

時系列横断の分散表現空間間の変換行列と単語の語意トラッキング

高橋 克郎[†] 大島 裕明[†]

[†] 兵庫県立大学 応用情報科学研究科 〒 650-0047 兵庫県神戸市中央区港島南町 7-1-28

E-mail: †{ab18y501,ohshima}@ai.u-hyogo.ac.jp

あらまし 我々は、既知情報の例をクエリとして未知の情報を検索するための手法を提案する。例えば、2000年代、*Bush*、1980年代という3つのエンティティが与えられた時に新聞記事のコーパスから「2000年代における *Bush*」という関係から「1980年代における *X*」という関係の *X* にもっとも適切な語を検索するという手法だ。提案手法は、語の分散表現空間間の変換行列を求める手法を基礎としている。特に、関係「2000年代における *Bush*」と関係「1980年代における *Reagan*」の間の関係に似ていることを利用して変換行列を補正する手法に着目した。

キーワード 変換行列, アナロジー検索, 類義語トラッキング

1 はじめに

若いユーザは自分が生まれていない時代の流行や政治家などの知識を知らなかったり、高齢者は今の流行を知らないことが多い。いずれも、ユーザは既知の知識を利用して自分が知らない情報を検索できると便利である。例えば、若者が「2000年代における *Bush*」から「1980年代における *Reagan*」を検索する状況である。あるいは、高齢者が「1980年代における *Reagan*」から「2000年代における *Bush*」を検索する状況である。既知情報検索は既知の例をクエリとして、クエリとの類似性を元に情報を検索をする分野である。以後「*A*における *B*」から最尤の「*C*における *D*」を検索する問題を $(A, B) \sim (C, D)$ 問題と表記する。このような $(A, B) \sim (C, D)$ 問題に関する関連研究としては以下のようなものが挙げられる。[1] は関係性の類似性を尺度とする手法を提案した。彼らは、語の組 (A, B) が与えられた時の言語的な $(A, B) \sim (C, D)$ を解こうとした。この手法は、SVMによる分類がもっとも精度が高かった [1]。Katoらはこの問題を Web 検索エンジンのインデックスに適用した。この研究で取り組んだ $(A, B) \sim (C, D)$ は、例えば $(iPod, Apple) \sim (X, Microsoft)$ における最尤の *X* を求める手法を提案した [2]。Ducらはこの問題を Web コーパスを用いた言語横断での潜在比喻検索に適用した。この研究で取り組んだ $(A, B) \sim (C, D)$ は、例えば、 $(Japan, Mt.Fuji) \sim (Germany, X)$ における最尤の *X* を求める手法を提案した [3]。

歴史言語学では異なる時代での語彙の意味の対応が重要な課題として提起されている ([4], [5], [6])。([7], [8], [9]) により語意の対応関係を自動で算出する手法がいくつか提案されている。[7] は文脈から語の年代を3つにクラスタリングした。[8] は意味の変化を語の組を比較し分析している。

Zhangらは時間横断の分散表現間の変換行列を算出する手法を提案した [9]。1980年代の *Reagan* を知らない若い人を想定してみる。2000年代の *Bush* という彼らにとって既知の情報をクエリとして、1980年代の *Reagan* を探すことは困難である。この問題で、クエリは起点領域の一語 2000年代の *Bush* とし

て、ターゲット領域を 1980年代とする。検出目標の語はターゲット領域の一語である（この場合は *Reagan*）。図1に概念図を示す。

[9] は基本的には二つの時代間の共通語を利用して、 $(A, B) \sim (C, D)$ 問題を解くための取り組みをした。[9] の手法は、最終的に精度が一番良かったのは、上記の手法にさらに、単語間の関係性を有向グラフとして表現し、そのグラフ間の比較を行う手法であった。この手法の計算コストはとても大きいため、実用には向かない。本研究では、より少ない計算コストで同様の性能が出ることを目的の一つとした。

目的の年代に一回で行き着くことは可能であるが、それが自然な変換である保証はどこにもない。入力年代と出力年代の隔たりが大きすぎて自然な変換になっていない場合は特にそうである。私たちの最終的な目的は、既知情報領域と未知情報領域の間に500年以上の隔たりがある場合の類義語トラッキングである。そのためには以下の問題の解決を必要とする。2000年代では二人称複数形 *You* は英語の基本単語の一つです。しかし、1500年代では *Thou* が使用され、*You* は使用されていなかった。これは一例であるが、基本語彙ですら時代の隔たりが大きいと変化する。これは、時代の隔たりが大きいと極端に共通語が少なくなることを意味する。実際に Python での自然言語処理のためのライブラリである nltk の中に収録されている York-Toronto-Helsinki Parsed Corpus of Old English Prose は、古英語のコーパスの例である。このコーパスの中には現代英語との共通語が100語に満たない。古英語におけるエンティティと現代英語におけるエンティティの対応関係は歴史言語学などの文脈ですで行われているが、この対応関係を理解するためには高度な専門知識を要する。システムがこの時間の隔たったエンティティの対応関係を保証する場合には、ユーザは専門知識を必要とせずに目的のエンティティを検索できる。我々は、このシステムの適用先として、これらの専門知識を持たない若年層の歴史教育を例として挙げる。用意された加工されていないコーパスから自動的にこれらの検索が可能になれば、ユーザにかかる負担は最小限となる。当然のこととして、本手法は使用するコーパスに依存した回答が出力される。コーパス

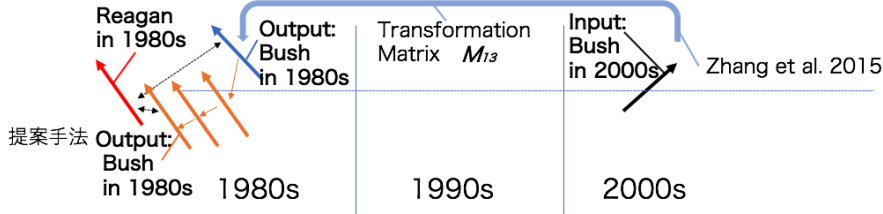


図1 アナロジー検索例. 入力 (*Bush, 2000s, 1980s*) と出力 (*Reagan*) .

に期待されるのは、時系列情報を含み、時間が隔たった場合の対応関係が取れるものが望ましい。この場合はいくつかに年代を分けて順に変換していくのが自然である。先行研究はいずれもこのような複数の時代間の変換を考慮に入れていない。まず我々は、一つのコーパス中での複数の時代間の変換に関する性質を精査する。我々は複数の変換を考慮に入れた [9] の拡張された手法を提案する。

2 クエリに対しての補正を考慮に入れた変換

2.1 問題定義: アナロジー検索

一般に、 $(A, B) \sim (C, D)$ 問題の入力はエンティティ a, b , および c からなるタプル (a, b, c) である。このタプル (a, b, c) を我々はクエリ q と呼ぶ。この解は適切なエンティティ d を出力することである。本紙では、 a, b, c および d を集合の元とし、以下で定義する。 $i, N \in \mathbb{N}$ である。 $D_{i \leq N}$ を文字列の全順序集合とする。 $\{D_i\}_{i \leq N}$ を $D_i (i \leq N)$ の族とする。 $D_i (i \leq N)$ を文書と呼ぶ。 D_i 中の文字列を文と呼ぶ。 W_i を D_i に出現する異なり単語をアルファベット順に並べた全順序集合とする。各 W_i 中の単語の順序はアルファベット順である。 $W_i = B, W_j = \mathbf{D} (W_i, W_j \in \{W_i\}_{i \leq N})$ とする。 $\{W_i\}_{i \leq N}$ は $W_i (i \leq N)$ の族である。 $W := \cup_{i \leq N} W_i$ 。 T_i は D_i に 1 対 1 対応する文字列である。 T_i を D_i のラベルと呼ぶ。 $T := \{T_i\}_{i \leq N}$ とする。 $A, C \in T$ とする。 $B, \mathbf{D} \in \{W_i\}_{i \leq N}$ とする。 \mathbf{Q} を直積 $A \times B \times C$ とする。 \mathbf{Q} をクエリ集合と呼ぶ。 \mathbf{D} はクエリ応答候補の全順序集合である。 $R: T \times W \rightarrow \mathbb{R}^n$ を写像とする。 $\text{Sim}: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ は実数値関数であり、 q と d_k 間の類似度 $\text{rank}(q, d_k)$ を返す。 $\text{rank}(q, d_k)$ は以下で定義される。

$$\text{rank}(q, d_k) := \text{Sim}(R(a, b), R(c, d_k)). \quad (1)$$

$\text{rank}(q, d_k)$ により \mathbf{D} をならべかえる。 $\text{rank}(q, d_k)$ によりならべかえられた \mathbf{D} を出力 $\text{Rank}(q)$ とする。特に、全順序集合 $\text{Rank}(q)$ の最初の元 d_1 を出力と呼ぶこともある。例えば、入力 $q = (2000s, Bush, 1980s)$ が与えられた時、*Reagan* が $d_1 \in \text{Rank}(q)$ として適切である。

2.2 提案手法の枠組み

提案手法は以下の手順により構成される。

手順 1. 語の分散表現への変換

手順 2. 分散表現空間間の変換行列 (直接変換/クエリに対する補正を考慮に入れた変換)

手順 3. 分散表現空間間の類似語のランキングづけ

2.2.1 語の分散表現への変換

[10] は語彙をニューラルネットワークを用いて、その分散表現に変換する手法である。Mikolov らは Skip-gram model を提案した [11], [12]。これは、[10] のモデルの構造をニューラルネットワークを用いて簡略化したものである。このモデルは文脈語の関係性から語意を精密に捉え、平均化する手法であり、語 1,000,000 語単位の語彙を入力として必要とする。 D_i は $T_i \in T$ 年代の文書をさす。Skip-gram モデルは文書 D_i を $n \times p$ 次の行列に変換する。 n は語彙数、 p は特徴ベクトルの次元で恣意的に指定可能である。 $n, p \in \mathbb{N}$ 。結果として、語 w_i の分散表現は p 次元の実数ベクトルである。

2.2.2 直接変換

我々は分散表現空間間の基底変換をその共通語の分散表現間の変換として算出する。共通語が k 語あり、それを $\{w_i\}_{(1 \leq i \leq k)}$ とする。 v_i を入力の分散表現空間での語 w_i をさす分散表現とする。 u_i を出力の分散表現空間での語 w_i をさす分散表現とする。 v_i に u_i が対応するように分散表現をそれぞれスタックし、 V と U とする。 V と U の間の変換行列 M を最小二乗法により以下の式を最小化することにより算出する。

$$M := \arg \min \sum_{i=1, v_i \in V, u_i \in U}^{|V|} \|M \cdot v_i - u_i\|_2^2 + \gamma \|M\|^2. \quad (2)$$

第 2 項は過学習防止のための正規化項である。ここで $\gamma = 0.02$ とする。

2.2.3 クエリに対しての補正を考慮に入れた変換

$U_i \subset W$ とする。 U_i は 2 種類考える。 U_i の元は W_i 中に $m_\phi \in \mathbb{N}$ 回以上出現するすべての $w_i \in W_i$ である。 $v_i^l \in \mathbb{R}^p$ は W_i の l 番目の元の分散表現であり、Skip-gram モデルによる写像 $v_i: W_i \rightarrow V_i; w_i^l \mapsto v_i^l$ により対応づけられる。 $R(a, b) := v_i(b)$ とする。 i は a に依存する。 V_i は任意の $w_i^l \in W_i$ に対応する v_i^l の集合である。 v_i の U_j による制限を、

$$v_i|_j := v_i(U_i \cap U_j), \quad (3)$$

とする。より正確には、 $v_i|_j$ は任意の $b \in B$ に対して $v_i|_j(b)$ アルファベット順にスタックした行列である。行列 $v_i|_j$ の次元は $|U_i \cap U_j| \times p$ とである。

提案手法は以下のように定義される。クエリ q に対して上記の最適化関数により求められた M_{ij} を利用して、ターゲットベクトルを求めるための M'_{ij} を以下のように求める。 $v_i(q)$ に対して $v_i(i \leq |W_i|)$ 中でコサイン類似度が高い上位 1000 個のベクトル V'_i とする。さらに $v_i(q) M_{ij}$ に対して $v_j(j \leq |W_j|)$ 中でコ

サイン類似度が高い上位 1000 個のベクトル V'_j をとる。この V'_i と V'_j の間の変換行列を直接変換で定義した手法により求めたものが M'_{ij} である。この操作をさらに M'_{ij} を利用して求めたものを M''_{ij} , M''_{ij} を利用して求めたものを M'''_{ij} とする。我々はクエリに対する最終的な変換行列を M'''_{ij} とする。提案手法は [9] の手法の自然な拡張と言える。

2.2.4 分散表現空間間の類似語のランキングづけ

以下に与えられる類似度により, $q = (a, b, c)$ を提案手法により検索し, \mathbf{D} のランキングが結果として出力される。 c における b が a となるような \mathbf{D} は期待される語 d の候補の集合である。 \mathbf{D} 中の d_k の位置は以下で定義される類似度により決定される。

$$\begin{aligned} & \text{Sim}(R_{A \times B}(a, b), R_{C \times D}(c, d_k)) \\ & := \cos(v_{1(j)}(b) \cdot M_{1(j)}, v_{(j)1}(d_k)) \quad (j = 2, 3). \end{aligned} \quad (4)$$

すなわち, $\text{Rank}(q, d_k)$ が $\text{Sim}(R_{A \times B}(a, b), R_{C \times D}(c, d_k))$ で与えられる。

3 分散表現の構成

3.1 New York Times Annotated Corpus

通常, Skip-gram モデルによる分散表現空間は Wikipedia を用いた Glove [13] などの汎用的なエンティティモデルを利用することが多い。しかし, エンティティモデルを使用した場合は, 学習コーパスの Wikipedia が記事に対して時系列情報を含んでいないため, 作成された単語の分散表現が 2000 年代のものか 1980 年代のものか特定できない。我々は New York Times Annotated Corpus ([14], 以後 NYTAC と記す) を分散表現を作成するためのコーパスとして利用した。New York Times Annotated Corpus は 1987 年から 2007 年まで発行された約 180 万記事が収録されている。我々はこのコーパスを後述の理由で 3 分割した。分割されたコーパスのラベルを T_3 : [1987-1993], T_2 : [1994-2000], T_1 : [2001-2007] とする。分割された記事はそれぞれ約 50 万記事, 約 380,000 語彙ずつ収録されている。

我々は, このコーパスから本文のみを抽出する。次にアルファベットと数字以外の文字を全て除去し, アルファベットを全て小文字に変換する。数字は全て 0 に直す。例えば, 1,200 は 0,0 のように変換する。その後, 2-グラムと 3-グラムにより, 頻出する複合 2 語と複合 3 語を部分的にフレーズ処理する。上記の処理が終わった文字列から Bag of Words を作成する。分散表現の次元を 600 にし, 文脈語の窓幅を 40, エポック数を 6 に設定した。分散表現の学習は文単位で行った。パラメータの決定は次節にてのべる。

3.2 パラメータ調整

学習は Skip-gram モデルを使用して分散表現を作成した。誤差逆伝播法は階層的ソフトマックス関数を使用した。分散表現を作成するにあたり, 一般的に考慮されているパラメータを調整するためにその性能を以下の方法で評価した。評価セット

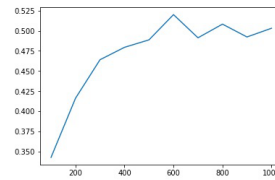


図 2 word vector の次元

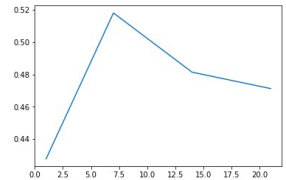


図 3 学習コーパスの学習量

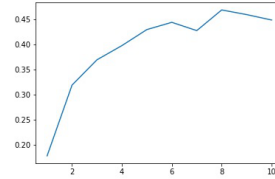


図 4 エポック数

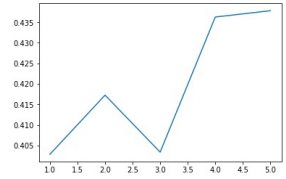


図 5 単語の最小出現頻度

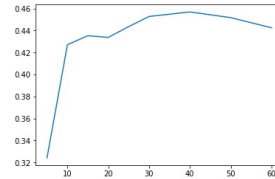


図 6 文脈語の窓幅

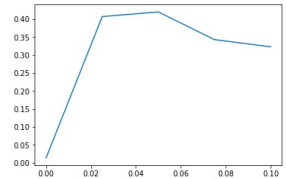


図 7 勾配降下法の学習率

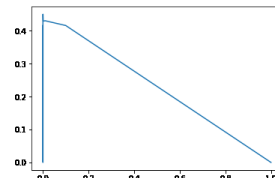


図 8 ダウンサンプリング率

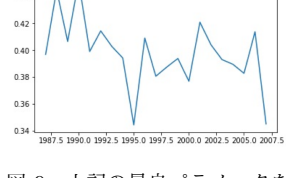


図 9 上記の最良パラメータを使用した場合の各年代のモデルの性能

は [11] で用意されている questions-words.txt¹ の意味カテゴリーを使用した。調整したパラメータは学習コーパスの語彙数, 分散表現の次元, エポック数, 単語の最小出現回数, 文脈語の窓幅, 最適化の学習率と誤差逆伝播法の時の単語のダウンサンプリング率である。分散モデル作成のための入力コーパスの語彙数を意味し, NYTAC の n 年分で表現した。分散表現の次元は出力される分散表現ベクトルの次元を表す。エポック数は順伝播学習 1 回と誤差逆伝播法 1 回を一サイクルとした場合の学習サイクル数を表す。単語の最小出現回数は入力される単語のコーパス中の最小出現回数をさす。ここで設定された数より小さい出現回数の単語は入力として無視される。文脈語の窓幅が l であるとは, Bag of Words を作成するとき注目する単語の前後 l 語を注目する単語の文脈語としてカウントすることを意味している。最適化の学習率とは, 順伝播法の時の学習率をさす。ダウンサンプリング率は誤差逆伝播法の時に重みが下げられる語彙の比率をさす。分散表現の次元を 200 にし, 文脈語の窓幅を 20, エポック数を 5, 学習率を 0.25, ダウンサンプリング率を 0.001 としたモデルを基本モデルとし, 上記のパラメータ一つずつを変動させた際の性能差を観察した。

図 3 が学習コーパスの語彙数を変動させた結果である。学習コーパスの語彙数が 7 年分の時適合率が最大になった。図 7 が

1: <https://code.google.com/archive/p/word2vec/>

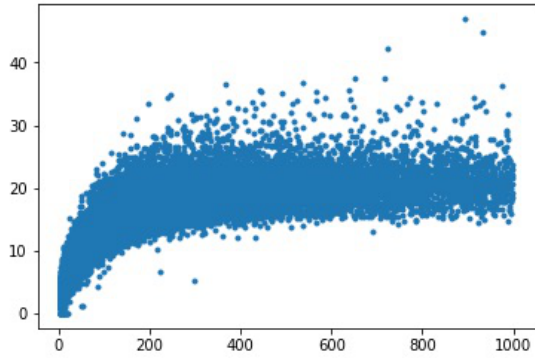


図 10 word2vec の word vector の長さと言語の出現回数の分布

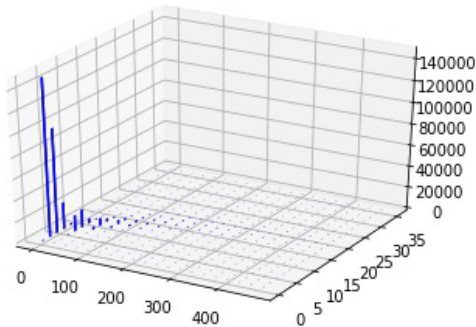


図 11 word2vec の word vector の長さと言語の出現回数の分布上のヒストグラム

表 1 評価セット：遷移類義語のカテゴリ

カテゴリ	リスト数	期間	時間間隔
leader-world	123	1987-2007	1
leader-common-countries	20	1987-2007	1
GDP-Top15	15	1985-2010	5
Sorp-Opera-Top10	10	1987-2007	1
Fortune-500-Top10	10	1987-2007	1
MVP	6	1987-2007	1
Person-of-Year	1	1987-2007	1

文脈語の窓幅を変動させた結果である。文脈語の窓幅が 40 の時適合率が最大になった。図 4 がエポック数を変動させた結果である。エポック数が 8 の時適合率が最大になった。図 8 がダウンサンプリング率を変動させた結果である。ダウンサンプリング率 0.001 の時適合率が最大になった。図 11 が上記で適合率が最大になったパラメータの値で設定した時のモデルの結果である。[11] のモデルは語彙数が 784M の学習コーパスを使用し、56%の適合率である。我々は語彙数が 210M を使用し、([11] のモデルの 30%の学習データ), ほぼ同等の 56% の適合率のモデルを作成した。

4 実験：類義語トラッキング

4.1 学習データおよびテストデータ

学習データは 3.1 節と同じものを使用する。テストデータに関しては以下を使用する。New York Times は政治、経済、

科学技術、スポーツを含む。先行研究 [9] では、グラウンドトゥールズデータが存在しないために人手により、95 組の正解データを作成した。しかし、組を作成した詳細に関する詳細説明が存在しないため、著者はテストデータを再現できなかった。このため、誰でも客観的に使用できるテストデータを使用するのがシステムの性能を見る上で必要なため、上記のトピックに関する類義語のリストを作成した。表 2 に評価セットのカテゴリと仕様を記す。leader-world カテゴリには 123ヶ国の国の指導者のリストを 1987 年から 2007 年までの 21 年間分 Wikipedia から抽出した。1987 から 2007 の期間で各項目を作成した。表 2 の各期間の間隔は 1 または 5 年である。同様にして、leader-common-countries は leader-world のうち NYTAC に頻出する上位 20ヶ国の指導者のリストである。GDP-Top15 は各年の GDP の上位 15 カ国の国名のリストである。Sorp-Opera-Top10 はアメリカで視聴率上位 10 位のテレビドラマの番組名のリストである。Fortune-500-Top10 はフォーチュン誌のフォーチュン 500 (アメリカ国内で売り上げの上位 500 位の企業名のリスト) の上位 10 位までの企業名のリストである。MVP は NYTAC に取り上げられている野球 (MLB), アメリカンフットボール (NFL), バスケットボール (NBA) の MVP プレーヤーの歴代受賞者リストである。Person-of-Year は歴代の New York Times の Person of Year 受賞者名のリストである。評価セットは 162 リスト作成した。各リストには 21 年分のエンティティが含まれるため、3402 個のエンティティが含まれている。我々は、時代によって単語が変化しやすい語義の類義語 (以後、遷移類義語) を対象とした類義語トラッキングを目的としている。したがって、我々は “man” などの時代によって意味が変化しづらい語ではなく、遷移類義語のリストを作成した。

4.2 評価手法

我々は平均逆順位 (MRR) と normalized discounted cumulative gain (nDCG) により性能を評価した。MRR は以下のように定義される。

$$\text{MRR} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{1}{\text{rel}_i}, \quad (5)$$

rel_i は $d_i \in D$ の順位適合性を表す。($B \times D$) に入出力が適合していた場合、 $\text{rel}_i = 1.0$ となり、それ以外の場合は $\text{rel}_i = 0.0$ となる。MRR はすべての $b \in B$ に関して評価する。

nDCG は以下で定義される。

$$\text{nDCG}(\mathbf{D}, r) := \frac{\text{DCG}_r}{\text{IDCG}_r} = \frac{\sum_{i=1}^r \frac{2^{\text{rel}_i^r} - 1}{\log_2(i+1)}}{\sum_{i=1}^{|\mathbf{D}|} \frac{2^{\text{rel}_i^r} - 1}{\log_2(i+1)}}, \quad (6)$$

ここで rel_i^r は D の r の順位 i の取得した順位適合性とする。我々は $r = 10$ とした。($B \times D$) に入出力が適合した場合は $\text{rel}_i^r = 1.0$ であり、それ以外の場合は $\text{rel}_i^r = 0.0$ とする。すべての $d \in D$ に関する $\text{nDCG}(\mathbf{D}, r)$ の平均を最終的な評価指標とした。これらの指標は高い値の時に高い適合性があることを示す。

表 2 遷移類義語のリスト例

年	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
アメリカの大統領	Reagan	Reagan	Reagan	Reagan	Bush	Bush	Bush	Clinton	Clinton	Clinton	Clinton

年	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
アメリカの大統領	Clinton	Clinton	Clinton	Bush	Bush	Bush	Bush	Bush	Bush	Bush

表 3 結 果

手法	学習セット	MRR	nDCG
Direct	Swadesh	0.015	0.074
Sim-Correct	Seadesh	0.025	0.082
Direct	10000	0.22	0.27
Sim-Correct	10000	0.24	0.29
Direct	1000	0.22	0.27
Sim-Correct	1000	0.24	0.29

4.3 実験結果と考察

表 4.3 に実験結果を示す。Direct が [9] の手法であり、Sim-Correct が提案手法である。学習セットとして Swadesh リストを使用した場合、10000 回以上出現した単語のうち二つの時代区分間の共通語、1000 回以上出現した単語のうち二つの時代区分間の共通語とした場合のいずれも提案手法が優位であった。

本手法の得られた結果に関して、以下のように考察する。本手法は MRR が高くなり、nDCG がそれほど変化しない特性を持つことがわかった。これは、正解データの一つが出力された候補ランキングの 1 位に来た場合、1 位の語の類義語が上位に引き上げられたためだと考えられる。したがって、正解データが 1 位に来ない場合には順位が下がる傾向にある。

5 おわりに

我々は、既知情報の例をクエリとして未知の情報を検索するための手法を提案した。例えば、2000 年代、*Bush*、1980 年代という 3 つのエンティティが与えられた時に新聞記事のコーパスから「2000 年代における *Bush*」という関係から「1980 年代における *X*」という関係の *X* にもっとも適切な語を検索するという手法である。提案手法は、語の分散表現空間間の変換行列を求める手法を基礎としている。特に、関係「2000 年代における *Bush*」と関係「1980 年代における *Reagan*」の間の関係に似ていることを利用して変換行列を補正する手法に着目した。提案手法は各時代の分散表現を基底変換する手法を基礎としている。我々は時代間に成立する推移律に着目した。提案手法は評価実験において類義語トラッキングに関してある程度の有意性を示した。この実験により、現在から過去への類義語トラッキングの性質と有用性を明らかにした。

謝 辞

本研究の一部は JSPS 科学研究費助成事業 JP16H02906, JP17H00762, JP16H01756, JP18H03243 による助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] P.D. Turney and M.L. Littman, “Corpus-based learning of analogies and semantic relations,” *Proceedings of Machine Learning*, Vol. 60, Issue 1–3, pp.251–278, 2005.
- [2] M.P. Kato, H. Ohshima, S. Oyama, and K. Tanaka, “Query by analogical example: Relational search using web search engine indices,” *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp.27–36, 2009.
- [3] N.T. Duc, D. Bollegala, and M. Ishizuka, “Cross-language latent relational search between japanese and english languages using a web corpus,” *Proceedings of TALLIP, ACM*, Vol.11, Issue 3, pp.11:1–11:33, 2012.
- [4] Aitchison J., “Language change, progress or decay?,” Cambridge University Press, pp.166–167, 2001.
- [5] W. Labov, “Principles of linguistic change, social factors,” 2010.
- [6] G. Hughes, “Words in time: A social history of the english vocabulary,” Basil Blackwell, pp.253–261, 1988.
- [7] R. Mihalcea and V. Nastase, “Word epoch disambiguation: Finding how words change over time,” *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Vol. 2*, pp.259–263, 2012.
- [8] A. Jatowt and K. Duh, “A framework for analyzing semantic change of words across time,” *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp.229–238, 2014.
- [9] Y. Zhang, A. Jatowt, S. Bhowmick, and K. Tanaka, “Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time,” *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Vol. 1:, pp.645–655, 2015.
- [10] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, “Learning representations by back-propagating errors,” *Neurocomputing: Foundations of Research*, pp.696–699, 1988.
- [11] T. Mikolov, W.-t. Yih, and G. Zweig, “Linguistic regularities in continuous space word representations,” *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.746–751, 2013.
- [12] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” *Advances in Neural Information Processing Systems 26*, pp.3111–3119, 2013.
- [13] J. Pennington, R. Socher, and C.D. Manning, “Glove: Global vectors for word representation,” *Empirical Methods in Natural Language Processing (EMNLP)*, pp.1532–1543, 2014.
- [14] E. Sandhaus, *The New York Times Annotated Corpus*, vol.6, No. 12, Philadelphia, 2008.