

意味と構造の構成演算と類似度学習における非線形性

椿 真史^{1,a)} Kevin Duh^{1,b)} 新保 仁^{1,c)} 松本 裕治^{1,d)}

概要 :

本稿で我々は、自然言語処理における単語ベクトル空間から、文の構成に伴って生じる新たな意味表現空間の類似度学習に焦点を当てる。この際、カーネルを用いた非線形類似度学習手法が効果的であることを示す。本研究の主な貢献は、単語ベクトルから文の意味を構成する演算と、カーネルを用いた非線形の類似度学習とを融合させた点である。実験結果は、ベクトル空間において単語からより複雑な文の意味を構成する際、単語が表現される空間とは異なる、適切な高次元の意味空間の学習が必要不可欠であり、カーネルを用いた非線形手法が重要な役割を果たすことを示唆している。

キーワード : 単語ベクトル空間モデル, 文の意味構成, カーネル法, 非線形類似度学習, 表現学習

1. はじめに

自然言語処理研究において、単語が持つ意味的な情報をベクトルによって表現する手法は古くから存在し、潜在意味解析 (Latent Semantic Analysis : LSA) [10], [24] から、近年ではニューラル言語モデル (Neural Language Model : NLM) [6], [9] に至るまで、様々な手法が研究され続けている。これにより単語間の意味的な類似度を、ベクトル空間における距離や内積から計算することができるため、様々なアプリケーションにおいて幅広く利用されている。そして特に近年では、単語ベクトル空間の自然な拡張として、意味の構成性 [13] の研究が盛んに行われている。この研究では、複数の単語から構成される句や文の意味を適切に表現するベクトルを、単語ベクトルの何らかの構成演算によって得ることが目標となる [4], [12], [18]。この研究分野は、特に近年大流行している Deep Learning (深層学習) の分野において、非常に活発に研究されている [21], [22], [23]。

しかしながら、これら Deep Learning を用いた意味の構成性の研究は、まったく異なる側面から捉え直すことが可能である。我々は Deep Learning における階層的な非線形性ではなく、カーネル法による無限次元空間を用いた非線形性手法に着目する。我々の研究は、以下の2つのフェー

ズに分けることができる。

- (1) 単語から句、句から文の意味表現を、それが持つ構造を考慮して計算する。
- (2) 2つの文が持つ構造と意味の双方を考慮して、その類似度を計算し学習する。

本稿では特に (2) の類似度学習に焦点を当て、単語ベクトル空間モデルにおける意味の構成性とカーネル法による非線形類似度学習という、2つのアプローチの融合と拡張を試みる。図1に我々提案する手法とモチベーションの概略図を示す。本稿の貢献は、以下の2点である。

- (1) 単語ベクトル空間から文の意味を構成する際、カーネルを用いた非線形類似度学習手法が有効であることを示した。これは、より複雑な意味構成に高次元空間が必要であることを示唆している。
- (2) 提案する非線形類似度学習手法によって、文の意味的類似度の評価データセットに対して、ベースラインを上回る性能の向上を確認した。我々の手法は、多くの既存研究と比較して非常に簡潔な手法でありながら、最高性能に迫る性能を達成した。

2. 背景

2.1 計量, 距離, あるいは類似度学習

計量あるいは距離学習 (以下, 計量距離学習) と呼ばれる分野では主に、データ間の距離をタスクに合わせて処理しやすい距離へと変換する行列を学習する。例えば、教師

¹ 情報処理学会

IPSSJ, Chiyoda, Tokyo 101-0062, Japan

^{†1} 現在, 奈良先端科学技術大学院大学

Presently with Nara Institute of Science and Technology

a) masashi-t@is.naist.jp

b) kevinduh@is.naist.jp

c) shimbo@is.naist.jp

d) matsu@is.naist.jp

あり計量距離学習では、同じラベルを持つデータ間の距離は近く、異なるラベルを持つデータ間の距離は遠くなるように、ベクトルデータ自体の変換を行う [14], [25]. 多くの手法はデータ間のマハラノビス距離

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{M} (\mathbf{x} - \mathbf{y})} \quad (1)$$

を最適化するもので、各々の手法において設定した目的関数に対して、最適な変換行列 M を学習する. また一方で、類似度学習においても同様のモチベーションで、距離ではなく主に内積

$$K_M(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{M} \mathbf{y} \quad (2)$$

を最適化する行列 M を学習する [5], [8], [19]. 近年ではさらに、Non-linear Similarity Learning (非線形類似度学習) が提案され、カーネル関数 $K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})$ による写像後に、写像された高次元空間のユークリッド距離 $d_\phi(\mathbf{x}, \mathbf{y}) = \|\phi(\mathbf{x}) - \phi(\mathbf{y})\|_2$ 学習するなどの手法も存在する [16]. これらの手法によって、分類やクラスタリングが困難であった元のデータ点から、より適切に処理可能なベクトル空間を得ることができる.

2.2 文の意味的類似度の評価

特に近年における自然言語処理の意味研究では、任意の長さの文の意味的類似度を評価する Sentences Involving Compositional Knowledge (以下 SICK) [17] と呼ばれるタスクが、SemEval の Shared Task おいて大きな注目を集めている. これらのデータセットは、2つの文の意味的な類似度を人手でスコアリングしたもの (表1を参照) から構成されている. 表1を見るとわかるようにこのデータセットには、ほぼ同じ単語と同じ構文からなる文でも類似度がまったく異なっていたり、あるいは異なる単語と異なる構文からなる文でも同一の意味を持つなど、様々な種類の文が存在する. このタスクは、同義文の検索や抽出、さらには含意関係認識や質問応答などのアプリケーションにとって非常に重要となる分野である. このタスクのより詳細な記述については [1] を参照されたい.

3. 提案手法

この章で我々はまず、幾つかの文の意味ベクトルの計算手法を述べる. そして、それらのベクトルに対する類似度学習手法をカーネルを用いて非線形手法へと拡張する. ここでの訓練データは、 $\{(S_1, S'_1), y_1\}, \dots, \{(S_n, S'_n), y_n\} \in (\mathbb{R}^d, \mathbb{R}^d) \times C$ の形式で与えられるものとする (2章参照). 我々はまず、2つの文 S のベクトル \mathbf{x} を計算し (3.1節)、それらの類似度である連続値 $C = [-1, +1]$ の値を予測するための学習手法 (3.2節) を提案する. 本稿では特に、3.2節で述べる類似度学習手法をカーネル法を用いて拡張し、その非線形性を議論することに焦点を当てる.

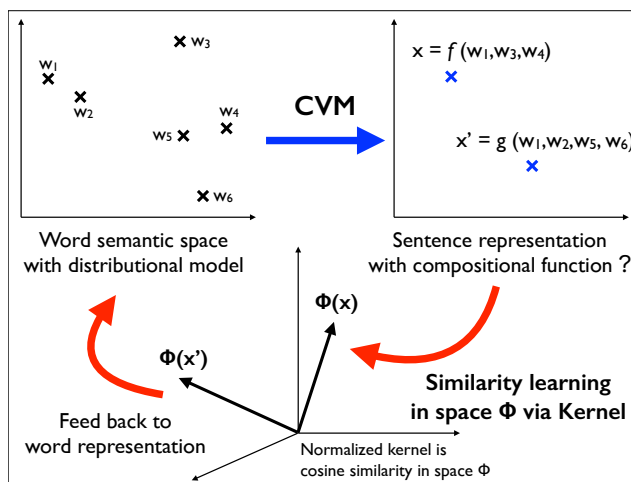


図1 本稿で提案する手法全体の概略図. 正規化されたカーネル関数 K は、高次元空間 ϕ におけるコサイン類似度となる. これを学習することで、元の単語ベクトル空間における意味表現が持ち得ない、文の意味表現構成力の高い単語ベクトルを学習することが可能となる. これを我々は非線形類似度学習と呼ぶ.

3.1 文の意味ベクトルの構成演算

この節では、文の意味ベクトルを計算するための構成演算について述べる. 構成演算である関数 f を $f: S \rightarrow \mathbb{R}^d$ とする. これは文 S を n 次元ベクトル \mathbf{x} へ写像する関数である. 我々はまず、最もシンプルな構成演算 f として、 $f_{ADD}(S) = \sum_{w \in S} \mathbf{d}(w)$ を考える. ここで $\mathbf{d}(w)$ は n 次元の単語ベクトル表現とする. これらのモデルは、文を単語のベクトルの単純な和で表現する Bag-Of-Words に基づくものであり、文が持つ語順や係り受け、さらには階層的な木構造等の情報は一切考慮されておらず、すべての文ベクトル構成演算のベースラインとなる.

次のモデルとして我々は、語順の情報を保持する文ベクトルの構成演算モデルとして、以下の Bi-gram モデルを考える.

$$f_{BI}(S) = \sum_{i=1}^{|S|} \tanh(\mathbf{d}(w_{i-1}); \mathbf{d}(w_i)) \quad (3)$$

ここで、 $;$ は2つのベクトルの連結を、 $|S|$ は文 S の長さを表す. 上記のようなベクトルの連結と非線形演算を適用することによって、前述の Bag-Of-Words では捉えられない語順の情報を保持し、文ベクトル表現を得ることが可能である.

さらにこのモデルは、Bi-gram という文の系列において連続した2単語の組み合わせだけでなく、係り受けあるいは述語項の関係にある単語間の演算にも容易に拡張できる. 我々はこれを Dependency モデルとし、以下のように計算する.

$$f_{DEP}(S) = \sum_{(i,j) \in D} \tanh(\mathbf{d}(w_i); \mathbf{d}(w_j)) \quad (4)$$

ここで w_i と w_j は係り受けの関係にある2つの単語を表

文 A と文 B	人手による類似度スコア
A : <i>A man is jumping into an empty pool.</i> B : <i>There is no biker jumping in the air.</i>	1.6
A : <i>Two children are lying in the snow and are making snow angels.</i> B : <i>Two angels are making snow on the lying children.</i>	2.9
A : <i>The young boys are playing outdoors and the man is smiling nearby.</i> B : <i>There is no boy playing outdoors and there is no man smiling.</i>	3.6
A : <i>A person in a black jacket is doing tricks on a motorbike.</i> B : <i>A man in a black jacket is doing tricks on a motorbike.</i>	4.9

表 1 Sentences Involving Compositional Knowledge (SICK) のデータセットに含まれる文と、それらに対する人手による類似度スコアの一例。データ自体は Amazon Mechanical Turk で複数のアノテータらによって付けられ、スコアはそれらの平均値となっている。類似度スコアは 1.0 から 5.0 となっており、値が高いほど意味的類似度が高いことを示している。訓練データとテストデータは各々 5000 文対からなる、大規模なデータセットとなっている。

し、 $|D|$ は係り受けの関係にある単語の集合の数である。前述の Bi-gram のモデルでは語順の情報を保持する一方、Dependency モデルではより意味的な情報を反映した形で、文ベクトル表現を得ることができる。

3.2 非線形類似度学習

まず我々は、最も基本となる類似度（カーネル）関数 $K(\mathbf{x}, \mathbf{x}')$ に、正規化された内積であるコサイン類似度を用いる。

$$\cos(\mathbf{x}, \mathbf{x}') = \frac{\mathbf{x}^T \mathbf{x}'}{\sqrt{\mathbf{x}^T \mathbf{x}} \sqrt{\mathbf{x}'^T \mathbf{x}'}} \quad (5)$$

このコサイン類似度は、単語ベクトル空間における意味的類似度として非常に幅広く用いられるものである。本稿では、以下に述べるすべてのカーネル関数 K についても、式内に現れる内積はすべて正規化し、またカーネル関数 K 自体も以下のように正規化する。正規化されたカーネルは写像された高次元空間 ϕ におけるコサイン類似度と等価となり得るため、以下のように表現できる。

$$\cos(\phi(\mathbf{x}), \phi(\mathbf{x}')) = \frac{K(\mathbf{x}, \mathbf{x}')}{\sqrt{K(\mathbf{x}, \mathbf{x})} \sqrt{K(\mathbf{x}', \mathbf{x}')}} \quad (6)$$

我々は、高次元空間に写像されたベクトルを陽に計算することなく、カーネルを通して類似度のみを計算し、その後以下に述べるような学習手法を適用することで、高次元空間に写像され得る適切な単語ベクトル表現を新たに学習する。図 1 は我々の提案手法におけるモチベーションと概略図である。

本稿では、以下の 2 つのカーネルを用いる。

$$K_{poly}(\mathbf{x}, \mathbf{x}') = (c + \cos(\mathbf{x}, \mathbf{x}'))^p \quad (7)$$

$$K_{gaus}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1 - \cos(\mathbf{x}, \mathbf{x}')}{\sigma^2}\right) \quad (8)$$

ここで式 (8) において、一般的に用いられるガウシアンカーネルではユークリッド距離を用いるが、これは

$\|\mathbf{x} - \mathbf{x}'\|^2 = \mathbf{x}^T \mathbf{x} + \mathbf{x}'^T \mathbf{x}' - 2\mathbf{x}^T \mathbf{x}'$ と展開することができ、そしてここに現れる内積を任意のカーネルに置き換えることが可能であるため、それをコサイン類似度に置き換えることで $\|\mathbf{x} - \mathbf{x}'\|^2 = 2 - 2\cos(\mathbf{x}, \mathbf{x}')$ を得る。そのため本稿でのガウシアンカーネルは、通常のユークリッド距離ベースでない式 (8) の形になることに注意されたい。

最終的なロス関数は以下の通りである。

$$L(\Theta) = \sum_{i=1}^n \frac{1}{2} \{y_i - K(\mathbf{x}_i, \mathbf{x}'_i)\}^2 + \frac{\lambda}{2} \|\Theta\|^2 \quad (9)$$

ここで、 Θ は学習するパラメータの集合であり、文ベクトルに含まれる各々の単語ベクトルとカーネル内のパラメータである。 n は訓練データ $\{(S_i, S'_i), y_i\}_{i=1}^n$ の数である。

4. 評価実験

本稿では、2 章で述べた SICK データセットを用いて実験を行った。評価には、提案する手法によって計算された 2 つの文ベクトル間の類似度と、人手の類似度スコアとのピアソン相関係数を用いる。これは SemEval で公式に用いられる指標であり、これによって Shared Task のランキングを算出している。

単語ベクトル表現には、Collobert ら [9] の SENNA*1 を用いた。SENNA はニューラルネットワークを用いた単語ベクトル表現学習アルゴリズムであり、本稿で用いるのは Wikipedia をコーパスとして学習された、50 次元の単語ベクトルである。比較として、同一次元のランダムなベクトルを初期値に用いても実験を行った。また、文の係り受け関係あるいは述語項構造を抽出するため、Enju*2 を使用した。

非線形類似度学習における単語ベクトルとカーネル内のパラメータの最適化には、Adaptive Gradient [11] を用いた。学習率の初期値は $\alpha = 0.1$ 、正則化項は $\lambda = 10^{-6}$ とし

*1 <http://ronan.collobert.com/senna/>

*2 <http://www.nactem.ac.uk/enju/index.ja.html>

Kernel	r (ADD)	r (BI)	r (DEP)
Cosine	0.740	0.722	0.755
Polynomial (p=2)	0.812	0.785	0.796
Polynomial (p=3)	0.817	0.793	0.796
Polynomial (p=4)	0.815	0.775	0.790
Gaussian	0.815	0.789	0.780

表 2 様々なカーネルを用いて学習した際の相関係数の結果.

Models	r (Rank)
Zhao et al., 2014	0.828 (1)
Jerva et al., 2014	0.827 (2)
Our best model	0.817 (3)
Jimenez et al., 2014	0.804 (4)

表 3 我々の手法と SICK における Shared Task の上位チームの結果との比較.

Kernel	r (NLM)	r (Random)
Polynomial (p=3)	0.817	0.792

表 4 単語ベクトルの初期値について、訓練済みの NLM とランダムの場合での違い.

た. データセットに対してはイテレーションを 100 に統一し実験を行った.

また、比較する既存研究として、SemEval2014 で行われた SICK の上位 3 チーム [7], [15], [26] の結果と比較した. これらの研究における手法や我々の手法との相違点等については、6 章の関連研究で詳細を述べる.

5. 結果と考察

5.1 Linear (cosine) vs. Non-linear (kernels)

表 2 を見るとわかるように、線形であるコサイン類似度よりも、非線形である多項式カーネルやガウシアンカーネルを用いた場合に相関係数の大幅な上昇が見られた. これは、非線形類似度学習が文の意味構成に必要な単語表現の学習に非常に有効であることを示している. カーネル関数によって写像された高次元空間におけるコサイン類似度を学習することを通して、元の単語ベクトル空間とは異なった、文の意味を適切に構成する単語表現を学習することができていると考えられる.

5.2 Bag-Of-Words vs. Bi-gram vs. Dependency

表 2 の通り、最もシンプルなモデルである Bag-Of-Words の ADD が最も良い性能を示した. 特に 3 次の多項式カーネルにおいて、最も高い相関係数 0.817 を示している. この結果は、カーネルを用いた非線形類似度学習が有効であることを示すと同時に、非線形カーネル埋め込みをする以前の段階で、文ベクトル構成に対して、意味と構造の情報をより適切に保持し学習する必要があることを示唆している.

5.3 Ours vs. Competitive Performers

表 3 に、Shared Task での上位のチームと我々の結果を比較して示す. 我々の提案手法では最高性能に達することはできなかったものの、ランキングでは 3 位に位置している. 我々の手法において強調したいのは、シンプルな単語ベクトルの構成演算のみで文の意味を計算し、そしてカーネルを用いた非線形類似度学習のみで、容易に大幅な性能の向上が見られている点である. 比較する既存手法では大量の素性を用いている点を考えても (6 章を参照), 我々の手法はそのような素性を必要としない点において、大きな優位性があると考えられる.

5.4 NLM vs. Random Vectors

表 4 に見るように、ニューラルネットワークによって訓練済みの単語ベクトルの方が、ランダム初期値の単語ベクトルよりも性能の良い結果となった. しかし仮に、単語ベクトルがランダム初期値であったとしても相関係数は 0.8 に迫っており、この結果を見ても非線形類似度学習が非常に有効であることを示している. 単語ベクトル表現については、初期値として事前に Wikipedia などの大規模なコーパスで学習する必要性はないものと考えられる.

6. 関連研究

従来文の意味的類似度の評価手法に対する主なアプローチは、本論文とはまったく異なる. 例えば、2 つの文に含まれる共通の単語や句の数、N-gram や木構造のアライメント、さらには WordNet などの外部リソースや機械翻訳に用いられる評価指標などの様々な素性を考え、それを用いてサポートベクター回帰で学習するものがほとんどである [2], [3], [20]. これらはある程度成功しているものの、単語ベクトル空間と文の構成性に基づいた意味的なアプローチとはなっていない. 近年行われた SemEval 2014 の Shared Task である SICK に関してもほぼ同様であり、単語ベクトル表現とそれを用いた意味構成モデルのみでは、相関係数が 0.7 程度に留まっているという報告がある [17]. 単語ベクトル表現以外には、前述と同様に多くの素性を用いてモデルを構築し、Shared Task の上位にランクするチームの提案手法では、相関係数は 0.8 以上にまで上昇する [7], [15], [26] これらの手法と比較して、我々の手法が顕著に異なる部分は、最もシンプルな単語ベクトルをベースとした意味構成演算のみを用いているところであり、かつ非線形類似度学習が非常に有効であることを示した点である. 我々の知る限り、単語ベクトル空間における意味の構成性と非線形類似度学習とを融合させた研究はなく、本研究がその基礎になると考えている.

7. 結論と今後の課題

本稿で我々は、単語の意味ベクトル空間から文の構成に

伴って生じる新たな意味ベクトル空間の類似度学習に焦点を当てた。そこで我々は、カーネルを用いた非線形類似度学習手法を提案し、その有効性を確認した。これはベクトル空間において、単語からより複雑な文の意味を適切に構成する際、単語が表現される空間とは異なる高次元の空間が必要となり、カーネル法が重要な役割を果たすことを示唆している。また我々の手法は、複数のデータ点から何らかの演算によって新たなデータ点が生成されるような性質を持つベクトル空間において、広く適用可能な最適化手法である。

今後の課題は以下の通りである。

- (1) 文の階層性を考慮するなどして、文の意味ベクトル計算を行う。
- (2) 深層カーネル (Deep Kernel) を用いて、非線形類似度学習手法をより拡張させる。
- (3) 最終的に、構造と意味のデータの双方をカーネル埋め込みによって学習するような、統一的な枠組みを構築する。

参考文献

- [1] Agirre, E., Diab, M., Cer, D. and Gonzalez-Agirre, A.: Semeval-2012 task 6: A pilot on semantic textual similarity, *International Workshop on Semantic Evaluation (SemEval)* (2012).
- [2] Banea, C., Hassan, S., Mohler, M. and Mihalcea, R.: Unt: A supervised synergistic approach to semantic text similarity, *International Workshop on Semantic Evaluation (SemEval)* (2012).
- [3] Bär, D., Biemann, C., Gurevych, I. and Zesch, T.: Ukp: Computing semantic textual similarity by combining multiple content similarity measures, *International Workshop on Semantic Evaluation (SemEval)* (2012).
- [4] Baroni, M., Bernardi, R. and Zamparelli, R.: Frege in space: A Program for Compositional Distributional Semantics, *Linguistic Issues in Language Technologies* (2013).
- [5] Bellet, A., Habrard, A. and Sebban, M.: Similarity learning for provably accurate sparse linear classification, *arXiv preprint arXiv:1206.6476* (2012).
- [6] Bengio, Y., Ducharme, R., Vincent, P. and Janvin, C.: A neural probabilistic language model, *Journal of Machine Learning Research (JMLR)*, Vol. 3, pp. 1137–1155 (2003).
- [7] Bjerva, J., Bos, J., van der Goot, R. and Nissim, M.: The Meaning Factory: Formal Semantics for Recognizing Textual Entailment and Determining Semantic Similarity (2014).
- [8] Chechik, G., Shalit, U., Sharma, V. and Bengio, S.: An online algorithm for large scale image similarity learning, *Advances in Neural Information Processing Systems (NIPS)* (2009).
- [9] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P.: Natural Language Processing (Almost) from Scratch, *Journal of Machine Learning Research (JMLR)*, Vol. 12, pp. 2493–2537 (2011).
- [10] Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R. A.: Indexing by latent semantic analysis, *Journal of the American Society for Information Science (JASIS)*, Vol. 41, No. 6, pp. 391–407 (1990).
- [11] Duchi, J., Hazan, E. and Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization, *The Journal of Machine Learning Research (JAIR)*, Vol. 12, pp. 2121–2159 (2011).
- [12] Erk, K.: Vector Space Models of Word Meaning and Phrase Meaning: A Survey, *Language and Linguistics Compass*, Vol. 6, No. 10, pp. 635–653 (2012).
- [13] Frege, G.: Über Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik*, 100 (1892).
- [14] Goldberger, J., Roweis, S., Hinton, G. and Salakhutdinov, R.: Neighbourhood components analysis, *Advances in Neural Information Processing Systems (NIPS)* (2004).
- [15] Jimenez, S., Duenas, G., Baquero, J., Gelbukh, A., Bátiz, A. J. D. and Mendizábal, A.: UNAL-NLP: Combining soft cardinality features for semantic textual similarity, relatedness and entailment (2014).
- [16] Kedem, D., Tyree, S., Sha, F., Lanckriet, G. R. and Weinberger, K. Q.: Non-linear metric learning, *Advances in Neural Information Processing Systems (NIPS)* (2012).
- [17] Marelli, M., Bentivogli, L., Baroni, M., Bernardi, R., Menini, S. and Zamparelli, R.: SemEval-2014 Task 1: Evaluation of Compositional Distributional Semantic Models on Full Sentences through Semantic Relatedness and Textual Entailment, *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Association for Computational Linguistics (ACL) and Dublin City University (2014).
- [18] Mitchell, J. and Lapata, M.: Composition in Distributional Models of Semantics, *Cognitive Science*, Vol. 34, No. 8, pp. 1388–1439 (2010).
- [19] Qamar, A. M., Gaussier, E., Chevallet, J.-P. and Lim, J. H.: Similarity learning for nearest neighbor classification, *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on* (2008).
- [20] Šarić, F., Glavaš, G., Karan, M., Šnajder, J. and Bašić, B. D.: Takelab: Systems for measuring semantic text similarity, *International Workshop on Semantic Evaluation (SemEval)* (2012).
- [21] Socher, R., Huval, B., Manning, C. D. and Ng, A. Y.: Semantic Compositionality through Recursive Matrix-Vector Spaces, *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (2012).
- [22] Socher, R., Pennington, J., Huang, E. H., Ng, A. Y. and Manning, C. D.: Semi-supervised recursive autoencoders for predicting sentiment distributions, *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2011).
- [23] Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y. and Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank, *Conference on Empirical Methods on Natural Language Processing (EMNLP)* (2013).
- [24] Turney, P. D.: Domain and Function: Dual-Space Model of Semantic Relations and Compositions, *CoRR*, Vol. abs/1309.4035 (2013).
- [25] Xing, E. P., Jordan, M. I., Russell, S. and Ng, A. Y.: Distance metric learning with application to clustering with side-information, *Advances in Neural Information*

Processing Systems (NIPS) (2002).

- [26] Zhao, J., Zhu, T. T. and Lan, M.: ECNU: One Stone Two Birds: Ensemble of Heterogenous Measures for Semantic Relatedness and Textual Entailment (2014).