

構成要素を考慮した論文の分散表現に基づく類似論文検索

小林 雄太^{1,a)} 松本 裕治^{1,b)}

概要: 科学論文の出版数の急激な増加に伴い、関連研究を絞り込むための論文検索技術の需要が高まっている。その一例として、「目的は異なるが手法は似た論文」といった、構成要素（目的・手法・結果）に基づく類似論文検索が挙げられる。しかし、論文は完全に構成要素毎に構造化されていないため、どのように構成要素の情報を抽出し論文の表現に反映させるかが問題となる。また、どの観点で類似しているかアノテーションされたデータセットが存在せず、検索結果の人手評価にコストがかかる。本研究では、類似する理由が含まれた類似論文検索のためのデータセットを独自に作成し、人手を介さず自動評価が可能な新しい検索タスクを提案する。さらに、論文の構成要素を考慮するための、テキストと引用グラフの両方の情報を用いた論文のベクトル分散表現学習手法を提案する。自然言語処理分野の論文における列挙共引用予測を用いた評価実験の結果、既存の分散表現手法と比べ精度の向上を確認した。

1. はじめに

ある研究分野において、論文の「研究の目的は何か」、「どのような手法が使われているのか」といった情報を網羅的に収集し整理することは、関連研究を概観し最新の研究を追い続けるために必要不可欠である。しかし、科学論文の出版数の急激な増加に伴い、各論文の内容理解には多大な時間と労力を要するため、効率的に関連研究を絞り込むための論文検索・推薦技術が必要となる。本研究では「目的は異なるが手法は似た論文」といった、構成要素（目的・手法・結果）に基づく論文検索システムの実現を目指す。

論文の構成要素に関する先行研究として、論文の各文に対して役割を付与する AZ (Argumentative Zoning) の研究が行われている [1], [2]。AZ を行うためには、アノテーションスキームとそれに基づく各文へのアノテーションが必要となるが、論文を初めとした専門文書への文単位のアノテーションはコストが高く、大規模なデータセットの構築が難しい。一方、セクションレベルの検索を実現するため、論文の各セクションの機能 (citation functionality) のラベルをルールベースで付与する Section tagger が提案されており、今後の課題として機械学習を用いた手法が望まれている [3]。

また、論文をベクトルで表現する先行研究では、主にテキストベースの手法と引用グラフベースの2種類の手法が

挙げられる。テキストベースの手法としては、Sugiyamaらは tf-idf ベースの手法で論文をベクトルで表現し、ユーザーの発表論文リストを元に論文推薦システムを構築した [4]。一方、引用グラフベースの手法では、Tangらはグラフの分散表現学習手法を提案し、論文の分類問題へ適用している [5]。これらの論文解析モデルは、単に本文のテキスト (bag-of-words) または、引用グラフの情報のみを用いており、Mikolovらの提案した単語の分散表現 [6]、Teufelらが提案した引用機能 (citation function) [7] を始めとした、より高次元の情報を使用した手法への拡張が期待できる。

論文検索や推薦の評価方法は、評価データセットの作成が困難であり、人手による関連論文の評価が一般的である [4]。また、類似タスクとして、引用文脈から参考文献の何番目の文献かを当てる擬似的な引用推薦タスクが提案されている [8]。

1.1 研究目的

本研究では構成要素（目的・手法・結果）に基づく論文検索のための、本文と引用グラフの両方の情報を用いた新しい論文の分散表現学習の手法を提案する。具体的には、テキストの分散表現と引用グラフの分散表現を組み合わせることにより、論文の構成要素の情報を反映したベクトル分散表現を得る。最終的に、既存の論文の分散表現と比較を行うため、独自に作成した自然言語処理分野における類似論文データセットを用いることで、類似論文検索タスクを通じて分散表現の自動評価を行う。

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology
a) kobayshi.yuta.kp1@is.naist.jp
b) matsu@is.naist.jp

1.2 貢献

本研究の貢献を以下に示す。

- 類似論文検索のための、自然言語処理分野の論文を対象とした類似論文データセットを作成し、列挙共引用予測タスクを提案した。これにより、従来では困難であった論文の構成要素の観点での類似度の評価が可能となった。
- 論文のテキストと引用グラフの両方の情報を用いた分散表現学習手法について、その性能を評価した。先行研究の、テキストまたは引用グラフのいずれか一方を用いて学習された論文の分散表現に比べ、両方の情報を用いて学習された分散表現が、類似論文検索タスクにおいて優れていることを確認した。

2. 類似論文検索のための列挙共引用予測タスク

本節では、類似論文検索のための新しい列挙共引用予測タスクの概要とデータセットについて述べる。先行研究での論文推薦 [4] や分類タスク [5] の評価法は、人手評価や論文のラベル分類問題による評価が挙げられる。しかし、前者は専門性の高い論文全体のアノテーションを行う必要がありコストが大きく、作成可能なデータ事例数に限りがある。後者は論文を7分野のみで分類しており、テキスト情報も bag-of-words と制限が多く、論文の全テキストを利用する本研究の趣旨と異なる。また、引用予測問題として論文を扱う研究もあるが、単に引用があれば関連しているという仮定に基づいており、「どの観点で」という情報が含まれておらず、「目的は異なるが手法は似た論文」を探すタスクへの応用は難しい [9]。

提案タスクに最も近いタスクとして擬似的な引用推薦タスクが提案されている [8]。これは引用文脈から参考文献のどの文献を参照しているかを当てる分類タスクである。しかし、実応用の場では論文集合は数千数万以上のオーダーであり、数十の参考文献から選択する分類タスクとは性質が異なる。

ある特定の観点での論文の類似性を扱った研究として Eto の共引用解析の研究が挙げられる [10]。彼は共引用関係にある論文の中でも、引用箇所が近ければ近いほど強い関係性があり、特に引用トークンが隣接した列挙共引用において論文の類似度が高くなることを示した。また、Gipp らは同様の共引用を扱った研究として Co-citation Proximity Analysis(CPA) を提唱し、同一文内に存在する共引用は最も強い関係性を持つと主張した [11]。

これらの研究を踏まえ、我々は類似論文検索のための列挙共引用予測タスクを提案する。図1に提案タスクの概略を示す。

論文中の列挙共引用において、その文脈(列挙共引用を含む文)と先頭の参考文献から、2つ目以降の参考文献を

全論文集合から当てるタスクとして定義する。図1では、論文 X 中の列挙共引用において、論文 A,B が引用されている。この場合、2つ目以降の参考文献である論文 B を隠蔽し、引用文脈と先頭の論文 A から隠された論文 B を予測することになる。

この際、全論文のテキストと引用グラフデータが利用可能であるが、列挙共引用で隠蔽された論文へのエッジ(X → B)は除外する。これは、実際に共引用関係にある論文を当てることが研究の趣旨ではなく、列挙共引用されそうな「ある観点で」類似した論文を当てることが趣旨であり、論文 X の参考文献による論文 B の絞り込みを防ぐためである。図1中では、論文 A, B, は単語分散表現を素性として使用した点で類似しているとして X に共引用されているが、論文 D でも同様に論文 A, B, C が共引用されていた場合、グラフ上の赤線で示した論文 A → D → B と辿ることで、論文 A と手法的な観点で類似した論文として、論文 B を予測することが期待できる。

LCP を解く利点を下記に示す。

- 列挙共引用は自動で抽出可能なため、人手によるデータ作成と異なり大量の類似論文セットを収集可能である。
- 引用文脈が共有されているため、「どの観点で」論文が類似しているかを考慮したタスクとなっており、先頭の論文及び引用文脈をクエリとみなした検索タスクとみなせる。
- 全論文集合から参考文献の論文を当てるという、より実応用に近いタスク設定のため、適切な引用を追加で推薦するシステムとして論文執筆やレビューへの応用が期待できる。

本研究では、ACL Anthology*¹ からクローリングされた2016年までの自然言語処理分野の論文 PDF データを対象としデータセットを作成した。PDF データからの xhtml への変換は独自にカスタマイズした Poppler*² を用いた。論文は会議や年代によってフォーマットが異なり、全ての論文の完全な変換は困難なため、下記の条件を満たす合計20496論文を使用する。

- 全セクションを抽出可能。
- セクションの数が3以上。
- 参考文献を抽出可能。

また、正規表現により論文中の引用と列挙共引用を検出し、引用グラフを構築した。この際、構築したデータセットに含まれる論文の引用のみ使用し、外部の論文への引用は除外した。本研究の評価実験では、列挙共引用に応じて引用グラフのエッジを除外するが、全ての列挙共引用について2番目以降の参考文献へのエッジを除外すると引用グラフが大幅に削減されるため、列挙共引用の中でも、1文

*¹ <http://aclweb.org/anthology/>

*² <https://poppler.freedesktop.org/>

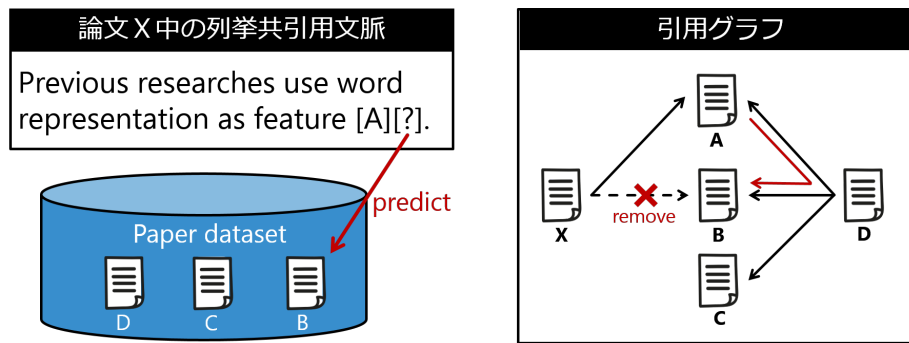


図 1 列挙共引用予測タスクの概要

表 1 抽出した ACL 論文データセットの統計量

項目	統計量
論文数 (ノード数)	20496
引用数 (エッジ数)	259743
平均次数	29.6
列挙共引用の平均参考文献数	2.38

表 2 列挙共引用 1000 例の参考文献数の分布

参考文献数	統計量
2	722
3	210
4	53
5	7
6	1
7+	2
average	2.38

中に一箇所のみ列挙共引用が出現する 34760 例中の 1000 例を評価データセットとして使用する。列挙共引用によるエッジ削除後の、最終的に使用する論文データセットの統計量を表 1 に、列挙共引用 1000 例の参考文献数のヒストグラムを表 2 を示す。列挙共引用の平均参考文献数は 2.38 のため、1 事例につき平均的に 1 つあるいは 2 つの隠蔽された論文を予測することになる。

3. 提案手法

本節では、類似論文検索のための論文ベクトル学習法の概略と各手順について述べる。図 2 に提案手法の概略を示す。手順 1~3 ではテキストからの論文の構成要素の抽出とベクトル化を行い、手順 4 で引用グラフに構成要素と対応したラベルを付与し、手順 5 でテキストとグラフの情報の統合を行う。

3.1 手順 1：論文データセットの前処理

2 節で述べた方法で ACL Anthology から抽出した論文の前処理を行い、各セクションのタイトルとテキスト抽出及び、参考文献情報と引用文脈の抽出によるグラフ構築を行う。また、英語 Wikipedia コーパスと ACL Anthology から抽出した論文のテキストを組み合わせたコーパスを作

成し、教師無しで fastText[12] の単語ベクトル分散表現の学習を行い、以後 fastText を使用する際はこのベクトルを単語ベクトルの初期値として使用する。

3.2 手順 2：NLMCM ルールによるセクションの分類

セクション単位で論文の機能がアノテーションされたデータセットである構造化アブストラクトを用いて、自然言語処理分野の論文のセクションを分類する。

構造化アブストラクトとは、論文の理解を助けるための、ラベル付けされたセクションを持つ要約であり、少なくとも 3 つのラベルを持つ。構造化アブストラクトは医療系論文のデータベースである PubMed に付与されており、National Library of Medicine Category Mappings (NLMCM)^{*3} ファイルは、構造化アブストラクトに使用された 3032 のセクションタイトルを broader NLM Categories (BACKGROUND, OBJECTIVE, METHODS, RESULTS, CONCLUSIONS) の 5 ラベルに対応付けたデータセットである。まず上記の NLMCM のルールに基づいて、論文の各セクションタイトルが NLMCM 中のルールと完全マッチするものについて、5 ラベルへの分類を行う。その後、モデルの単純化のため Nanba らの引用機能の 3 ラベルの定義 (OBJECTIVE, METHODS, OTHERS) に則って、さらに最低限の 3 つに分類する [13]。具体的には、BACKGROUND と OBJECTIVE を OBJECTIVE に、RESULTS と CONCLUSIONS をまとめて OTHERS とする。

例えば、セクションタイトルが「Introducton」の場合、まず、NLMCM ルールに基づいて BACKGROUND に分類された後、3 ラベルの中の OBJECTIVE に分類される。以後この 3 ラベル (OBJECTIVE, METHODS, OTHERS) を各セクションの機能ラベルとして利用する。NLMCM に存在しないセクションタイトルは、次に述べる分類器により分類される。

ルールにより分類されなかったセクションを、ルールによりラベル付けされたセクションのテキストを教師データとして教師ありの分類器によって分類する。分類器には

^{*3} <https://structuredabstracts.nlm.nih.gov/>

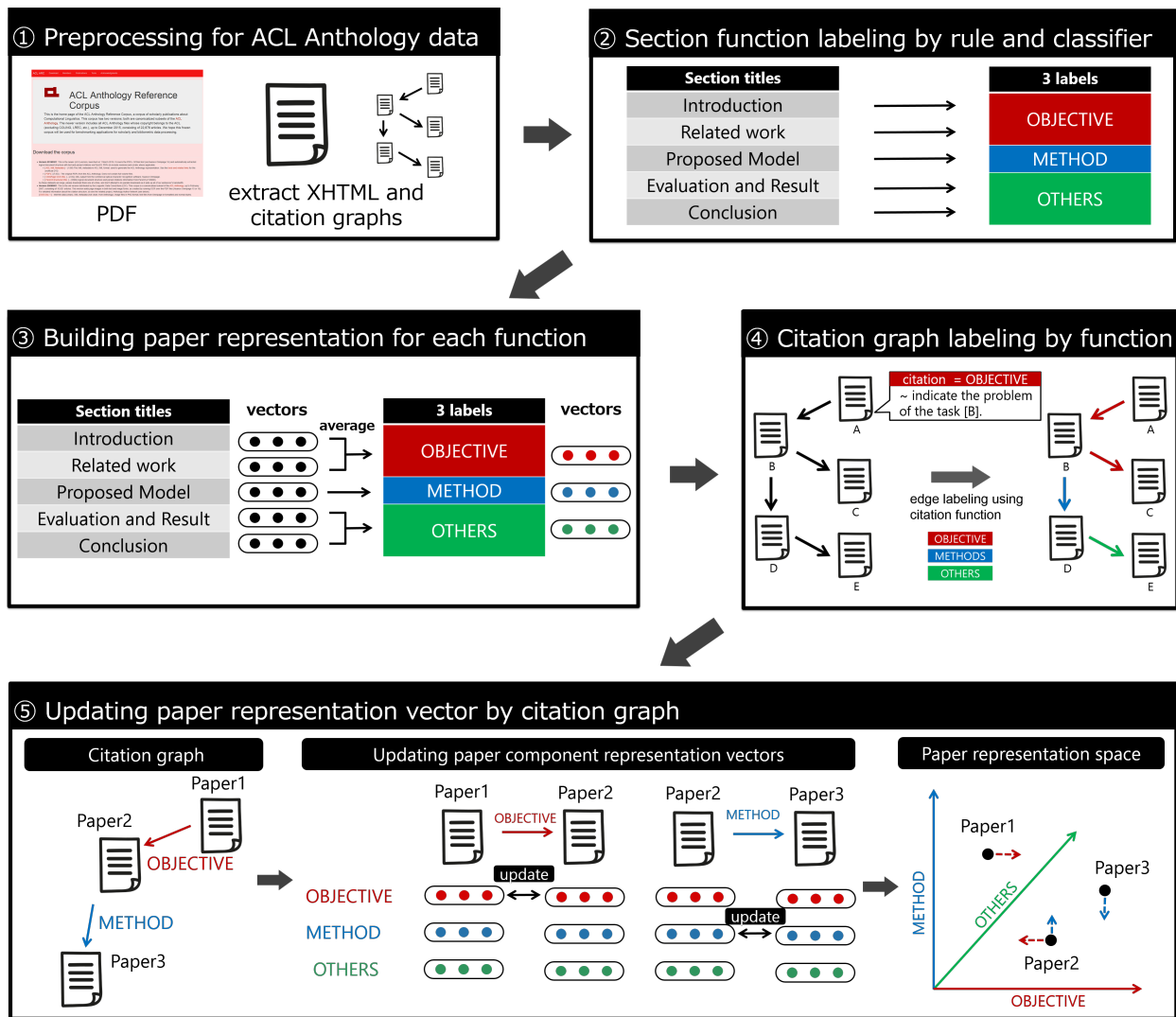


図 2 類似論文検索のための論文ベクトル学習法

文字 n-gram を考慮した単語ベクトルを使用可能で、bag-of-words の情報に加え、単語 n-gram の情報をエンコードした fastText[12] を用いる。入力はセクションの本文テキスト、出力は各セクション機能ラベル (OBJECTIVE, METHODS, OTHERS) となる。fastText は隠れ層を通じてセクション機能ラベル情報を共有する。また、この際に各単語のベクトル表現が学習される。

本研究では、部分文字列を考慮することで論文中の専門語に対応可能な点や、単語 n-gram が語彙構造の認識に有効であるという先行研究に基づいている点 [14]、事前実験において tf-idf や Doc2Vec[15] に比べ精度の高かった点を考慮し、以後 fastText を分類器として使用する。

また、セクション機能ラベルの有効性を検証するため、手順 2 でラベリングされた 72721 のセクションタイトルを用いて 5 分割交差検定を行った。単語ベクトルの次元数は 100、ウィンドウサイズ及びネガティブサンプリングのパラメータはともに 5 とした。

表 3.4 に各 n-gram での交差検定の結果を示す。tri-gram

の場合に 96.1 ポイントの高い精度でセクションをラベル付けすることを確認した。以後 n-gram は最も性能の高かった tri-gram で学習された単語ベクトルを使用する。分類器の精度を確認した後、ルールによってラベル付けされなかった 11118 セクションについて、セクションタイトルと本文を結合したテキストを入力としてラベル付を行った。

3.3 手順 3: 論文のベクトル表現の構築

まず、手順 2 で教師あり学習した fastText の 100 次元の単語ベクトルにより、各セクションのベクトルを各単語ベクトルの平均として計算する。次に、各セクションのベクトルを機能ラベル (OBJECTIVE, METHODS, OTHERS) 毎に平均し、最終的に $100 \times 3 = 300$ 次元のベクトルが得られ、これを論文のベクトル表現とする。ベクトル表現は前述の 3 つの機能ラベル毎に 100 次元を持つ。

3.4 手順 4: 引用グラフのエッジへの引用機能ラベル付与

ここでは、手順 1 で抽出した引用グラフの各エッジに、

表 3 セクション機能ラベルの予測

Type of fastText	Precision
uni-gram	89.6
bi-gram	94.3
tri-gram	96.1

表 4 引用機能ラベルの分布と予測精度

Citation function label	Precision
OBJECTIVE	200
METHOD	1214
OTHERS	204
Total	1618
Accuracy	90.1

対応する引用の文脈を基にした引用機能ラベルを付与する。この引用ラベルに沿って次の手順 5 で論文の構成要素のベクトルが更新される。

引用に関する情報の分類は大きく分けて 3 項目あり、どういった理由で引用したかを表す引用機能 (citation function)[7], 極性 (ポジティブ/ネガティブ) [16] や重要度 [17] が挙げられる。しかし, 引用の極性はネガティブな事例の絶対数が少なく, 重要度に関してはアノテーションデータが少ない。

従って, 本研究では引用機能に関するラベルを使用する。引用のラベルには最低で 3 種類 [13], 多いもので 20 を超えるもの [18] があるが, セクション機能との対応付や, アノテーションデータの収集の容易さから, 引用機能ラベルについてシンプルな 3 種類の Nanba らの先行研究を踏襲する [13]。難波らは他の研究者が提唱する手法等を用いて新しい理論を展開する場合を論説根拠型, 関連研究との比較や既存研究の問題点の指摘を行う場合を問題点指摘型, 上記 2 つにあてはまらないものはその他型とした。

本研究では, 自然言語処理分野の論文の各引用文脈に機能ラベルがアノテーションされた, Teufel らの Citation Function Corpus[7] 及び, 要約タスクのための CL-SciSumm-2017 SharedTask[19] のコーパスを組み合わせたデータセットを使用した。これらのコーパスの部分的なデータはアノテーションスキームの特性上, 手順 3 と同様に引用機能ラベルを本研究の 3 種の引用機能へ一部転用が可能である。引用機能ラベルのアノテーションデータセットのラベル分布, 及び fastText による引用機能ラベル予測の 5 分割交差検定の結果を表 3.4 に示す。

METHOD にラベルが偏っているため, OBJECTIVE や OTHERS のラベル予測精度が低い傾向が見られたが, 交差検定では 90.1 ポイントの精度が得られた。

3.5 手順 5 : 引用グラフによる論文ベクトルのアップデート

引用の機能ラベルが付与されたグラフを用いて, 手順 4 で得られた論文ベクトルを初期値として各構成要素 (100

次元) 毎に更新を行う。具体的には, グラフの分散表現手法である LINE を用いる [5]。LINE は引用グラフ上でランダムウォークを行い, 1 ホップの経路上 (引用を 2 回辿った経路) にあるノードを近傍と考え, それらのノードのベクトルが近くなるように学習を行う。本研究ではグラフに 3 種類の引用機能のラベルが付与されているため, このラベルに沿ってベクトルが伝搬する。

図 2 手順 5 では, 論文 A から論文 B への引用に OBJECTIVE の引用機能があるため, 論文 A と論文 B それぞれの OBJECTIVE の 100 次元ベクトルの内積が大きくなるように更新される。同様に論文 B と論文 C は METHOD の引用機能があるため, METHOD の 100 次元ベクトルが更新される。また, 仮に論文 B から論文 C への引用機能が OBJECTIVE であった場合, 1 ホップの関係にあるため論文 A と論文 C も OBJECTIVE の関係であるとみなし, OBJECTIVE の 100 次元ベクトルの内積が大きくなるようにベクトルの更新が行われる。手順 4 のベクトルを初期値としているため, 次元数は 100, ネガティブサンプリングのパラメータは 5 とした。

これらの手順を踏むことにより, 最終的に類似論文検索のための構成要素を考慮した論文のベクトル分散表現が得られる。

4. 実験

4.1 近傍探索による論文の表現ベクトルの評価

提案手法によって得られた論文の表現ベクトルが従来の表現ベクトル fastText または LINE と比べ有用か, 列挙共引用予測タスクによって検証を行う。具体的には, 表 1 で示した 20496 の論文データセットに対して, 1000 事例分の列挙共引用予測を行う。各事例において列挙共引用の先頭の論文ベクトルからコサイン類似度を基に近傍探索を行い, 上位 100 位に正解の論文が含まれた数とその順位で評価を行った。評価指標として, 順位付け問題の精度評価指標である nDCG (normalized Discounted Cumulated Gain) を用いた。

4.2 ベースライン

ベースラインの表現ベクトルとして, 2 節手順 1 で得られた教師なしの fastText [12] を用いて, 論文を全単語の平均ベクトルとしたものをテキストベースのベースラインとする。また, LINE [5] を引用グラフベースの表現ベクトルとして比較する。ハイパーパラメータは 2 節と同様だが, 論文の各構成要素を持たないため, 両ベクトルの次元は 300 とした。

4.3 提案手法

提案手法は 2 種類のベクトルを用意した。1 つはセクション機能ラベルを予測するように fastText を用いて教師

表 5 論文の表現ベクトルを用いた列挙共引用予測タスクの結果

論文のベクトル表現	nDCG@100
fastText [12]	0.43
LINE [5]	0.46
提案手法 1 (fastText + functions)	0.50
提案手法 2 (fastText + functions + LINE)	0.56

あり学習を行った表現ベクトル (提案手法 1 : fastText + functions).

2つ目は提案手法 1 のベクトルを初期値とし, グラフの分散表現手法である LINE を用いて, 引用グラフの引用機能ラベルにも基づいて更新を行った表現ベクトルである (提案手法 2 : fastText + functions + LINE). 各論文の表現ベクトルは OBJECTIVE, METHODS, OTHERS 各 100 次元のベクトルで表現されており合計 300 次元のベクトルを持つ.

4.4 実験結果

表 4.4 に論文の表現ベクトルを用いた列挙共引用予測タスク結果を示す. 論文の構成要素の情報を使用した提案手法 1 の fastText + functions がベースラインに比べ精度が改善していることが確認できる.

さらにテキスト情報に加え引用グラフの情報を付け加えた提案手法 2 の fastText + functions + LINE はさらに精度の改善が認められた. これらの結果から, テキストの構成要素の情報と引用グラフの情報が科学技術論文の類似度計算に有用であることが確認できた.

5. おわりに

本研究では構成要素 (目的・手法・結果) に基づく類似論文検索のための自動評価タスクとして, 類似する理由が引用文脈に含まれた類似論文検索のためのデータセットを独自に作成し, 列挙共引用予測タスクを提案した. また, 論文のテキストと引用グラフの両方の情報を用いた論文のベクトル分散表現学習手法について, その性能を評価した.

先行研究の, テキストまたは引用グラフのいずれか一方を用いて学習された論文の分散表現に比べ, 両方 (セクション機能, 引用機能) の情報を用いて学習された分散表現が, 類似論文検索タスクにおいて優れていることを確認した.

今後の課題として, 引用文脈の論文中での位置の利用や, ニューラルネットワークを使った End-to-End 型の手法が挙げられる.

参考文献

[1] Guo, Y., Korhonen, A., Liakata, M., Silins, I., Sun, L. and Stenius, U.: Identifying the Information Structure of Scientific Abstracts: An Investigation of Three Different Schemes, *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pp. 99–107 (2010).
[2] Contractor, D., Guo, Y. and Korhonen, A.: Using Argu-

mentative Zones for Extractive Summarization of Scientific Articles, *Proceedings of COLING 2012*, pp. 663–678 (2012).
[3] Kafkas, S., Pi, X., Marinos, N., Morrison, A., McEntyre, J. R. et al.: Section level search functionality in Europe PMC, *Journal of biomedical semantics*, Vol. 6, No. 1, p. 7 (2015).
[4] Sugiyama, K. and Kan, M.-Y.: A comprehensive evaluation of scholarly paper recommendation using potential citation papers, *IJDL*, Vol. 16, No. 2, pp. 91–109 (2015).
[5] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J. and Mei, Q.: LINE: Large-scale Information Network Embedding., *WWW* (2015).
[6] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, pp. 3111–3119 (2013).
[7] Teufel, S., Siddharthan, A. and Tidhar, D.: Automatic classification of citation function, *EMNLP*, pp. 103–110 (2006).
[8] Duma, D. and Klein, E.: Citation Resolution: A method for evaluating context-based citation recommendation systems, *ACL*, pp. 358–363 (2014).
[9] Shibata, N., Kajikawa, Y. and Sakata, I.: Link prediction in citation networks, *Journal of the Association for Information Science and Technology*, Vol. 63, No. 1, pp. 78–85 (2012).
[10] Eto, M.: Spread co-citation relationship as a measure for document retrieval, *Proceedings of the fifth ACM workshop on Research advances in large digital book repositories and complementary media*, pp. 7–8 (2012).
[11] Gipp, B., J. E. ranBeel : Citation Proximity Analysis (CPA) : A New Approach for Identifying Related Work Based on Co-Citation Analysis, *Proceedings of the 12th International Conference on Scientometrics and Informetrics, vol. 1*, pp. 571–575 (2009).
[12] Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T.: Bag of Tricks for Efficient Text Classification, *arXiv preprint arXiv:1607.01759*.
[13] Nanba, H. and Okumura, M.: Towards multi-paper summarization using reference information, *IJCAI*, Vol. 99, pp. 926–931 (1999).
[14] Athar, A.: Sentiment Analysis of Citations using Sentence Structure-Based Features, *Proceedings of the ACL 2011 Student Session*, pp. 81–87 (2011).
[15] Lau, J. H. and Baldwin, T.: An empirical evaluation of doc2vec with practical insights into document embedding generation, *arXiv preprint arXiv:1607.05368* (2016).
[16] Case, D. O. and Higgins, G. M.: How can we investigate citation behavior? A study of reasons for citing literature in communication, *Journal of the Association for Information Science and Technology*, Vol. 51, No. 7, pp. 635–645 (2000).
[17] Valenzuela, M., Ha, V. and Etzioni, O.: Identifying Meaningful Citations, *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015).
[18] Garzone, M. and Mercer, R.: Towards an automated citation classifier, *Advances in Artificial Intelligence*, pp. 337–346 (2000).
[19] Jaidka, K., Chandrasekaran, M. K., Rustagi, S. and Kan, M.-Y.: Overview of the CL-SciSumm 2016 Shared Task, *In Proceedings of Joint Workshop on Bibliometric-enhanced Information Retrieval and NLP for Digital Libraries (BIRNDL 2016)* (2016).