G-107

語彙関係知識と単語間表現パターンを用いた単語分散表現の改良

Exploiting Lexical Relation Knowledge and Lexical Pattern in Word Embedding Learning

坂田 亘† 柴田 知秀† 黒橋 禎夫†
Wataru Sakata Tomohide Shibata Sadao Kurohashi

1 はじめに

大規模テキストからの単語のベクトル表現の学習が注目を浴びている。単語のベクトル表現は単語分散表現もしくは word embedding と呼ばれ、各単語は数十から数百次元のベクトルで表現される。word embedding は、構文解析 [1] や固有表現解析 [20]、関係解析 [19, 8] など、様々な自然言語処理のタスクの精度向上に寄与している。

word embedding の学習は「文脈が類似した語は類似している」という分布仮説に基づいている。例えば、「dog」と「animal」はともに「bark」「leg」などの語とよく共起しているため、「dog」と「animal」のベクトルが類似しているように学習される。word embedding の学習は word2vec[14] や GloVe[17] などの方法で行われる。

語彙関係知識を利用して word embedding を改善する研究がある ([6] や [2] など)。語彙関係知識としては WordNet, FrameNet, PPDB などが用いられ、それらの研究では、語彙知識において 2 語が関係を持てばそれらのベクトルが近づくような目的関数を用いる。例えば、WordNet で「dog」と「animal」が上位下位関係にあるので、「dog」のベクトルと「animal」のベクトルが近くなるような目的関数を用いる。

また、単語間表現パターンは上位下位関係などの抽出に古くより用いられてきた [9]。例えば、「animal, such as dog...」のように、「animal」と「dog」の間に単語間表現パターン「such as」が現れると、これらの関係は上位下位関係を表していることがわかる。

本研究では語彙関係知識とコーパス中の単語間表現パターンの両方を使って単語分散表現を改良する手法を提案する。本研究で利用する情報を図 1 に示す。まず、コーパスから共起情報を手がかりとして GloVe で単語ベクトルを学習する (図 1 の左)。そして、学習された単語ベクトルを初期値として、語彙関係知識を考慮する目的関数とコーパス中の単語間表現パターンを予測する目的関数を加えることにより単語ベクトルを改良する (図 1 の右)。

学習された単語ベクトルをアナロジータスクで評価し、提案手法の有効性を示すことができた。

†京都大学, Kyoto University

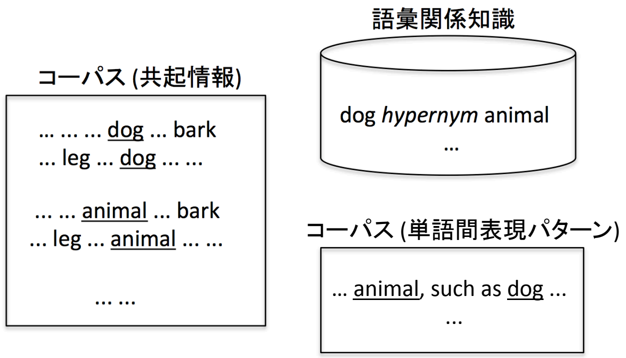

図 1: 本研究で用いる情報

2 関連研究

まず、本研究のもととなる word embedding の学習方法から述べる。word2vec の skip-gram モデルでは単語 w_t から文脈語 w_{t+j} が予測される確率を $p(w_{t+j}|w_t)$ とおき、次式で表される目的関数を最大にする単語ベクトル表現を学習する [14]。

$$\sum_{t=1}^{T} \sum_{-c\le j\le c, j\neq 0} \log p(w_{t+j}|w_t) \tag{1}$$

ここで、T はコーパスの単語数、c は文脈のサイズ (通常、5 単語程度) を表す。$p(w_{t+j}|w_t)$ を定義するために、単語ベクトル $\boldsymbol{v}$ と単語予測ベクトル $\tilde{\boldsymbol{v}}$ の 2 種類の単語ベクトルを導入し、$p(w_{t+j}|w_t)$ は以下のように定義される。

$$p(w_{t+j}|w_t) = \frac{\exp(\boldsymbol{v}_{w_t}^T \tilde{\boldsymbol{v}}_{w_{t+j}})}{\sum_{w\in W} \exp(\boldsymbol{v}_{w_t}^T \tilde{\boldsymbol{v}}_w)} \tag{2}$$

ここで、W はコーパスに出現する語彙の種類数を表す。この式の分母は語彙の種類数分 (通常、数万語) だけ、計算しないといけないため、階層ソフトマックスや負例サンプリングの高速化方法が用いられる。

GloVe は skip-gram よりも広域の共起情報を用いて、word embedding を学習する手法である [17]。次式のような二乗誤差を目的関数として、これの最小化を行う。

$$J = \sum_{i,j=1}^{W} f(X_{i,j})(\boldsymbol{v}_{w_i}^T \tilde{\boldsymbol{v}}_{w_j} + b_i + \tilde{b}_j - \log(X_{i,j}))^2 \quad (3)$$

ここで $X_{i,j}$ は単語 i と j の共起頻度である。文中でウインドウ内に単語 i と j が含まれたときに、単語 i と j の距離を $d_{i,j}$ とすると $\frac{1}{d_{i,j}}$ を共起回数として総和をとったものが共起頻度 $X_{i,j}$ である。b_i, b_j は単語 i, j のバイアス項である。$f(X_{i,j})$ は次式であらわされ、低頻度の共起より高頻度の共起を重視して学習するために用いる。また、$f(X_{i,j})$ は一定以上に高頻度の共起情報は重視しすぎないように $X_{i,j} < X_{max}$ では $f(X_{i,j})$ は一定にする。

$$f(X_{i,j}) = \begin{cases} (\frac{X_{i,j}}{X_{max}})^{\alpha} & (X_{i,j} < X_{max}) \\ 1 & (otherwise) \end{cases} \quad (4)$$

$\boldsymbol{v}$ と $\tilde{\boldsymbol{v}}$ は目的関数において対称なので、ベクトルを学習した後、両者を足すことでオーバーフィッティングやノイズを防ぎ、単語ベクトルを改善することができる。GloVe の方が skip-gram よりも性能がよいと言われていることから、本研究では GloVe をベースラインの word embedding として用いる。

語彙関係知識を利用して word embedding を改善する研究がある [2], [7], [23], [13], [22], [10], [18]。例えば、Faruqui らは Word embedding を学習した後、後処理として語彙関係知識を用いて同義や上位下位関係にある語のベクトルを近くすることにより、word embedding を改善する手法を提案している [6]。また、Bollegala らはコーパスの共起情報から単語ベクトルを学習するモデルにおいて、目的関数に語彙関係知識のうちのある関係にある語ペアのベクトルが近くなる項を加えることにより単語ベクトルの学習を改善している [2]。本研究では語彙関係知識のうちのある関係だけでなく、全ての関係を用いる。

また、単語間表現パターンは古くより用いられており、上位下位関係抽出 [9] や反義語認識 [12] などに利用されている。word embedding の学習に用いるものとして、Bollegala らの研究がある [3]。この研究ではまず、“large Y such as X” と “X is a huge Y” のように類似した意味関係を表すパターン対を、X と Y に出現する単語の類似性を手がかりとしてコーパスから取得する。そして、各パターンのベクトルを X と Y に現れる単語のベクトルの差から計算し、2 つのパターンが類似しているかどうかの判別を目的関数とし、単語ベクトルを学習している。本研究ではそのような 2 ステップを踏まずに、単語間表現パターンを直接的に単語ベクトルの改善に用いる。

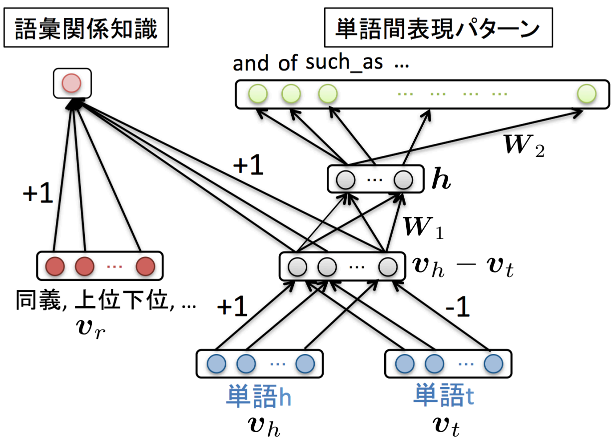

図 2: 提案手法で用いるニューラルネットワーク

3　提案手法

コーパスから学習された word embedding を初期値として、語彙関係知識を考慮する目的関数と、コーパスから得られた 2 語の間のパターンを予測する目的関数を加えることにより、単語分散表現を改善する手法を提案する。図 2 に提案手法で用いるニューラルネットワークを示す。

3.1　語彙関係知識

語彙関係知識には WordNet[16] を用いる。WordNet は人手で構成された大規模なシソーラスである。WordNet は同義の単語がまとめられた synset というグループによって構成されている。各 synset 間には上位下位や反義などの関係を示すリンクが張られている。

WordNet から、2 語 h(head), t(tail) とその間の関係 r を表す三つ組 (h, r, t) の形式で語彙関係知識を取り出す。本研究では Bollegala[2] のデータセットを用いる。表 1 に語彙関係の種類と、それぞれの関係に対して例と数を示す。

語彙関係	例	数
Synonyms	kid↔child	70,018
Antonyms	heaven↔hell	4,064
Hypernyms	computer↔machine	119,029
Hyponyms	machine↔computer	122,926
Member-holonyms	America↔NATO	11,506
Member-meronyms	NATO↔America	11,431
Part-holonyms	eye↔face	13,082
Part-meronyms	face↔eye	13,251

表 1: WordNet に記載されている語彙関係知識

例	単語間表現パターンとその頻度		
insect↔ant	such as:8	(:3	include:3
heaven↔hell	and:14	or:14	&:4
kid↔child	and:72	,:46	club for:43

表 2: 抽出された単語間表現パターンの例

3.2 単語間表現パターンの抽出

コーパス中には単語対の意味関係を表すパターンが存在する。word embedding の学習ではある語とある語が共起することのみを用いているが、それに加えて、ある語とある語がある単語間表現パターンを伴なって出現することを考慮する。

例えば、意味関係を表す単語間表現パターンとして次のようなものが考えられる。

(1) ... insect such as ant ...

(2) ... heaven or hell ...

1 つ目の例では “such as” が「insect」と「ant」の間に上位下位の関係を示唆しており、2 つ目の例では “or” が「heaven」と「hell」の間に反義関係を示唆している。

コーパス中の単語間表現パターンを計数すると表 2 のようなデータを抽出することができる。例えば、表 2 の insect↔ant という単語対について見ると “insect such as ant” という文字列がコーパス内に 8 回、“insect (ant” が 3 回、“insect include ant” が 3 回登場している。なお、本研究ではパターンは 4 語以下のものを利用する。

3.3 語彙関係知識と単語間表現パターンを用いた単語分散表現の改善

語彙関係知識の考慮については Knowledge Base Completion の手法を用いる。Knowledge Base Completion は WordNet や FreeBase などの知識ベースに対して欠けている情報を補うタスクである。本研究ではそのうちの一つの手法である TransE[4] を用いる。

TransE では 2 語 h(head), t(tail) とその間の関係 r を表す三つ組 (h, r, t) が与えられ、以下のスコアが小さくなるように各種 embedding を学習する。

$$score(h, r, t) = |\boldsymbol{v}_h - \boldsymbol{v}_t + \boldsymbol{v}_r|_2 \quad (5)$$

ここで、$\boldsymbol{v}_h$, $\boldsymbol{v}_t$ はそれぞれ語 h, t の word embedding、$\boldsymbol{v}_r$ は関係 r の embedding を表し、すべて k 次元のベクトルとする。

以下のロス関数が小さくなるように各種 embedding を学習する。

$$\sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'_{(h,l,t)}} [\gamma + score(h, r, t) - score(h', r, t')]_+ \quad (6)$$

ここで、$[x]_+$ は x の正の部分を表す。$S'_{(h,r,t)}$ は以下で与えられる疑似負例であり、r, t を固定し、h をランダムに得た h' と置き換えたものと、同様に、h, r を固定し、t をランダムに得た t' と置き換えたものからなる。

$$S'_{(h,r,t)} = \{(h', r, t)|h' \in E\} \cup \{(h, r, t')|t' \in E\} \quad (7)$$

ここで、E は語彙関係知識に含まれる単語の集合を表す。式 (6) の $\gamma(> 0)$ はマージンを表し、正例に対するスコアと疑似負例に対するスコアが γ 以上離れるようにするためのものである。

単語間表現パターンの考慮には語彙関係知識とは異なるニューラルネットワークを構成する。その理由としては語彙関係知識は関係を持つかどうかの 2 値であったが、単語間表現パターンは頻度があるためであり、2 単語のベクトルの差からそれらの 2 語の間に出現する単語間表現パターンの分布を予測するニューラルネットワークを構成する。

システムは以下のように単語間表現パターンの分布の予測ベクトル $\boldsymbol{out}_p$ を計算する。

$$\boldsymbol{h} = f_1(\boldsymbol{W_1}(\boldsymbol{v}_h - \boldsymbol{v}_t)) \quad (8)$$

$$\boldsymbol{out}_p = f_2(\boldsymbol{W}_2\boldsymbol{h}) \quad (9)$$

ここで、$\boldsymbol{h}$ は中間層、$\boldsymbol{W}_1$ は単語ベクトルの次元 × 中間層の次元数の行列、$\boldsymbol{W}_2$ は中間層の次元数 × パターン数の行列である。f_1, f_2 は活性化関数であり、f_1 には $tanh$、f_2 には softmax 関数を用いた。そして、$\boldsymbol{out}_p$ と、実際の単語間表現パターンのベクトルの 2 乗誤差をロス関数とする。これを最小化することで、単語間表現パターンの分布が類似した 2 語のベクトルの差が類似することが期待される。

学習時には語彙関係知識、単語間表現パターンそれぞれのロス関数において、単語ベクトルが初期値から大きく外れないように、学習中の単語ベクトルと初期単語ベクトルの二乗誤差項を追加する。

学習は 1 エポックで語彙関係知識に関するものと単語間表現パターンに関するものの両方を行い、評価実験では学習を 10 エポック行った。

4 実験と結果

4.1 実験設定

実験において、単語ベクトルの初期値として GloVe で学習したものを用いた。コーパスには ukWac*(269 万文書) を使い、前処理として各単語は小文字にし原形に直した。学習のパラメータはウインドウサイズ 15 単語、次元数 300、$\alpha = 0.75$、$X_{max} = 100$ とした。学習率の初期値は 0.05 とし、AdaGrad[5] の手法を用いて学習率を適応させた。学習は 50 エポック行った。

3.3 節で提案したニューラルネットワークにおいて、単語ベクトルならびに関係ベクトルの次元数は GloVe と同じ 300、隠れ層の次元数は 100 とし、単語間表現パターンとして頻度上位 3,000 個を用いた。また、マージン γ は TransE の設定にならい 2 とした。SGD で単語の embedding と各種行列を学習した。

4.2 評価セット

提案手法の有効性を示すためにアナロジータスクで評価した。アナロジータスクでは以下の 3 つを用いた。

- Google dataset[15]

  以下のような「a:b:c」の 3 語が与えられ、「a に対する b は c に対する何?」という問題に答えるものである。

  – athens:greece:tokyo:?
  – boy:girl:father:?

  上記の例での答えはそれぞれ、japan, mother となる。Google dataset は 10,675 個の syntactic analogy タスクと 8,869 個の semantic analogy タスクからなるが、本研究では syntactic な関係は学習しないので、上記の例にあげたような semantic な関係の評価だけを用いた。

- SemEval 2012 Task 2 [11]

  各カテゴリにその関係を表す典型的な語ペアが 3,4 つ与えられる。例えば、“REVERSE” カテゴリには (attack,defend) や (buy,sell) などが与えられる。そして、各カテゴリに語ペアのインスタンスが与えられ、それらが典型的な語ペアにどれくらい近いかを評価し、それらが人間が付与したものとどれくらい相関があるかでシステムの評価値を算出する。69 個のカテゴリがあり、評価尺度 Maxdiff で評価する。

- SAT dataset [21]

  SAT(アメリカの大学入学適正試験) から作られたデータセットで、例えば 2 語 “mason:stone” が与えられ、(a)teacher:chalk, (b)carpenter:wood, (c)soldier:gun, (d)photograph:camera, (e)book:word の 5 つの選択肢から、与えられた 2 語の間の関係と最も類似した関係を表すものを選ぶ問題である。この例では (b) が答えとなる。このデータセットは 374 問からなる。

上記の 3 つのアナロジータスクに対して、システムの出力は以下のように求める。まず、Google dataset に対しては、以下の cosine 類似度が最も大きくなる単語 w をシステム出力とする。

$$cos(\boldsymbol{w}_b - \boldsymbol{w}_a + \boldsymbol{w}_c, \boldsymbol{w}_w) \tag{10}$$

SemEval と SAT に対しては、単語 a と単語 b の関係と単語 c と単語 d の関係の類似度を以下の式で計算する。

$$cos(\boldsymbol{w}_b - \boldsymbol{w}_a, \boldsymbol{w}_d - \boldsymbol{w}_c) \tag{11}$$

そして、SemEval については人間がアノテーションしたものとシステムの出力する類似度の間の MaxDiff を計算し、SAT については 5 つの選択肢の中から類似度が最大となる選択肢を選ぶ。

*http://wacky.sslmit.unibo.it/doku.php

	Google	SemEval	SAT
コーパスのみ (GloVe)	73.5	42.1	42.2
+ 語彙関係知識	77.2	42.4	**43.4**
+ 単語間表現パターン	73.5	42.1	42.2
+ 語彙関係知識 + 単語間表現パターン	**77.5**	**43.0**	43.1
Bollegala[2] (コーパスのみ, GloVe)	61.5	38.0	–
Bollegala[2] (+ 語彙関係知識)	61.7	38.7	–
Bollegala[3] (コーパスのみ, GloVe)	86.8	44.7	44.7
Bollegala[3] (+ 単語間表現パターン)	86.9	44.9	46.0

表 3: 上部: 提案手法の実験結果, 下部: 関連研究の実験結果 (Google と SAT は正解率、semeval は Maxdiff を示す。いずれの評価尺度も 0 から 100 までの値をとり、大きい値の方がよい性能である。)

4.3 実験結果

提案手法の実験結果を表 3 の上部に示す。表中の数字は Google と SAT は正解率、semeval は Maxdiff を示す。学習にはランダム性があることから 3 回の実験の平均値を示している。Google と SemEval データセットについてはコーパスのみ (GloVe) に語彙関係知識と単語間表現パターンを利用することにより、最も精度が高い結果となったが、SAT データセットについては語彙関係知識を利用するものが最も精度が高い結果となった。また、いずれのデータセットについても単語間表現パターンのみを用いた場合はコーパスのみの場合と同じ精度となったが、語彙関係知識を併用することにより、精度向上がみられた。

表 3 の下部に関連研究の精度を示す。GloVe の学習パラメータの相違などでベースラインの精度が異なるため、一概には比較できないが、関連研究と比較して提案手法は同等またはそれ以上の精度を達成していることがわかる。

5 おわりに

本論文では語彙関係知識と単語間表現パターンを用いて単語分散表現を改善する手法を提案した。3 種類のアナロジータスクで評価したところ、Google と SemEval データセットについては語彙関係知識と単語間表現パターンを利用することで最も高い精度が得られた。

今後の課題としては、Wikipedia のリダイレクトやカテゴリなど、他の語彙関係知識の利用や、構文解析や文書分類など、アナロジータスク以外で提案手法の有効性を確認することなどがあげられる。

謝辞

本研究は JST CREST「知識に基づく構造的言語処理の確立と知識インフラの構築」の支援を受けたものである。

参考文献

[1] Mohit Bansal, Kevin Gimpel, and Karen Livescu. Tailoring continuous word representations for dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 809–815, Baltimore, Maryland, 2014.

[2] Danushka Bollegala, Mohammed Alsuhaibani, Takanori Maehara, and Ken-ichi Kawarabayashi. Joint word representation learning using a corpus and a semantic lexicon. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pp. 2690–2696, 2016.

[3] Danushka Bollegala, Takanori Maehara, and Ken-ichi Kawarabayashi. Embedding semantic relations into word representations. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pp. 1222–1228. AAAI Press, 2015.

[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 2787–2795. Curran Associates, Inc., 2013.

[5] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. Technical Report UCB/EECS-2010-24, EECS Department, University of California, Berkeley, Mar 2010.

[6] Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1606–1615, Denver, Colorado, May–June 2015. Association for Computational Linguistics.

[7] Daniel Fried and Kevin Duh. Incorporating both distributional and relational semantics in word representations. *CoRR*, Vol. abs/1412.5836, , 2014.

[8] Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka. Jointly learning word representations and composition functions using predicate-argument structures. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1544–1555, Doha, Qatar, 2014.

[9] Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: The 15th International Conference on Computational Linguistics*, 1992.

[10] Stephanie L. Hyland, Theofanis Karaletsos, and Gunnar Rätsch. A generative model of words and relationships from multiple sources. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pp. 2622–2629, 2016.

[11] David Jurgens, Saif Mohammad, Peter Turney, and Keith Holyoak. Semeval-2012 task 2: Measuring degrees of relational similarity. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pp. 356–364, Montréal, Canada, 7-8 June 2012. Association for Computational Linguistics.

[12] Dekang Lin and Shaojun Zhao. Identifying synonyms among distributionally similar words. In *In Proceedings of IJCAI-03*, pp. 1492–1493, 2003.

[13] Quan Liu, Hui Jiang, Si Wei, Zhen-Hua Ling, and Yu Hu. Learning semantic word embeddings based on ordinal knowledge constraints. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1501–1511, Beijing, China, July 2015. Association for Computational Linguistics.

[14] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. Curran Associates, Inc., 2013.

[15] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–751, Atlanta, Georgia, 2013.

[16] George A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, Vol. 38, No. 11, pp. 39–41, November 1995.

[17] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.

[18] Sascha Rothe and Hinrich Schütze. Autoextend: Extending word embeddings to embeddings for synsets and lexemes. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1793–1803, Beijing, China, July 2015. Association for Computational Linguistics.

[19] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems 26*, pp. 926–934. 2013.

[20] Joseph Turian, Lev-Arie Ratinov, and Yoshua Bengio. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 384–394, Uppsala, Sweden, 2010.

[21] Peter D. Turney. Similarity of semantic relations. *Comput. Linguist.*, Vol. 32, No. 3, pp. 379–416, September 2006.

[22] Chang Xu, Yalong Bai, Jiang Bian, Bin Gao, Gang Wang, Xiaoguang Liu, and Tie-Yan Liu. Rc-net: A general framework for incorporating knowledge into word representations. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pp. 1219–1228, New York, NY, USA, 2014. ACM.

[23] Mo Yu and Mark Dredze. Improving lexical embeddings with semantic knowledge. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 545–550, Baltimore, Maryland, June 2014. Association for Computational Linguistics.