

3 レビュー閲覧履歴からの価値観モデリング

3.1 価値観モデリングのためのレビュー提示

評価一致率に基づく価値観モデリングでは、アイテムに対する総合評価だけでなく、各属性に対する評価も必要とする。これらの情報を取得する方法として、前述のとおりユーザに直接尋ねる明示的手法、ユーザが投稿したレビューを利用する暗黙的手法が提案されている。近年、属性毎の評価も付与されているオンラインレビューサイトが多く存在し、その場合には暗黙的手法が利用可能である。また、価値観モデリングでは少数のレビューからでもモデリング可能であることが報告されている [2] が、通常レビュー投稿者よりもレビュー閲覧のみを行うユーザの方が多数である。レビュー閲覧者は、レビュー記事の中で自身がこだわりを持つ属性についての言及や評価を確認し、意思決定に利用していると想定される。従って、レビュー閲覧履歴からモデリングを行うことができれば、適用範囲を拡大する効果が期待できる。

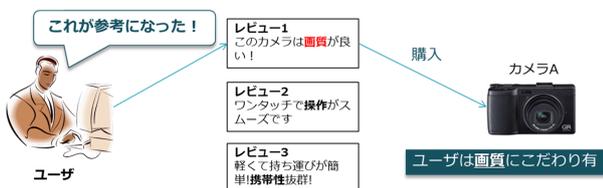


図 1: ユーザによるレビュー評価の例。

提案手法のコンセプトを図 1 に示す。図ではユーザがデジタルカメラの評価情報として、レビュー 1 を参考に購入した場合を想定している。レビュー 1 には「画

質」についての評価が書かれており、ユーザは「画質」にこだわりを持ってデジタルカメラを購入したと推測できる。実際には、レビューテキストを解析して言及されている属性およびその評価極性を抽出するのではなく、レビューに付与された評価値を利用する。従って、総合評価だけでなく属性別評価も利用可能なオンラインサイトを対象とする。

上述の通り、提案手法においてユーザはレビュー記事を読んでアイテムを評価するが、評価一致率の計算はレビューに付与されている属性評価値に基づき行う。従って、価値観モデリングを効率的かつ精度よく行うために、ユーザに提示すべき有用なレビューの条件を以下のように定義する。

1. レビュー内容と属性別評価の極性が一致していること
2. 特定の属性に対して根拠を示し言及していること
3. 全属性が高評価(低評価)に偏っていないこと

(1) に関して、提案手法では属性別評価に基づき評価一致率を計算するため、レビュー内容と属性別評価が対応しないものは適さない。また、(2) に関しては、どの属性にも根拠が言及されていないレビューは属性に関する有用な情報を提供していないため適さない。(3) については、例えば属性別評価が全て 5 のレビューを参考にしてアイテムを評価した場合、どの属性に着目したかが判断できないため、価値観モデリングには適さない。

先行研究 [12] では、デジカメに関するレビュー 3 件を 1 組とし、合計 20 組のレビューセットを作成して、ユーザモデリングおよびその結果を用いたアイテム推薦についてのユーザ実験を行っている。20 名のユーザに対し実験を行った結果、満足度ランキング上位のアイテムと同程度の推薦精度が得られることを示している。

これに対し本稿では、固定的なレビューセットを用いるのではなく、対象ユーザの価値観モデルを逐次更新しながらその値に基づき提示レビューを決定する、動的なレビュー提示手法を提案する。これにより、ユーザの価値観モデルをより効率的に構築可能となる効果だけでなく、事前にレビュー組を作成する必要がないことからシステム構築コストの削減効果も期待できる。

動的レビュー提示によるユーザモデル構築のフローチャートを図 2 に示す。最初に提示するレビューに関しては、ユーザの回答情報を得られていないためランダムに 3 件選択し、1 組として提示する。その際、どの属性も必ず 1 件以上は言及されているレビュー組となるようにする。提示されたレビューに対し、ユーザは評価を行う。具体的には、参考になったレビューの一つ選択するとともに、対象アイテムに対する評価を 5 段階(5:高評価)で回答する。

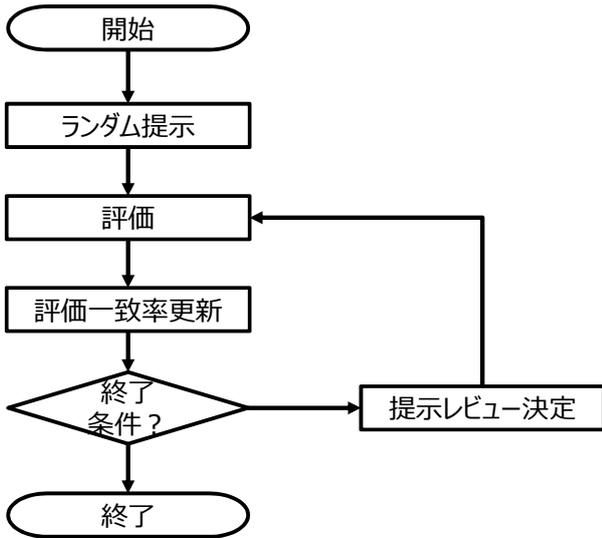


図 2: 動的レビュー提示のフローチャート.

評価が得られるたび、式 1 に基づき各属性の評価一致率を更新する。2 組目以降に提示するレビューは、更新された評価一致率に基づき決定する。提示するレビューの選択に関して、こだわりの強い属性を早期に推論するため、その時点での評価一致率が高い属性に着目し、その属性に関して好不評両方のレビューを同時に提示する。これにより、当該属性に関するユーザーフィードバックを効率的に獲得する効果が期待できる。

具体的には、推定対象となっている属性の中で評価一致率が最も高い属性に着目し、以下の条件を満たす様に提示するレビュー 3 件を決定する。ここで、評価一致率の推定には 7 件程度の投稿レビューが得られれば十分という先行研究 [2] の知見に基づき、ユーザーからの評価が 10 件得られた属性については、その時点で評価一致率を固定し、推定対象から除外する。これは、なるべく多くの属性について効率よく推定を行うためである。

1. 着目した属性について好評、不評のレビューをそれぞれ 1 件以上提示する。
2. (1) を満たす中で、式 2 の値が高いレビュー r を提示する。

$$Score(r, u_i) = \frac{\sum_k \{ |e_r^k - e_r| \} \cdot RMR_{ik}^2}{N_r} \log n_r \quad (2)$$

ここで、対象ユーザーを u_i 、 r の属性 a_k に付与されている評価値を e_r^k 、 r の全属性の平均評価値を e_r 、文字数を n_r 、言及されている属性数を N_r とする。先行研究 [12] の実験結果を分析した結果、文章量の多いレ

ビュー程ユーザーが参考にする傾向がみられたことから、文書長も考慮に入れている。

終了条件に関して、本稿の評価実験では 20 回のレビュー提示を終了条件としたが、推定が完了した属性数などを終了条件とすることも考えられる。

なお、ユーザーに提示したレビューのうち、ユーザーが選択したものは以降の提示対象から除外する。評価しなかったものに関しては、他のレビューとの比較では参考になる可能性があると考え、以降の提示対象とすることにより、収集したレビューの有効活用を図る。

3.2 情報推薦システムへの適用

ユーザーモデルを用いた推薦では、推薦対象ユーザー u_i に対し各アイテム x_j のスコアを式 3 の様に定め、このスコアが高い順に推薦する。 x_j が所属するカテゴリ c における属性 a_k の平均評価値を e_c^k 、アイテム x_j に対する a_k の平均評価値を e_j^k とする。これらは提示用に収集したレビューのみから求めるのではなく、オンラインサイトに登録されている全アイテム・レビューから求めることが望ましい。スコアの計算においては、対象ユーザー u_i の全属性に対する平均評価一致率以上の評価一致率を取る属性だけを使用する。

$$Score(x_j, u_i, c) = \sum_{a_k \in A_i} \{ e_j^k - e_c^k \} \cdot RMR_{ik}^2 \quad (3)$$

$$A_i = \left\{ a_k \mid RMR_{ik} \geq \frac{\sum_{a_l \in A} RMR_{il}}{|A|} \right\} \quad (4)$$

4 評価実験

4.1 実験概要

提案手法の有効性を検証するため、20 人の実験協力者を対象として評価実験を行った。実験は、3.1 節で述べたレビュー提示によるユーザーモデリングフェーズと、3.2 節で述べたアイテム推薦フェーズからなる。

モデリングフェーズでは、提案する動的提示手法と、提示するレビューをランダムに決定する手法（ランダム提示）の 2 種類の手法を用い、その結果を比較する。両提示手法において、異なるホテル 3 件に対するレビュー 1 件ずつを組み合わせて 1 組とし、20 組について評価を行ってもらう。

実験には、ホテルレビューサイト 4travel¹ からレビューとホテルの情報を収集し、データセットを構築して用いた。総レビュー数は 592 件である。評価一致率を求めるためには、属性評価を好評・不評の二値に変換

¹<http://4travel.jp/>

する必要がある。そのため、レビュー毎に属性評価の平均値を求め、平均値以上の属性を好評、平均値未満の属性を不評と判定した。また、4travelに掲載されているホテルの属性別評価の一つに「バリアフリー」があるが、今回の実験協力者においてはバリアフリーに着目してホテルを選択することはないと判断し、除外した。

表 1: 実験計画.

| グループ | 動的提示 | ランダム提示 |
|--------|----------|----------|
| 観光 A | 東京・神奈川 | 北海道 |
| 観光 B | 北海道 | 東京・神奈川 |
| ビジネス A | 大阪・京都 | 東京・愛知・福岡 |
| ビジネス B | 東京・愛知・福岡 | 大阪・京都 |

ホテル利用目的としてビジネスと観光の2種類を想定し、それぞれについて地域の異なるデータセットを2種類ずつ構築した。実験におけるデータセットと提示手法の組み合わせ（実験計画）を表1に示す。実験協力者は5名ずつ、表に示すグループのいずれかに振り分けられ、ユーザモデル構築のためにレビューを閲覧する。この時、データセットや提示手法の順番に偏りが出ないようにラテン方格計画を用いて調整している。

アイテム推薦フェーズでは、動的提示、ランダム提示それぞれにより構築したユーザモデルを用いてホテルの推薦を行う。各ユーザモデルについて10件ずつホテルを推薦する他、レビューサイトにおいて公開されている地域ごとの満足度ランキングを用いて10件の推薦を行う。推薦対象として、モデリングフェーズとは異なる地域のホテルからデータセットを作成した。作成したデータセットと実験条件の組み合わせを表2に示す。総アイテム数は3176件である。なお、価値観モデルはユーザのこだわりを推定することが目的であるため、全ての属性において評価の高いホテルは推薦対象にふさわしくないと考える。そこで、全属性の評価が5段階で4以上のホテルは推薦対象から除外する。

3つの手法それぞれについて10件の推薦アイテムを決定した後、どの手法で推薦されたものかをわからないように統合して実験協力者に提示して、好評・不評のいずれかで評価してもらった。なお、複数の手法で同じアイテムが推薦対象となる場合もあるため、実験協力者に提示するアイテムは30件未満となる場合もある。

4.2 モデリング結果の評価

モデリング結果の一例として、観光 A グループの実験協力者5名について動的提示、ランダム提示で得ら

表 2: アイテム推薦に用いたデータセット.

| グループ | 動的提示 | ランダム提示 |
|--------|----------|----------|
| 観光 A | 大阪・兵庫・京都 | 沖縄 |
| 観光 B | 沖縄 | 大阪・兵庫・京都 |
| ビジネス A | 神奈川 | 兵庫・京都 |
| ビジネス B | 兵庫・京都 | 神奈川 |

れたユーザモデル（評価一致率）を表3, 4にそれぞれ示す。括弧内の数値は、ユーザが選択したレビュー中で言及されていた回数を示す。両表で、同じIDは同一の実験協力者に対応する。

実験後にアンケートを実施し、実験協力者が興味あると回答した属性は表中で太字にしてある。興味はなくても意思決定に影響を与えている属性の存在や、レビューの書き方によっては参考にならない場合などもあるため、アンケート結果が正解とは必ずしも言えないと考えるが、評価一致率が0.7以上の場合にこだわりがある属性と判断した場合、表3ではアンケート結果と推定結果が一致する場合（正解）が5回、表4では2回となっている。また、ユーザが興味あると回答しなかった属性で評価一致率が0.7以上となっている場合を誤検出とみなすと、表3では4件、表4では6件、反対に興味ある属性で評価一致率が0.7未満となる場合を見逃しとみなすと、それぞれ11件、14件あることがわかる。

同様にして、4つのグループについて正解、誤検出、見逃しの回数を集計した結果を表5に示す。これより、ランダム提示の方が動的提示よりもアンケート回答に一致する結果が得られている。また、両提示手法に共通する傾向として、アンケート結果を正解とみなした場合には誤検出よりも見逃しの方が多結果となっている。この理由として、例えば表3, 4では、実験協力者2はアクセス、コスパについて興味ありと回答しているが、両提示手法で評価一致率は0.7未満であり、レビュー中での言及回数も10回未満と少なめになっている。従って、推定に必要な情報が十分取得できなかったことが原因の一つとして考えられる。前述の通り先行研究では、投稿レビューの場合には7件以上で推定可能とされているが、閲覧履歴の場合は本人が記述したレビューではないため、より多くのレビュー閲覧履歴を必要とする可能性があると考えられる。

実験協力者全グループについて、興味があると回答した属性について得られた評価一致率の平均を表6に示す。表より、0.7以上となったのはランダム提示手法のビジネス B グループだけであるが、全てのグループについてランダム提示の方が評価一致率が高くなっている。

表 3: 観光 A・動的提示の実験結果.

| ID | アクセス | コスパ | 接客 | 部屋 | 風呂 | 食事 |
|----|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1 | 0.500 (10) | 0.429 (7) | 0.800 (12) | 0.500 (15) | 0.800 (10) | 0.800 (11) |
| 2 | 0.375 (8) | 0.667 (6) | 0.900 (15) | 0.300 (17) | 0.500 (8) | 0.700 (10) |
| 3 | 0.444 (8) | 0.625 (8) | 1.000 (12) | 0.400 (17) | 0.571 (7) | 0.700 (10) |
| 4 | 0.600 (10) | 0.600 (10) | 0.900 (12) | 0.300 (17) | 0.429 (7) | 0.429 (7) |
| 5 | 0.800 (15) | 0.000 (4) | 0.500 (14) | 0.300 (15) | 0.200 (11) | 0.500 (4) |

表 4: 観光 A・ランダム提示の実験結果.

| ID | アクセス | コスパ | 接客 | 部屋 | 風呂 | 食事 |
|----|----------------------|---------------------|---------------------|----------------------|----------------------|----------------------|
| 1 | 0.333 (6) | 1.000 (2) | 0.700 (10) | 0.273 (11) | 0.636 (11) | 0.600 (15) |
| 2 | 0.571 (7) | 0.500 (6) | 0.455 (11) | 0.642 (14) | 0.583 (12) | 0.385 (13) |
| 3 | 0.500 (2) | 0.750 (8) | 0.556 (9) | 0.643 (14) | 0.583 (12) | 0.533 (15) |
| 4 | 0.400 (10) | 0.875 (8) | 0.289 (7) | 0.455 (11) | 0.500 (14) | 0.833 (18) |
| 5 | 0.667 (6) | 0.714 (7) | 0.727 (11) | 0.143 (14) | 0.385 (13) | 0.750 (16) |

表 7 は、高評価一致率の属性について、実験協力者が選択したレビューで言及されていた回数を比較している。表より、こだわりありと推定された属性については、十分な数のレビューが収集されており、推定の信頼性という点では効果があると考えられる。以上より、動的提示手法ではこだわりがありそうな属性に注目して効率的に情報を収集可能である一方、ランダム提示と比較して属性の見逃しが発生しやすいといえる。推薦システムでの利用を考えた場合、対象ユーザがこだわりを持つ属性をすべて明らかにしてから推薦を行いたい場合にはランダム提示の方が適しているといえるが、少数の属性で十分な場合、迅速に推薦を行いたい場合には動的提示の方が適していると考えられる。

4.3 アイテム推薦結果の評価

各手法による推薦精度を表 8 に示す。推薦精度は、提示したアイテムのうちユーザが好評と判断したアイテムの割合である。表より、ランダム提示、動的提示ともに満足度ランキングによる推薦よりも良い結果が得られていることがわかる。この結果より、レビュー閲覧履歴から価値観モデルを構築するアプローチは情報推薦に有効といえる。

表 5: アンケート回答との対応 (正解/誤検出/見逃し).

| グループ | 動的提示 | ランダム提示 |
|--------|----------|----------|
| 観光 A | 5/4/11 | 2/6/14 |
| 観光 B | 4/2/10 | 7/4/7 |
| ビジネス A | 1/6/12 | 4/3/9 |
| ビジネス B | 5/4/5 | 8/5/2 |
| 合計 | 15/16/38 | 21/18/32 |

表 6: 興味がある属性の平均評価一致率.

| グループ | 動的提示 | ランダム提示 |
|--------|-------|--------|
| 観光 A | 0.536 | 0.547 |
| 観光 B | 0.567 | 0.675 |
| ビジネス A | 0.483 | 0.636 |
| ビジネス B | 0.538 | 0.744 |

動的提示はランダム提示よりも推薦精度が低くなっているが、これはユーザモデリングの実験結果と同様、効率を優先した結果であると考えられる。

5 むすび

本稿では、ユーザがレビューを閲覧・評価した履歴に基づき価値観をモデル化する手法を提案した。提案手法では、ユーザが参考にしたレビューにおける属性評価の極性と、アイテムに対し下した総合評価の極性に基づき、評価一致率を計算する。得られたユーザモデルに基づき、推薦アイテムを決定する手法、評価一致率が高くなる可能性のある属性を優先的に推定するために、提示するレビューを動的に決定する手法を提案した。

ユーザ実験により提案手法を評価した結果、レビュー閲覧履歴から構築したユーザモデルに基づく推薦の精度は満足度ランキングによる推薦を上回る結果が得られた。また、レビューの動的提示手法では、ランダム提示手法よりも推薦精度などは低下する結果となったが、評価一致率の高い属性については十分なユーザフィードバックが得られており効率的なモデリングには効果があることを示した。

従来の価値観モデリング手法では、対象ユーザが投稿したレビューや、興味ある属性についての明示的なフィードバックを必要とした。これに対し、本研究で着目したレビュー閲覧はオンラインショッピングサイトなどで多数ユーザにより一般的に行われる行為であるため、価値観モデリングに基づく情報推薦システムの適用範囲拡大に貢献することが期待できる。

表 7: 高評価一致率の属性におけるレビュー数の比較.

| 提示条件 | 評価一致率 | 10 件以上 | 10 件未満 |
|--------|--------|--------|--------|
| 動的提示 | 0.7 以上 | 28 | 2 |
| | 0.8 以上 | 17 | 0 |
| ランダム提示 | 0.7 以上 | 13 | 21 |
| | 0.8 以上 | 9 | 12 |

表 8: 推薦精度の比較.

| グループ | 動的提示 | ランダム提示 | 満足度 |
|------|-------|--------|-------|
| 観光 | 0.720 | 0.800 | 0.670 |
| ビジネス | 0.630 | 0.710 | 0.580 |
| 合計 | 0.675 | 0.755 | 0.625 |

謝辞

本研究は JSPS 科研費 JP16K12535 の助成を受けたものです.

参考文献

- [1] P. T. Costa, R. R. McCrae: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI), Psychological Assessment Resources, 1992.
- [2] S. Hattori, Y. Takama: Recommender system employing personal-value-based user model, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 18, No. 2, pp. 157–165, 2014.
- [3] 服部俊一, 高間康史: 属性に対する価値判断に基づく評判情報からのユーザモデル作成手法, 第 97 回数値モデル化と問題解決研究発表会, No. 14, 2014.
- [4] 伊藤ゆかり, 波多野賢治, 松本尚宏: 商品の価値とユーザの嗜好を考慮した商品推薦システムの提案, 電子情報通信学会第二種研究会技術報告, vol. WI2201027, pp.101–106, 2010.
- [5] 神瀧敏弘: 推薦システムのアルゴリズム (1), 人工知能学会誌, Vol. 22, No. 6, pp. 826–837, 2007.
- [6] 三澤遼理, 服部俊一, 高間康史: 価値観に基づくユーザモデルによる協調フィルタリングの拡張手法の提案, 第 28 回人工知能学会全国大会 (JSAI2014), 1H4-NFC-01a-5, 2014.
- [7] M. A. S. N. Nunes, R. Hu: Personality-based Recommender Systems: an Overview, Rec-Sys’12, pp. 5–6, 2012.
- [8] F. Pachet, P. Roy, D. Cazaly: A combinatorial approach to content-based music selection, IEEE International Conference on Multimedia Computing and Systems, pp. 457–462, 1999.
- [9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl: GroupLens: an open architecture for collaborative filtering of netnews, 1994 Conference on Computer Supported Cooperative Work, pp. 175–186, 1994.
- [10] M. Rokeach: The Nature of Human Values, New York: The Free Press, 1973.
- [11] A. I. Schein, A. Popescul, L. H. Ungar, D. M. Pennock: Methods and metrics for cold-start recommendations, Special Interest Group on Information Retrieval, pp. 253–260, 2002.
- [12] 清水涼人, 服部俊一, 高間康史: レビュー閲覧履歴からの価値観に関するユーザモデル構築手法の提案, 第 28 回人工知能学会全国大会, 3B4-OS-10b-3, 2014.
- [13] W. Wu, L. Chen, L. He: Using Personality to Adjust Diversity in Recommender Systems, 24th ACM Conf. on Hypertext and Social Media, pp. 225–229, 2013.

語の分散表現と上位下位関係—研究動向と今後への試案— Recent Research Trends and Proposal of Distributional Word Representations for Hypernymy Detection

鷲尾光樹^{1*}
Koki Washio¹

¹ 東京大学大学院総合文化研究科

¹ Graduate School of Arts and Sciences, The University of Tokyo

Abstract: Distributional representation for words is an important model of word senses which reflects several types of semantic relations, not only similarity relations. In this paper, we provide an overview of unsupervised/supervised approaches with word distributional representations to detect hypernym-hyponym relations and a recently reported problem associated with these approaches. Moreover, we propose a future research direction to solve this problem and prove the validity of this direction with small experiments.

1 はじめに

語の分散表現では、分布意味論に基づいて、語の分布の情報からベクトルを作成し、単語ベクトルとして語の意味を表現する [22][7]。単語ベクトル間の距離などを用いて、語の意味の類義性などが判断できることから、語の分散表現は質問応答システムや情報抽出などの様々なタスクに貢献している。また、語の分散表現はコーパスから自動的に獲得できるため、人手で作成したリソースに存在しない語に対応できることも大きな利点となっている。近年では、語の類義関係に留まらず、アナロジーなど、単語間の様々な意味関係を分散表現を用いて推測するという意味タスクの研究がなされている。そのような研究の一つとして、分散表現を用いた二語の上位下位関係の推測・学習が注目されている。これは、「動物」が「犬」の上位語であり、「犬」が「動物」の下位語であるというような意味関係を、それぞれの語の分散表現から推定するタスクである。

本稿では、分散表現からある二語の上位下位関係を推測する研究と問題点について概観し、今後の研究の方向性を提案するとともに、その提案の妥当性を裏付ける簡単な実験の報告を行う。

2 分散表現

分布意味論においては、語の出現文脈によって、単語の意味を捉えるため、分散表現を獲得する際は、ま

ず単語の分布を見る際の文脈を規定する。文脈に何をを用いるかは様々だが、代表的なものとしては、近傍共起や依存構造に基づく関係を利用するものが挙げられる。近傍共起では、対象の語の前後数語を文脈窓として共起を測る。一方、依存構造においては、主語と動詞の関係など、文中において対象の単語が他の語とどのように関わっているかを文脈として用いる。分散表現の性質はこのような文脈の選択により決定される。たとえば、文脈窓を採用した場合は話題的・領域的な類似性がベクトル空間上で表現され、依存構造などの統語的な文脈を用いた場合は、語の品詞などを考慮した機能的な類似性が捉えられるという報告がある [9]。しかし、様々な意味タスクにおける性能の良さや、コーパスを処理するコストの低さから、自然言語処理における分布意味論の分野では文脈窓が用いられることが多い。

以上から、本稿では文脈窓の利用を前提としつつ、本節においては二つの分散表現の獲得法について説明する。一つは共起頻度に基づく古典的な共起頻度ベクトル(カウントベースの分散表現)であり、もう一つは近年台頭してきたニューラルネットワークによる学習から獲得する単語埋め込みベクトル(ニューラルベースの分散表現)である。

なお、本稿では以下の記法を用いる。単語 w の集合を V_W 、単語の文脈(文脈窓中に出現する語) c の集合を V_C とし、 w の単語ベクトルを \vec{w} 、文脈 c の単語ベクトルを \vec{c} とする。また、コーパスで観察された w と c の共起 (w, c) の集合を D とし、コーパスで観察された要素の頻度を返す関数を f とする。

*連絡先： 東京大学総合文化研究科言語情報科学専攻
〒 153-8902 東京都目黒区駒場 3-8-1
E-mail: kkwashio3333@gmail.com

2.1 カウントベースの分散表現

カウントベースの分散表現獲得では、文脈に基づいて共起頻度を集計し、 $|V_W| \times |V_C|$ の共起頻度行列 M を作成する。単語 w に該当する行を行列 M から切り出すことで、単語ベクトル \vec{w} が得られる。カウントベースの分散表現は、高次元でスパースな表現であり、 M のほとんどの要素は0である。また、後に述べるニューラルベースの分散表現と異なり、カウントベースの分散表現においては \vec{w} の各次元が単語 w と文脈 c の結びつきの強さを表しており、各次元の持つ意味が明確であることが特徴である。

行列 M の要素として、生の共起頻度をそのまま用いた場合、 w や c の頻度のばらつきが単語ベクトルに悪影響を及ぼす可能性がある。このような問題を回避するために、共起頻度行列 M の各要素を以下で定義される PMI(pointwise mutual information, 相互情報量) に変換することがよく行われる。

$$\begin{aligned} PMI(w, c) &= \log_2 \frac{P(w, c)}{P(w)P(c)} \\ &= \log_2 \frac{f(w, c)|D|}{f(w)f(c)} \end{aligned}$$

PMI は w と c が共起する確率からそれぞれの出現確率を差し引いており、正確に二語の結びつきの強さを測ることができる。しかし、 w と c の共起が観察されなかった場合、 $PMI(w, c) = -\infty$ になるため、実際は以下の PPMI(positive pointwise mutual information, 正の相互情報量) が用いられることが多い。

$$PPMI(w, c) = \begin{cases} 0 & (PMI(w, c) \leq 0) \\ PMI(w, c) & (PMI(w, c) > 0) \end{cases}$$

PPMI は生の共起頻度を用いるより、様々な意味タスクにおいて良い成績を残す一方で、低頻度な w や c を含むペアに高い値を返してしまうバイアスがあることが知られている [12]。

また、行列 M のスパースネスを解消するために、特異値分解によって $|V_W| \times |V_C|$ の行列 M を $|V_W| \times d$ (d は数百次元) に圧縮して用いることもある。この場合、文脈 c を表していた単語ベクトルの各次元の意味は失われる。

2.2 ニューラルベースの分散表現

ニューラルベースの分散表現獲得では基本的に、ニューラルネットワークの入力層と出力層に単語や文脈を配置して学習を行い、 d 次元の隠れ層を低次元で密な単語の表現として取り出す。カウントベースの分散表現とは対照的に、各次元の意味は不明瞭であり、直観

を働かせることはできない。本稿では、文献 [14] と文献 [15] で提案された、ネガティブサンプリングを用いた SkipGram モデル (SGNS) を説明する。SGNS は意味タスクにおいて汎用的に用いられている分散表現獲得法である。

SGNS は、コーパスで観察された共起 (w, c) に対し、 d 個のユニットからなる隠れ層を持つニューラルネットワークの入力層に、 w に該当する次元のみ 1 (他の次元は 0) の $|V_W|$ 次元 one-hot ベクトルを配置し、出力層には c に該当する部分のみ 1 の $|V_C|$ 次元 one-hot ベクトルを配置して学習を行う。その結果において、入力層側にある $|V_W| \times d$ のパラメータ行列の w に該当する行を単語ベクトル \vec{w} 、出力層側にある $d \times |V_C|$ のパラメータ行列の c に該当する列を文脈ベクトル \vec{c} とみなす¹。

以下では SGNS の目的関数について説明する。いま、 (w, c) が D に含まれる確率を、

$$P(D = 1|w, c) = \frac{1}{1 + \exp(-\vec{w} \cdot \vec{c})}$$

のようにモデル化すると、目的関数は、

$$\begin{aligned} \arg \max_{\vec{w}, \vec{c}} \prod_{(w, c) \in D} P(D = 1|w, c) \\ = \arg \max_{\vec{w}, \vec{c}} \sum_{(w, c) \in D} \log \frac{1}{1 + \exp(-\vec{w} \cdot \vec{c})} \end{aligned}$$

となる。このままだと、すべての (w, c) に対し $P(D = 1|w, c) = 1$ になってしまうパラメータが存在するため、各 (w, c) に対し、 D に存在しない k 個のペア $((w, c_1), \dots, (w, c_k))$ を負例として考慮して学習する。 c_j は $P(c)$ の $3/4$ 乗に従ってサンプリングを行う。負例の集合を D' とすると、ネガティブサンプリングを用いた場合の目的関数は、

$$\begin{aligned} \arg \max_{\vec{w}, \vec{c}} \prod_{(w, c) \in D} P(D = 1|w, c) \prod_{(w, c) \in D'} P(D = 0|w, c) \\ = \arg \max_{\vec{w}, \vec{c}} \sum_{(w, c) \in D} \log \frac{1}{1 + \exp(-\vec{w} \cdot \vec{c})} \\ \sum_{(w, c) \sim P(c)^{\frac{3}{4}} \in D'} \log \frac{1}{1 + \exp(\vec{w} \cdot \vec{c})} \end{aligned}$$

となる。この目的関数を確率的勾配降下法などを用いて最適化すると、 d 次元の単語埋め込みベクトルが得られることになる。

ニューラルベースで獲得した低次元で密な分散表現はカウントベースの分散表現と同様に、様々な意味タスクに用いることができる。

¹ w の one-hot ベクトルに入力層側のパラメータ行列をかけると、 d 次元の単語埋め込みベクトル \vec{w} が得られる。

2.3 カウントベース vs. ニューラルベース

文献 [2] では、様々な類義語の認識やアナロジーの計算などの様々な意味タスクにおいて、カウントベースの分散表現とニューラルベースの分散表現を用いた際の性能が比較されており、基本的にはニューラルベースの分散表現の性能が良いと報告されている。

一方で、ニューラルベースの分散表現獲得法がカウントベースの獲得法よりアルゴリズムとして優れているという主張に疑問を投げかける、Levy らによる一連の報告がある [11][10][13]。

たとえば、文献 [11] では、SGNS などの word2vec 系の分散表現獲得法は、カウントベースの PMI 行列を行列分解していることと等価であると述べられている。この報告によると、SGNS の目的関数を展開し、 $\vec{w} \cdot \vec{c}$ で微分することで、

$$\begin{aligned}\vec{w} \cdot \vec{c} &= \log \left(\frac{f(w, c) \cdot |D|}{f(w) \cdot f(c)} \right) - \log k \\ &= PMI(w, c) - \log k\end{aligned}$$

が得られる。このとき、SGNS における入力層側のパラメータ行列を W 、出力層側のパラメータ行列を C とすると、

$$W_i \cdot C_j = PMI(w_i, c_j) - \log k$$

となる。この式の右辺はカウントベースの共起頻度行列 M の各要素を PMI 行列に変換し、各要素からネガティブサンプリングに用いた負例ペアの数の対数をとった値を差し引いた行列である。結局、SGNS で行われている最適化は、単語ベクトルと文脈ベクトルの内積が PMI 行列の各要素から $\log k$ を引いた値になるように学習しているということになる。以上のことから、SGNS は特異値分解とは異なるが、同じ行列分解の一種とみなすことができる²。

また、ニューラルベースの分散表現が注目された理由のひとつに、アナロジーの計算ができることが挙げられる。アナロジーの計算とは、たとえば、「女性」のベクトルから「男性」のベクトルを引き「王様」のベクトルを足すと、「女王」のベクトルに非常に近いベクトルになるというものであり、ニューラルベースの分散表現は意味関係を簡単足し引きで表現できると言われていた [14][15][16]。しかし、アナロジーを計算するための式を改良することで、ニューラルベースの分散表現とカウントベースの分散表現の両表現でアナロジータスクの性能が向上し、またそれぞれの表現を用いた際の性能が拮抗するという報告がある [10]。この

²word2vec 系でも CBoW モデルや、別のニューラルネットワークモデルでは事情が異なる。一方、同じくニューラルベースの分散表現獲得法である GloVe は、行列分解とみなせるとの報告がある [20]。

ことから、アナロジーの計算に必要な情報は、ニューラルベースの分散表現のベクトル空間のみにおいて捉えられているわけではなく、カウントベースの分散表現獲得法でも捉えられていることがわかる。

さらに、文献 [13] では、SGNS などのニューラルベースの分散表現獲得法に実装されているネガティブサンプリングやサブサンプリング（高頻度語の文脈からの確率的な排除）などを、分散表現獲得の際のハイパーパラメータとみなし、同等の処理をカウントベースの分散表現獲得法にも適用することで、様々な意味タスクにおける両分散表現の性能が拮抗するという実験結果が報告されている。この結果を踏まえて Levy らは、ニューラルベースの分散表現の優位性は、アルゴリズムそのものが優れているのではなく、獲得手法にデフォルトで設定されているハイパーパラメータが性能の向上に寄与した結果であり、カウントベースの分散表現獲得法とニューラルベースの分散表現獲得法に本質的な差はないと結論づけている。一方で Levy らは、SGNS は分散表現の獲得が早く、かつ、非常にロバストな表現であり、意味タスクにおいて際立った性能を出すことは少ないが、性能が急激に落ちることもなく、実験を行う際のベースラインとして優れていると報告している。

以上の一連の報告から、カウントベースの分散表現とニューラルベースの分散表現のどちらが優れているかは一概に論じることが難しく、意味タスクごとに最適な分散表現とハイパーパラメータを探索することが重要であることがわかる。

3 上位下位関係の学習

上位下位関係を分散表現から学習する場合、類義関係などの対称的な意味関係と異なり、二つの分散表現に対して非対称な指標や関数を求める必要がある。本節では、2 節の手法を用いて獲得した分散表現を用いて語の上位下位関係を学習するための二つのアプローチ（教師なし学習と教師あり学習）と、文献 [12] で報告された教師あり学習の問題点について概観する。

3.1 教師なし学習

分散表現を用いた上位下位関係推測研究の初期においては、二語の分布の包含性を測る指標が研究されていた。これらの一連の研究は、内省的な分析に基づいた指標の提案を行うものであり、訓練データを利用した機械学習によらないという意味で、上位下位関係の教師なし学習と呼ばれている。このアプローチにおいては、基本的に以下の二つの分布意味論的直観が前提とされている。

- 分布一般性 (Distributional generality)[22]
- 分布包含仮説 (Distributional inclusion hypotheses)[5]

分布一般性とは、意味の広い語はコーパス上に広く分布するという直観である。たとえば、「動物」という単語は「犬」という単語に比べて、「犬」とは共起しにくそうな「泳ぐ」や「飛ぶ」といった単語とも共起するはずである。この直観をもとに、広く分布する語はより上位語らしく、分布が狭い語はより下位語らしいと判断することができる。一方、分布包含仮説とは、ある二語が上位下位関係にあるならば、下位語の出現文脈がある程度上位語の出現文脈に含まれているという直観である。たとえば、「犬」は「走る」や「吠える」などと共起するはずだが、「動物」も同様の文脈に出現するはずである。ここから、分布の包含関係を見ることで、上位下位関係を判断できると期待できる。この二つの前提のもとで、二語の分散表現の各次元の値を見比べることで、上位下位関係性を判断するための指標が提案されてきた。なお、教師なし学習においては各次元の意味が明確である必要があるため、基本的に分散表現はカウントベースのものを用いる。以下では、代表的な指標を紹介する。なお、以降では w_1 と w_2 の二語を扱うとする。 $\vec{w}_i = (w_{i1}, \dots, w_{in})$ であり、関数 F をベクトルの 0 ではない素性の集合を返す関数とする。

分布意味論的観点から提案された最初の指標は、*Weeds* という指標である [22]。

$$WeedsP(w_1, w_2) = \frac{\sum_{i \in F(w_1) \cap F(w_2)} w_{1i}}{\sum_{i \in F(w_1)} w_{1i}}$$

$$WeedsR(w_1, w_2) = \frac{\sum_{i \in F(w_1) \cap F(w_2)} w_{2i}}{\sum_{i \in F(w_2)} w_{2i}}$$

これは、情報検索などの評価指標として用いられる precision と recall を二語の関係性を捉える際に適用し、分布の包含性を捉えようとしたものである。いま w_1 が下位語、 w_2 が上位語であるとする、下位語の分布は上位語の分布にある程度包含されるため、*WeedsP* の値は 1 に近づき、*WeedsR* の値は 0 と 1 の間に収まるという指標である。判別においては、*WeedsP* のみ、あるいは *WeedsP* - *WeedsR* を用いて、閾値を設定して上位下位関係の有無を判断する。

Weeds から派生した指標として、*Clarke*[3] と *invCL*[8] という指標がある。

$$ClarkeP(w_1, w_2) = \frac{\sum_{i \in F(w_1) \cap F(w_2)} \min(w_{1i}, w_{2i})}{\sum_{i \in F(w_1)} w_{1i}}$$

$$ClarkeR(w_1, w_2) = \frac{\sum_{i \in F(w_1) \cap F(w_2)} \min(w_{1i}, w_{2i})}{\sum_{i \in F(w_2)} w_{2i}}$$

$$invCL(w_1, w_2) = \sqrt{ClarkeP(w_1, w_2)(1 - ClarkeR(w_1, w_2))}$$

Clarke は *Weeds* に近い指標であり、*ClarkeP* のみ、あるいは *ClarkeP* - *WeedsR* を用いて、閾値を設定する。*invCL* は *Weeds* や *Clarke* と異なり、下位語が上位語にどれだけ包含されているかのみならず、下位語の出現文脈の以外の上位語の分布の広さを考慮した指標になっている。*invCL* も閾値を設定して、二語の上位下位関係の有無を判断させる。

また、*invCL* のような指標の派生として、以下の様な指標も考えることができる [21]。

$$simdiff(w_1, w_2) = \frac{\sum_{i \in F(w_1) \cup F(w_2)} \min(w_{1i}, w_{2i})}{\sum_{i \in F(w_1) \cup F(w_2)} \max(w_{1i}, w_{2i})} \cdot \frac{\sum_{i \in F(w_2) - F(w_1)} w_{2i} - \sum_{i \in F(w_1) - F(w_2)} w_{1i}}{\sum_{i \in F(w_1) \cup F(w_2)} \max(w_{1i}, w_{2i})}$$

これは Jaccard 尺度 (第 1 項) と分布が重なっていない部分の文脈の広さの差 (第 2 項) の積である。つまり、二語がどれくらい似ているかと上位語と下位語にどれくらい分布の広さに差があるかを考慮している。

また、ここでは深く触れないが、 w_1 と w_2 の分布の包含を測る際、AP (Average Precision) を応用することで、 w_1 が w_2 とより結びつきが強い文脈語と共起する場合により重みを割り振った、*balAPinc* という指標もある [7]。

教師なし学習の性能を測るために、Erk が行った実験 [4] を参考にして、以下の実験を行った。対象の語を動詞にしぼり、BNC (British National Corpus) コーパス前半の 5000 万語から、動詞の主語と目的語を依存構造文脈として集計し、PPMI 行列を作成した。データセットとしては、WordNet3.0 から同義語集合を一つしか持たない単義の動詞と、その直接の上位語のペアを正例とし、負例は下位語と上位語の正例以外の組み合わせにより作成した。条件ごとに、負例から正例と同じ数のペアをランダムにサンプリングし、それぞれの指標を用いて最適な閾値を設定した場合の正解率は以下のようになった。なお、*Weeds* と *Clarke* に関しては precision から recall を引いたものを指標として用いている。

表 1: 教師なし学習の性能

| | 全負例 | 負例 1 | 負例 2 | 負例 3 |
|----------------|------|------|------|------|
| <i>Weeds</i> | 0.74 | 0.59 | 0.83 | 0.72 |
| <i>Clarke</i> | 0.74 | 0.61 | 0.83 | 0.72 |
| <i>invCL</i> | 0.71 | 0.64 | 0.83 | 0.65 |
| <i>simdiff</i> | 0.74 | 0.64 | 0.83 | 0.72 |

全負例はすべての負例ペア、負例 1 はちぐはぐな上

位下位関係のペア、負例2は下位語同士のペア、負例3は上位語同士のペアからそれぞれサンプリングした場合の正解率である。ちぐはぐな上位下位関係のペアとは、下位語の集合と上位語の集合から、それぞれ一語ずつ抜き出し、正例にないペアを作ることによって作り出した負例である。用いるデータセットにもよるが、頻度の少ない下位語同士のペアを負例と判断するような簡単な場合(負例2)を除くと、分布の包含関係に基づく指標は、だいたい6割から7割5分くらいの正解率であることがわかる。

また、分布包含仮説に依拠せず、文脈語の分布に着目した指標として、文献[19]で提案されたSLQSという指標がある。この指標は、上位語はより広い意味を持つ文脈語と共起し、下位語は狭い意味を持つ文脈と共起するという直観に基づいている。たとえば、「動物」という語は「走る」や「飛ぶ」といった一般的な語と共起するが、「犬」は「吠える」などの具体的な語と共起しやすい。このような直観に基づけば、二語がよく共起する文脈語のベクトルを見て、それらのエントロピーを比べることで上位下位関係性が特定できるはずである。SLQSでは、対象の語と結びつきが強い上位N個の文脈語のエントロピーを計算し、その中央値をとって比較する。いま、文脈語ベクトル $\vec{c} = (c_1, \dots, c_n)$ のエントロピーを以下のように定義する。

$$H(c) = - \sum_{i=1}^n p(c_i|c) \cdot \log_2(p(c_i|c))$$

$p(c_i|c)$ は文脈語 c の頻度と各共起頻度の割合である。これをMinMaxスケールリングにより0から1の値を取るようにスケールリングし、その値を $H_n(c)$ と定義すると、単語 w_i の文脈語のエントロピーの中央値は、

$$E_{w_i} = Me_{j=1}^N(H_n(c_j))$$

と定義される。ただし、 Me は中央値を返す関数である。このときSLQSは、

$$SLQS(w_1, w_2) = 1 - \frac{E_{w_1}}{E_{w_2}}$$

となる。 $SLQS(w_1, w_2) > 0$ のとき、 w_1 は w_2 の下位語と判断される。

3.2 教師あり学習

分散表現を用いた上位下位関係の教師あり学習では、二つの単語ベクトルに何らかの演算を施して特徴ベクトルとして扱い、二語が上位下位関係を持つか否かの二値分類を学習する。学習アルゴリズムには、SVMを用いるもの[18][23]、ロジスティック回帰を用いるもの[18]などがある。二つのベクトルに施す演算としては、

差をとったり、二つのベクトルを結合して特徴量とする方法の性能がいいことが知られている。

教師あり学習の性能を確認するために、以下の上位下位関係性の二値分類実験を行った。

分散表現には、BNC約1億語から、前後二語の文脈窓を採用したSGNSを用いて獲得したものをを用いた³。データセットにはBLESS[1]を用いた。このデータセットは曖昧性のない200語の名詞について、上位下位関係や類義関係、部分全体関係、ランダムな関係にある語などを収集して作成されたものである。

BLESSにおいて、上位下位関係にある1337ペアから、分散表現が獲得できた1155ペアを正例として扱い、同じく分散表現が獲得できた上位下位関係以外の名詞ペアを負例とした。正例ペアと同じ数だけ負例ペアをサンプリングし、SGNSによって獲得した500次元の単語ベクトルの差を特徴量として、10分割交差検定のロジスティック回帰を行う。これを50回繰り返したとき、平均スコアとして0.93の分類正解率、0.93のF値、また、ペアに割り当てた確率をもとにデータを並べ替えAP(Average Precision)を計算したところ、0.97をマークした。文脈や分散表現の性質、用いているデータセットが異なるものの、教師あり学習は3.1節で紹介した教師なし学習の手法よりも良い成績を残している。

3.3 教師あり学習の問題点

教師あり学習は高い分類精度を誇る一方で、上位語に位置しやすい単語を覚えているだけであるという問題点が指摘されている。文献[12]では、このような教師あり学習の振る舞いを調べるために、様々なデータセットを用いて二つの実験が行われている。

ひとつは、訓練データとテストデータの単語ペアの語彙の重なりをなくして分類を行う実験である。これによって、教師あり学習の分類性能が大きく下がり、データセットによっては教師なし学習を下回るの性能になってしまうことが観測されている。また、二つのベクトルの差や結合を用いた場合と、上位語のみを用いて学習した場合の性能の比較を行うと、性能差がごく小さいことがわかった。この実験によって、教師あり学習は下位語の情報をほぼ無視していることが明らかになった。

もうひとつの実験は、様々な条件で学習を行った分類器に、ちぐはぐな上位下位関係のペアを判断させる実験である。ちぐはぐな上位下位関係を誤って正例として分類してしまう割合をmatch errorとして算出し、様々なモデルでrecallとの相関を調べたところ非常に強い相関があり、 $matcherror = 0.935 \cdot recall$ と線形回帰できてしまうことが明らかになった。つまり、正

³分散表現の獲得には、Omer Levyが公開しているhyperwordsを用いた。<https://bitbucket.org/omerlevy/hyperwords>

しく上位下位関係を持つペアを正例として多く分類できる分類器は、その分、ちぐはぐな上位下位関係を持つペアも誤って正例として分類してしまうことになる。

これらの実験から文献 [12] は、教師あり学習では典型的な上位語を覚えているだけであり、二語の関係性は学習できていないと結論づけている。

この結論を追証するために、3.2 節で行った実験と、コーパスや分散表現は同じ条件で、訓練データとテストデータの語彙の重なりをなくした場合の結果を以下に示す。

表 2: 教師あり学習における語彙の重なりの影響

| | 重なりあり | 重なりなし | Δ |
|-------|-------|-------|----------|
| 分類正解率 | 0.93 | 0.68 | 0.25 |
| F 値 | 0.93 | 0.61 | 0.32 |
| AP | 0.97 | 0.77 | 0.20 |

(Δ は「重なりあり」と「重なりなし」の差を表す)

表 5 を見ると重なりがある場合と重なりがない場合に性能差があることがわかる。教師あり学習が二語の関係性を学習できていないとすると、そのようなモデルは訓練データにない語彙に対応できない。これは、コーパスからの自動的学習のそもそものメリットである、「人手によるリソースに存在しない語彙に対応できる」という点においては致命的である。

4 考察と今後の方向性

3.3 節で述べたような教師あり学習の典型的な上位語の記憶という問題への対処としては、二つの方向性が考えられる。ひとつは分散表現獲得法の最適化であり、もうひとつは学習法の見直しである。

4.1 分散表現の最適化

分散表現は、語の文脈によって語の意味を表現するという分布意味論に基づいて獲得されるが、そのモデル自体が特定の意味タスクを志向しているわけではない。よって、ある意味タスクの性能の向上のためには、それぞれの意味タスクが着目している語の意味の側面を正しく反映するようなハイパーパラメータの調整や、モデルそのものの変更が必要になる。たとえば、上位下位関係の認識においては話題的・領域的な類似性だけでなく、語の機能的な類似性も見ることが必要である。現在は他の様々な意味タスクでの性能の良さや処理コストの低さから文脈窓を用いられることが多いが、上位下位関係認識においては、文脈に依存構造を採用することによって、ちぐはぐな上位下位関係を

正例と判断してしまう割合が減少することが考えられる⁴

4.2 学習法の見直し

上位下位関係認識の教師あり学習においては、特徴量として二つのベクトルの差や結合を用いるだけでは、二語の関係性を学習できていないことがわかっている [12]。これに対しては、上位下位関係に関して分布的な意味付けがより明確な特徴を用いることによって、二語の関係性の学習を促進できる可能性がある。そのような特徴として、今まで研究されてきた教師なし学習の指標を挙げるができる。これらの指標を特徴として採用することで、二語の分布の包含関係や形状の違いなどの関係性が学習されるはずである。

4.3 実験

これらのアプローチの妥当性を検討するために、3.1 節で述べたような、教師なし学習として提案された指標群を、教師あり学習における特徴量として採用した場合の性能評価を行った。

分散表現として BNC 約 1 億語から前後二語の文脈窓を共起として獲得した PPMI 行列を採用し、データセットは 3.2 節と同じく BLESS における名詞ペアを用いた。

特徴量としては以下のものを採用した。

- 二語の類似度: \cos 類似度
- 分布の包含関係: $WeedsP/R$, $ClarkeP/R$, $invCL$
- 分布の形状: 二語の分布のエントロピーの差と比、それぞれのベクトルにおいて値が高い上位 50 次元の値の平均の差
- 文脈の分布: $SLQS$

まず、二語の類似度を見るために \cos 類似度を特徴として考慮し、分布の包含関係を見るために、 $WeedsP/R$ 、 $ClarkeP/R$ 、 $invCL$ を採用した。また、分布の形状の違いを見るために二語の分布のエントロピーの差と比、それぞれの単語ベクトルにおいて値が高い上位 50 次元の値の平均の差を採用した。これは上位語は分布がなだらかで広く、下位語は狭い文脈にしか出現しないという直観に基づく。さらに結びつきが強い文脈語の分布を見るために $SLQS$ を特徴として考慮する。これによって、二語の分散表現から 10 次元の特徴ベクトルを算出することができる。

⁴ただし、大規模コーパスに対し依存構造解析を行うコストや、精度の良いパーサーが存在する言語の少なさを考えると、文脈窓の採用には妥当性がある。

表 3: 提案手法における語彙の重なりの影響

| | 重なりあり | 重なりなし | Δ |
|-------|-------|-------|----------|
| 分類正解率 | 0.61 | 0.55 | 0.06 |
| F 値 | 0.59 | 0.50 | 0.09 |
| AP | 0.64 | 0.59 | 0.05 |

表 4: cos 類似度のみ (baseline)

| | 重なりあり | 重なりなし | Δ |
|-------|-------|-------|----------|
| 分類正解率 | 0.54 | 0.49 | 0.05 |
| F 値 | 0.51 | 0.41 | 0.10 |
| AP | 0.56 | 0.56 | 0.0 |

この特徴ベクトルをもとに、上位下位関係のペアを正例として、正例と負例の数を揃えてロジスティック回帰を行い、10 分割交差検定を行った場合の性能と、3.3 節で行った実験と同様に、訓練データとテストデータの語彙の重なりをなくした場合の性能を比較すると、表 3, 表 4 のようになった。

なお、baseline は cos 類似度のみを用いた場合とした。提案手法はいずれの場合でもベースラインを上回りつつ、SGNS よりも語彙の重なりの影響が少ないことがわかる。この結果から、二語の関係性としての意味付けが明確なものの特徴量を用いると、性能は低いものの、語彙の重なりの影響が少ないことがわかる。

さらに、SGNS で獲得した二語の単語ベクトルの差に、今回算出した特徴ベクトルを結合して学習した場合の性能 (SGNS+提案手法) と、単語ベクトルの差のみを特徴ベクトルとして学習した場合の性能 (SGNSのみ) を比較した。訓練データとテストデータの語彙の重なりをなくした際の結果は以下ようになった。

表 5: 提案手法を付加した場合の性能

| | SGNS のみ | SGNS+提案手法 |
|-------|---------|-------------|
| 分類正解率 | 0.68 | 0.73 |
| F 値 | 0.61 | 0.68 |
| AP | 0.77 | 0.83 |

提案手法の特徴ベクトルを従来の教師あり学習の特徴ベクトルに結合した場合、正解率、 F 値、 AP において性能が向上している。この結果から、従来の特徴ベクトルに教師なし学習の指標を付加することで、二語の関係性の学習が促進されていることがわかる。これによって、教師あり学習の際に、上位下位関係に関して意味付けが明確な分布的特徴を用いることで、二語の関係性の学習が可能であることが示された。

5 おわりに

本稿では、分散表現を用いてある語のペアが上位下位関係を持つか否かを判別する研究とその問題点を概観した。今後の研究の方針として、二語の分散表現の差をや結合を用いて教師あり学習を行った場合、二語の関係性を学習できないが、分散表現から分布的特徴を適切に抽出すれば、二語の関係性の学習を促進できることを示した。これからは、上位下位関係の認識に最適な分散表現の獲得法と、影響力のある分布的特徴を模索していきたい。

謝辞

本研究に関して、終始あたたかくご指導ご鞭撻をして頂き、また本稿の執筆の機会を与えてくださった、東京大学大学院総合文化研究科加藤恒昭教授に心より感謝いたします。

参考文献

- [1] Baroni, M., Lenci, A.: How we BLESSed distributional semantic evaluation., In *Proc. of the ofthe GEMS 2011Workshop on GEometrical Models of Natural Language Semantics*, pp. 1–10 (2011)
- [2] Baroni, M., Dinu, G., Kruszewski, G.: Dont count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors., In *Proc. of the Annual Meeting of the Association for Computational Linguistics(ACL)*, Vol. 2, long paper, pp. 238–247 (2014)
- [3] Clarke, D.: Context-theoretic semantics for natural language: an overview., In *Proc. of the EACL 2009Workshop on GEMS: GEometrical Models of Natural Language Semantics*, pp. 112–119 (2009)
- [4] Supporting inferences in semantic space: representing words as regions., In *Proc. of the International Conference on Computational Semantics(ICCS)*, pp. 104–115 (2009)
- [5] Geffet, M., Dagan, I.: The distributional inclusion hypotheses and lexical entailment., In *Proc. of the Annual Meeting of the Association for Computational Linguistics(ACL)*, pp. 107–114 (2005)

- [6] Goldberg, Y., Levy, O.: word2vec explained: deriving mikolov et al. 's negativesampling., *arXiv preprint*, arXiv:1402.3722. (2014)
- [7] Kotlerman, L., Dagan, I., Szpektor, I., Geffet, M., Directional Distributional Similarity for Lexical Inference., *Natural Language Engineering*, Vol. 16(4), pp. 359–389 (2010)
- [8] Lenci, A., Benotto, G.: Identifying hypernyms in distributional semantic space., In *SEM 2012 The First Joint Conference on Lexical and Computational Semantics*, Vol. 2, pp. 75–79 (2012)
- [9] Levy, O., Goldberg, Y.: Dependencybased word embeddings., In *Proc. of the Annual Meeting of the Association for Computational Linguistics(ACL)*, Vol. 2, Short Paper (2014)
- [10] Levy, O., Goldberg, Y.: Linguistic regularities in sparse and explicit word representations., In *Proc. of the Conference on Computational Natural Language Learning.*, pp. 171-180 (2014)
- [11] Levy, O., Goldberg, Y.: Neural word embeddings as implicit matrix factorization., In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, pp. 2177-2185 (2014)
- [12] Levy, O., Remus, S., Biemann, C., Dagan, I.: Do Supervised Distributional Methods Really Learn Lexical Inference Relations?, In *Proc. of the 2015 North American Chapter of the Association for Computational Linguistics(NAAACL): Human Language Technologies*, pp. 970–976 (2015)
- [13] Levy, O., Goldberg, Y., Dagan, I. Ramat-Gan, I.: Improving distributional similarity with lessons learned from word embeddings., *Transactions of the Association for Computational Linguistics*, 3 (2015)
- [14] Mikolov, T., Chen, K., Corrado, G. S., Dean, J.: Efficient estimation of word representations in vector space., *CoRR*, abs/1301.3781. (2013)
- [15] Mikolov, T., Sutskever, I., Chen, K. Corrado, G. S., Dean, J.: Distributed representations of words and phrases and their compositionality., In *Advances in neural Information Processing Systems*, pp. 3111–3119 (2013)
- [16] Mikolov, T., Yih, W., Zweig, G.: Linguistic regularities in continuous space word representations., In *Proc. the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–751 (2013)
- [17] Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation., In *Proc. of the Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Vol. 2, Short Paper (2015)
- [18] Roller, S., Erk, K., Boleda, G.: Inclusive yet selective: Supervised distributional hypernymy detection., In *Proc. of the International Conference on Computational Linguistics(COLING)*, pp. 1025–1036 (2014)
- [19] Santus, E., Lenci, A., Lu, Q., Walde, S.: Chasing hypernyms in vector spaces with entropy., In *Proc. of the Conference of the European Chapter of the Association for Computational Linguistics*, pp. 38–42 (2014)
- [20] Suzuki, J., Nagata, M.: A Unified Learning Framework of Skip-Grams and Global Vectors., In *Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL/IJCNLP)*, Vol. 2, Short Paper (2015)
- [21] 鷲尾光樹, ベクトル空間表現における動詞の意味関係の現れ, 東京大学教養学部教養学科超域文化科学分科 卒業論文 (2015)
- [22] Weeds, J., Weir, D., McCarthy, D.: Characterising measures of lexical distributional similarity., In *Proc. of the International Conference on Computational Linguistics(COLING)*, pp. 1015–1021(2004)
- [23] Weeds, J., Clarke, D., Reffin, J., Weir, D., Keller, Bill.: Learning to distinguish hypernyms and cohyponyms., In *Proc. of the International Conference on Computational Linguistics(COLING): Technical Papers*, pp. 2249–2259 (2014)