

サプライチェーンデータの言語属性に基づいた埋め込み表現の学習

白石 望[†] 江口 浩二^{††} 水野 貴之^{†††} 高須 淳宏^{†††}

[†] 広島大学情報科学部 〒739-8527 広島県東広島市鏡山 1-4-1

^{††} 広島大学大学院先進理工系科学研究科 〒739-8527 広島県東広島市鏡山 1-4-1

^{†††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†]{b184654,kxeguchi}@hiroshima-u.ac.jp, ^{††}{mizuno,takasu}@nii.ac.jp

あらまし 近年、企業間物流状況を表すサプライチェーンデータを解析する技術の深化が求められている。とりわけ本研究では、サプライチェーンにおける各企業間取引の特徴が複数の語からなるキーフレーズの集合で表現された状況を想定し、そのようなサプライチェーンの言語属性に着目した分析を可能とするようなモデルを開発する。各取引のキーフレーズ集合を語彙空間上で表現することは、その高次元性やスパース性の観点から望ましくない。そこで、各キーフレーズを構成する単語系列に着目した文脈化単語埋め込みモデルについて検討する。より具体的には、(1) 事前学習済み BERT を用いて取引キーフレーズの文脈化単語埋め込み表現ベクトルを得て、(2) 同一取引または同一サプライヤーに対応づけられたキーフレーズ（アンカー、正例）とランダムに抽出されたキーフレーズ（負例）を用いてトリプレット損失に基づいた単語埋め込みモデルの微調整（fine tuning）を行う。現実のサプライチェーンデータを用いた実験により有効性を評価し、提案手法の有効性を示唆する結果を得た。

キーワード 文脈化単語埋め込み、距離学習、サプライチェーン

1 はじめに

近年サプライチェーンにおいては国を超えて様々な企業が複雑に結びついている。そのような状況下で、貿易相手国において災害や障害の発生や政治的緊張などにより貿易困難な事態となった場合に、自企業に限らず、自企業を取り巻く他企業からも間接的な影響を受けることにより、販売状況の悪化につながる可能性がある。このように潜在的なリスクのあるサプライチェーンにおいて各企業の取引の関係を分析することで、そのようなリスクにたいして柔軟に対応し課題の解決に繋がること期待されている。

本論文においては、サプライチェーンにおける各企業間取引の特徴が複数の語からなるキーフレーズの集合で表現された状況を想定し、そのようなサプライチェーンの言語属性に着目した分析を可能とするようなモデルを開発する。企業間の取引はサプライヤー 1 社とコンシューマー 1 社とキーフレーズ集合で構成される。キーフレーズは各企業間の取引に対し複数個紐づけられ、各キーフレーズは 1 単語以上から構成されるとする。各取引を語彙空間上のベクトル（以下、単語ベクトル）として表現するにあたり、その高次元性やスパース性の点から望ましくないものとなっている [6]。そこで、キーフレーズを構成する単語系列に着目した文脈化単語埋め込みモデル [2], [7] を効果的に学習する手法について検討する。

2 関連研究

2.1 単語埋め込み表現

単語埋め込みとは高次元の単語ベクトルを低次元の実ベクトル（以下、単語分散表現または単語埋め込み）で表現する技術のことである。よく用いられるものとしては先行研究でも用いられていた Word2Vec [5] が挙げられる。その中でも skip-gram モデルと CBOW モデルをベースとしたものがよく知られている。CBOW モデルは文脈語から対象語推測することを目的としたニューラルネットワークであり、この CBOW モデルをできるだけ正確な推測ができるように訓練することで単語分散表現を獲得することができる。一方で skip-gram モデルは対象語からその前後の文脈語を予測するモデルとなり CBOW モデルを逆転させたような形となる。しかし、これらの文脈依存なしの手法では、多義語の場合に問題が生じる場合がある。また、文脈依存の単語埋め込み表現ができない。

2.2 文脈化単語埋め込み表現

文脈化単語埋め込み表現とは対象語に対して文脈を考慮して行う単語埋め込み表現である。よく用いられるものとしては BERT [2] が挙げられる。BERT は注視機構に基づく Transformer モデル [8] を用いて自然言語に対して事前学習を行う手法であり、事前に教師ラベルが付与されていない単語系列から文脈を考慮した分散表現を学習することを可能としている。masked language modeling タスクと next sentence prediction タスクを仮定した事前学習によって、教師ラベルなしに単語系列にお

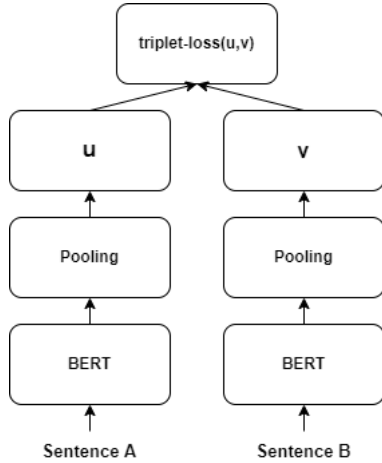


図 1 Sentence BERT の学習

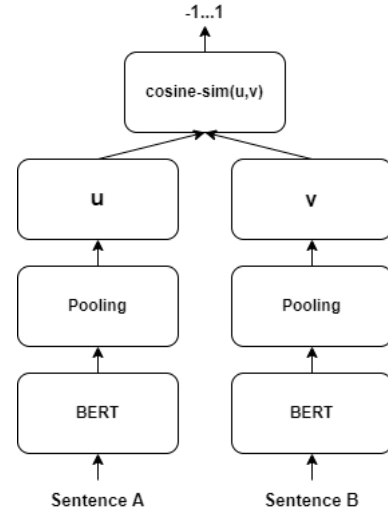


図 2 Sentence BERT による推論

けるパターンを学習し、それを微調整（fine tuning）することにより転移学習を容易に実現できるという利点を有する。

2.3 文レベルの埋め込み表現

文レベルの埋め込み表現を学習することにより、文を単位とした自然言語処理タスクを実現することが可能になる。これらのタスクには、意味論的類似性比較、クラスタリング、意味に基づく情報検索などが含まれる。Sentence BERT [7] では事前学習済みの BERT モデル [2] の出力にプーリング層を追加し、トリプレット損失（triplet loss）[1] に基づいて互に関連した文（アンカーと正例）を単語埋め込み空間上で近くに配置し互に関連しない文（アンカーと負例）を遠くに配置するよう、距離学習（metric learning）に基づいてパラメータの微調整（fine tuning）を行っている。図 1, 2 にそれぞれ Sentence BERT の学習と推論のためのモデル構成を示す。学習において用いるトリプレット損失を以下に示す。

$$\begin{aligned} d_+ &= \|q - k_+\| \\ d_- &= \|q - k_-\| \\ \text{loss} &= \max(d_+ - d_- + \epsilon, 0) \end{aligned}$$

学習では、基準となるアンカー文 q 、正の文 k_+ 、負の文 k_- の 3 つを 1 組として行われる。埋め込み空間で、アンカー文と正の文の距離 d_+ 、アンカー文と負の文の距離 d_- を計測する。マージン ϵ は k_+ が k_- よりも q に少なくとも近いことを保証するものである。最終的には $d_+ + \epsilon \leq d_-$ を満たすように最適化する。

距離の指標としてはコサイン類似度を用い、4 章で述べる実験ではマージン ϵ を 1.0 とした。ただし、これにはアンカー文、それと関連した文（正例文）、関連しない文（負例文）の組が所与であることが想定される。本論文では、Sentence BERT から着想を得て、サプライチェーンに付与されたキーフレーズ属性の埋め込みのため、サプライチェーンの構造を利用した距離学習による微調整を行う手法を提案する。

3 キーフレーズの埋め込み表現学習

本論文では、テキストで表現された取引属性を伴う企業間取引データを対象とした分析を行う。Sentence BERT [7] から着想を得て、サプライチェーンに付与されたキーフレーズ属性の埋め込みのため、サプライチェーンの構造を利用した距離学習による微調整（fine tuning）を行う手法を提案する。

各取引はサプライヤー 1 社とコンシューマー 1 社とキーフレーズの集合で表現されると仮定する。キーフレーズ集合は企業間取引の内容を示し、取引毎に複数のキーフレーズが紐づけられる。また、各キーフレーズは 1 単語以上で表現される。

まず、企業を頂点、企業間取引を有向辺としたグラフ（以下、サプライチェーンネットワーク）を構築する。また、企業間取引毎のキーフレーズ集合を抽出する。

次に、キーフレーズの埋め込み表現学習を考える。前章で述べた Sentence BERT と同様に、トリプレット損失に基づいて互に関連したキーフレーズ（アンカーと正例）を埋め込み空間上で近くに配置し、互いに類似しないキーフレーズ（アンカーと負例）を遠くに配置するような距離学習に基づく距離学習を行う。ただし、この距離学習にはアンカーキーフレーズ、それと関連した文（正例キーフレーズ）、関連しない文（負例キーフレーズ）の組が学習データとして必要となるが、これらが明示的に与えられるような状況は一般的でない。そこで以下に示すような距離学習を考える。

(1) 事前学習済み BERT を用いて、取引キーフレーズの文脈化単語埋め込み表現ベクトルを得る。

(2) 次に示す 2 つの条件それぞれについてトリプレット損失に基づいた単語埋め込みモデルの微調整（fine tuning）を行う。

(a) 同一取引に対応づけられたキーフレーズ（アンカー、正例）とランダムに抽出されたキーフレーズ（負例）を用いる（以下、**非緩和条件**と呼ぶ）。

(b) もう一つの条件として、上に述べた条件を緩和させて、

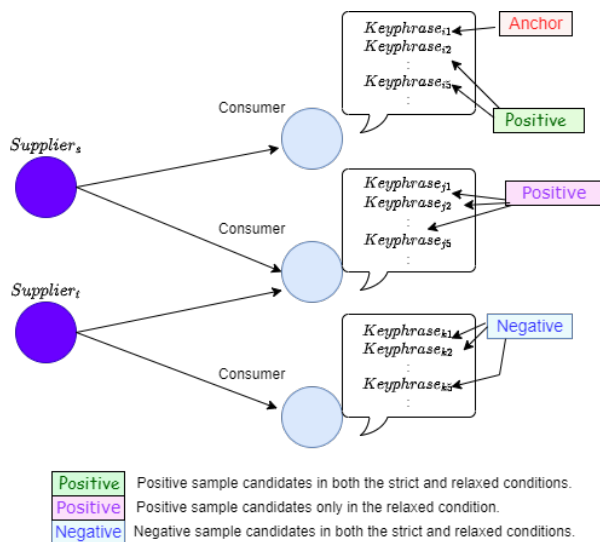


図 3 学習データのサンプリング手法 (a) および (b)

同一の供給企業に対応づけられたキーフレーズ（アンカー、正例）とランダムに抽出されたキーフレーズ（負例）を用いる（以下、緩和条件と呼ぶ）。

本節の残りの部分で、上記の非緩和条件と緩和条件の詳細について述べる。

3.1 非緩和条件

上述の非緩和条件について例を用いて説明する。取引 $Trade_i$ に対して図 3,4 のようにキーフレーズが付与されているとする。取引 $Trade_i$ に着目すると、アンカーキーフレーズを $Keyphrase_{i1}$ としたときの正例キーフレーズは $Keyphrase_{i2}, Keyphrase_{i3}, \dots$ となり、アンカーキーフレーズを $Keyphrase_{i2}$ としたときの正例キーフレーズは $Keyphrase_{i1}, Keyphrase_{i3}, \dots$ となり、アンカーキーフレーズを $Keyphrase_{i3}$ としたときの正例キーフレーズは $Keyphrase_{i1}, Keyphrase_{i2}, \dots$ となる。また、それぞれにおいて負例キーフレーズは $Trade_i$ 以外に対応づけられたキーフレーズからランダムに抽出する（例、 $Keyphrase_{j2}$ ）。このように、非緩和条件ではアンカーキーフレーズと正例キーフレーズの対は、同一の取引に対応づけられたキーフレーズ集合から抽出したものであるため、互いに強く関連した内容のものであると期待される反面、サンプル数は限定される。

3.2 緩和条件

次に、緩和条件について例を用いて説明する。サプライヤ企業 $Supplier_s$ による任意の取引に対して以下のようにキーフレーズが付与されているとする。サプライヤ $Supplier_s$ に着目すると、アンカーキーフレーズを $Keyphrase_{i1}$ としたときの正例キーフレーズは $Keyphrase_{i2}, Keyphrase_{i3}, Keyphrase_{j1}, Keyphrase_{j2}, \dots$ となり、負例キーフレーズは $Supplier_s$ 以外に対応づけられたキーフレーズからランダムに抽出する（例、 $Keyphrase_{k2}$ ）。このように、緩和条件ではアンカーキーフレーズと正例キーフレーズ

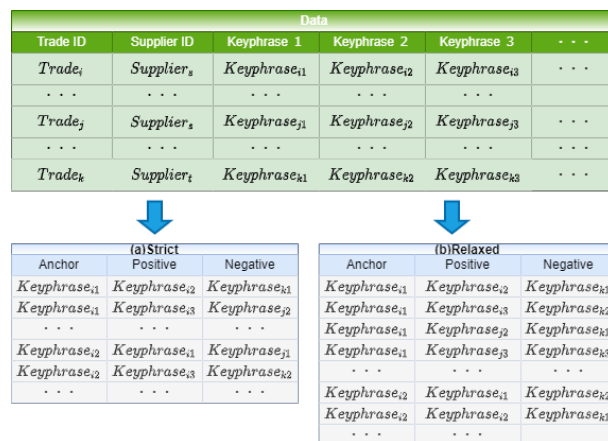


図 4 サンプリング手法 (a) および (b) における正例と負例

ズの対は、同一サプライヤによる任意の取引に対応づけられたキーフレーズ集合から抽出したものであるために、非緩和条件ほどではないかもしれないものの一定の関連性を維持できると考えられる。つまり、同一のサプライヤによる複数の取引内容に関して均質性を仮定しており、この仮定が常に成り立つとは限らないが多くの場合に有効であると考えられ、それにより学習サンプル数を増やすことの効果をねらったものである。

4 評価実験

4.1 データセット

実験のため、FactSet Revere LiveData¹のサブセットである「Feed Supply Chain Relationships」データセットを用いた。本データセットは企業間の取引データとして、製品を供給するサプライヤとその顧客であるカスタマ間の取引について、取引開始日、取引終了日、サプライヤ企業、カスタマ企業、取引の内容を表すキーフレーズ（最大 10 フレーズ）などで構成される。この企業間取引データセットから 2014 年 1 月 1 日の時点で取引中であるものを抽出し、さらに取引の内容を表すキーフレーズが全く付与されていない取引のデータを取り除いた。その結果、データセットに含まれる取引数は 65,438 件であった。

4.2 評価方法と結果

前節で述べたようにして得られたキーフレーズ埋め込みの性能を検証するため、予め held-out したアンカーと正例と負例が埋め込み空間上で適切な距離関係に配置されているかどうかに着目して評価する。held-out セットはランダムに 1 割を抽出し、残りを訓練セットとしてモデルの学習に、held-out セットをモデルの評価に用いた。この一連の工程は 5 通り行った。

比較対象として、キーフレーズ毎に事前学習済み BERT によって得た文脈化単語埋め込みベクトルの平均を用いる。held-out セットにおいてアンカーと正例の間のコサイン類似度がアン

1 : <https://www.factset.com/>

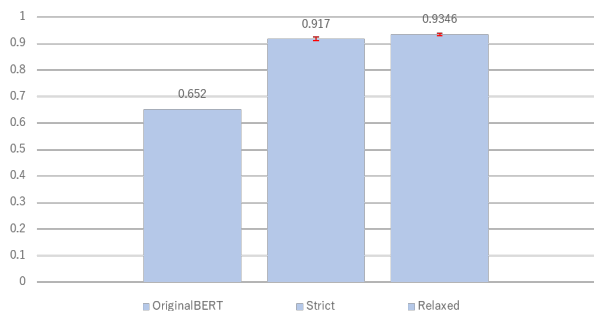


図 5 成 功 率

カーと負例の間のコサイン類似度よりも高いケースを成功と見なし、それが全体を占める割合を成功率（success rate）とし、その結果を図 5 に示す。これらは 5 通り抽出した held-out セットのそれぞれに対する評価結果である。また、これらに対応した正例・負例間のコサイン類似度の平均差を図 6 に示す。これらの図より、提案手法は比較対象の事前学習済み BERT よりもキーフレーズ埋め込みの性能が高く、提案手法でもとりわけ緩和条件の性能が高かったことが確認できる。これは緩和された条件を適用することにより、アンカー、正例、負例からなる学習サンプルの数が増えたことによるものと考えられる。

5 おわりに

サプライチェーンの言語属性に着目する埋め込み学習に関する研究は、著者らの知る限り他に類を見ない。取引内容を示すキーフレーズの高次元性、スパース性、ならびに、同義性と多義性に起因する問題のため、キーフレーズをそのまま用いても期待した効果が得られにくいと考えられる。本論文では、各キーフレーズにおける単語の文脈を考慮しつつ、サプライチェーンの構造に基づいた取引内容の関連性を仮定して構成した疑似的な訓練データを利用してキーフレーズの埋め込み学習を行う手法を提案した。キーフレーズの埋め込みが適切に行われているかどうかに着目した評価実験により、提案手法の有効性が確認された。また評価実験から、学習の精度は訓練データサンプルの数に起因することも明らかとなり、サプライチェーンの構造を考慮したキーフレーズ埋め込みにより従来の BERT より精度が上がることも分かった。キーフレーズの埋め込みを利用したサプライチェーンの予測分析とその評価が今後の課題として挙げられる。また、対照学習（Contrastive Learning）[3], [4] によるキーフレーズ埋め込みについての比較検討は今後の課題である。さらに、提案手法を学術コンテンツや画像・映像データに付与されたキーフレーズ集合に適用することも今後の展開として考えられる。

謝 辞

本研究の一部は 2021 年度国立情報学研究所公募型共同研究（21FS01）の援助による。

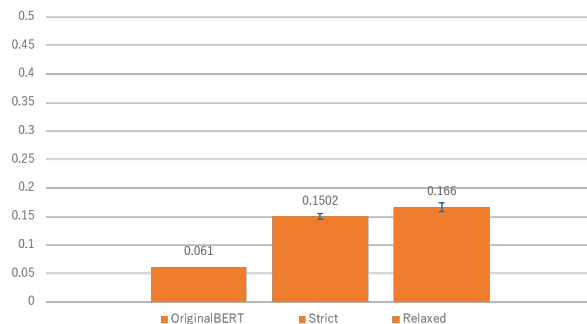


図 6 正例・負例間のコサイン類似度の平均差

文 献

- [1] Gal Chechik, Varun Sharma, Uri Shalit, and Samy Bengio. Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, 11(3), 2010.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186, 2019.
- [3] Michael U. Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010.
- [4] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A survey on contrastive self-supervised learning. *arXiv preprint arXiv:2011.00362*, 2021.
- [5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [6] Mohammad Taher Pilehvar and Jose Camacho-Collados. Embeddings in natural language processing: Theory and advances in vector representations of meaning. *Synthesis Lectures on Human Language Technologies*, 13(4):1–175, 2020.
- [7] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.