語義曖昧性解消のための語義分散表現

梶澤 優希　山本 和英
長岡技術科学大学
{gumizawa, yamamoto}@jnlp.org

1 はじめに

語義曖昧性解消はコンピュータが意味理解するために必要であり、機械翻訳を始めとする様々な自然言語処理のタスクにおいて大きな問題となっている。そのため、語義曖昧性解消は教師あり学習や知識ベースなど様々な手法が活発に研究されている自然言語処理における大きなタスクの一つである [1]。

一方で、現状の語義曖昧性解消システムが実際に後段の言語処理タスクにおいて応用された例は非常に少ない。この理由として de Lacalle らは「語義曖昧性解消の精度」の問題を挙げている [2]。語義曖昧性解消の手法には主に教師あり学習によるものと知識ベースによる教師なし学習の 2 つが存在しているが、どちらも実用的な精度には至っていない。前者は後者に対して比較的性能が良いことが知られているが、それらすべてが規模の小さい訓練データに依存したものとなっているため、語義の被覆率や一語義当たりの訓練データ数を考えると実用的とは言えない。知識ベースによる手法では訓練データを必要としない反面、それらの殆どが大規模な辞書データに依存しており、辞書データの少ない言語においては手法自体の適用が困難になってしまう問題がある。

我々はこの問題に対して、平文コーパスと小規模な辞書資源から学習できる語義の分散表現を活用した新たな知識ベースの語義曖昧性解消手法である SV4D[1](Sense Vector for Disambiguation) を提案する。本論文で提案する新しい手法において必要な言語資源は語義に強く関係する単語のリストと平文コーパスのみであるため、大量の訓練データや構築に専門知識が必要な辞書等に依存しない。我々はこの提案手法に関して英語の語義曖昧性解消タスクを用いて他手法と比較し、有効性を示す。また、学習された分散表現に関して周辺文脈付き単語類似度のデータセットを用いて評価を行う。

[1]本論文で使用したコードとモデルは次のサイトで公開されている。https://github.com/gumigumi4f/sv4d

2 関連研究

我々が提案するモデルは Chen らの提案した手法 [3] と近い。Chen らの手法では語義ベクトルと文ベクトルの距離で語義曖昧性解消を行い、その結果に基づいて語義ベクトルを更新することでそれぞれのモデルの精度を相互に高めていく。この手法では文中に含まれる内容語の分散表現の平均ベクトルと語義ベクトルの距離のみで語義曖昧性解消を行っているが、我々は語義曖昧性解消において特徴量となる語が近い文にも出現することを日本語において確認している [4]。また、対象単語の前後の単語が特徴量として重要であることは広く知られており、これらをより考慮することは語義曖昧性解消の精度向上につながるはずである。さらに、Chen らが提案した手法では語義の選択と語義の分散表現が完全に分かれている。辞書で定義されている語義にそった分散表現の学習を行う場合、語義の選択が正しいかどうかによらず一意に決定した語義によって分散表現を学習することは意味のずれを招く。

語義の分散表現と語義曖昧性解消のモデルを教師なしにおいて学習する手法としては強化学習を用いた Lee らの MUSE が存在する [5]。これは語義の分散表現から得られる信号を強化学習の枠組みを用いて語義曖昧性解消のモデルに伝えることで確率的に語義を学習していく手法である。確率的に語義を学習することで Chen らの手法に比べてより良い語義ベクトルが学習できている一方で、MUSE では教師なしの枠組みを用いているため我々の欲しい粒度での曖昧性解消が行えないという欠点が存在する。

我々はここで上げた問題点を解決し、より高い精度で語義曖昧性解消が行える SV4D を提案する。本手法では辞書資源を語義曖昧性解消モデルと語義の分散表現を学習するモデル両方に適用することで、辞書の定義に沿った分散表現と語義曖昧性解消を同時に学習することができるものとなっている。

3 手法

3.1 語義曖昧性解消を行うモデル

語義曖昧性解消を行なうモデルは 1 層のニューラルネットから構成される。入力する特徴量には語義曖昧性解消の対象となる単語の前後 c 単語の単語ベクトルの平均 v_{ctx}、文に含まれる単語のベクトルの平均 v_{sent}、前後の文に含まれる単語のベクトルの平均 $v_{surrsent}$ の 3 つをつなぎ合わせたものを用いる。既存の研究と異なりこれら 3 つの特徴量を使うことによって幅広い文脈を考慮でき、高い精度での語義曖昧性解消が期待できる。式で表すと、語彙を W、対象の文章中に含まれる t 番目の単語を $w_t \in W$ と表したとき、単語 w_t の語義が $s_{tk} \in S_t$ である確率はソフトマックス関数を用いて以下のように表される。

$$\hat{p}(s_{tk}|\overline{C_t}) = \frac{\exp\left(\frac{Q^T_{s_{tk}} \cdot \overline{C_t} + P_{s_{tk}}}{\tau}\right)}{\sum_{s'_{tk} \in S_t} \exp\left(\frac{Q^T_{s'_{tk}} \cdot \overline{C_t} + P_{s'_{tk}}}{\tau}\right)} \quad (1)$$

$$\text{where } \overline{C_t} = v_{ctx_t} \oplus v_{sent_t} \oplus v_{surrsent_t}$$

ここで Q は出力層の行列、P はバイアス項を表す。また、$\oplus$ はベクトルのつなぎ合わせの演算を表し、式中の τ はソフトマックス関数の温度を表すものである。

ソフトマックスにおける温度 τ は出力確率を調整する項である。ソフトマックスの温度が高い場合は低い確率を強調し、逆に温度が低い場合は高い確率をより高く出力するようになる。このパラメータは正しく語義ベクトルと語義曖昧性解消を行うためのものである。温度が高い状態での学習を行っても語義が混ざったような学習しか行えず、精度の良い語義ベクトルの学習は難しい。逆に語義曖昧性解消を行うモデルの学習が進んでいない状態において低い温度を用い一意に語義を決定することは誤った語義ベクトルの学習につながると考えられる。そこで、温度パラメータを学習が進むにつれて減少させることで、学習前半では探索的に語義の確率を出力するようにし、学習後半では低い温度パラメータで正しい語義の分散表現が学習できるようにする。

3.2 語義の分散表現を学習するモデル

語義曖昧性解消を行うモデルによって計算された語義の確率を元に、語義の分散表現を学習する。語義の分散表現の学習には既存の手法である Negative Sampling を適用した Skip-gram を用いる。2 層のニューラルネットの隠れ層と出力層をそれぞれ U と V で表したとき、語義の分散表現を学習するモデルでは式 (2) に沿って学習を行う。

しかし、式 (2) には Negative Sampling を適用することができない。これは語義を選択するモデルで計算された確率 $\hat{p}(s_{tk}|\overline{C_t})$ が log 関数の内部に含まれているためである。そこで、式 (2) に対してイェンセンの不等式を適用し、対数尤度の下限を最大化するように式を変形することでこの問題を回避する。

$$\frac{1}{T}\sum_{t=1}^{T} - \sum_{\substack{c \le j \le c \\ j \neq 0}} \sum_{s_{tk} \in S_t} \log \hat{p}(s_{tk}|\overline{C_t}) p(w_{t+j}|s_{tk}) \quad (2)$$

$$\geq \frac{1}{T}\sum_{t=1}^{T} - \sum_{\substack{c \le j \le c \\ j \neq 0}} \sum_{s_{tk} \in S_t} \hat{p}(s_{tk}|\overline{C_t}) \log p(w_{t+j}|s_{tk}) \quad (3)$$

変分ベイズ法のように対数尤度の下限を最大化するような目的関数を考えることで、ソフトマックスの近似である Negative Sampling が適用できる。Negative Sampling を $\log p(w_{t+j}|s_{tk})$ に対して適用すると以下のようになる。

$$\begin{aligned} \log p(w_{t+j}|s_{tk}) = {} & \log \sigma(U^T_{s_{tk}} \cdot V_{w_{t+j}}) \\ & + \sum_{\substack{w'_v \in W_{neg} \\ w'_v \notin W_{pos}}} \log \sigma(-U^T_{s_{tk}} \cdot V_{w'_v}) \\ & + \beta_d \sum_{w''_v \in W_{pos}} \log \sigma(U^T_{s_{tk}} \cdot V_{w''_v}) \end{aligned} \quad (4)$$

ここで、W_{neg} は語彙の集合から適当な数だけサンプルしたものを表す。また W_{pos} は辞書資源から抽出した語義に関連するペアの集合を表す。

式 (4) の最後の項は dict2vec[6] で用いられている Positive Sampling を表すものである。これは、事前に作った語義に強く共起するような単語のペアに対する出現確率を最大化するような学習をする項である。例として、銀行を表す「bank」であれば、「deposit」や「money」といった単語と強く共起することが予想される。そこで、このようなペアの情報を制約として目的関数に加えることで、語義の定義に沿った分散表現の学習が行えると考えた。

3.3 語義曖昧性解消を行なうモデルの学習

語義曖昧性解消を行なうモデルに対して式 (3) の目的関数を用いても、Lee ら [5] が述べているように正しい語義の選択を学習することはできない。

そこで、我々は語義曖昧性解消を行うモデルの学習に対して式 (5) の目的関数を用いる。なお、式中の c' は語義曖昧性解消において周辺何単語を考慮するかを表す窓幅である。

$$
\begin{aligned}
J(P,Q) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s_{tk'}\in S_t} L_{s_{tk}} \log p(s_{tk}|\overline{C_t}) \quad (5)\\
\text{where } L_{s_{tk}} &= \frac{\exp(l_{s_{tk}})}{\sum_{s'_{tk}\in S_t}\exp(l_{s'_{tk}})}\\
l_{s_{tk}} &= \max_{\substack{-c'\le j\le c'\\ j\neq 0}} U_{s_{tk}}^{T}\cdot V_{w_{t+j}}\\
&+ \beta_r \sum_{w_{v''}\in W_{pos}} \max_{\substack{-c'\le j\le c'\\ j\neq 0}} U_{w_{t+j}}^{T}\cdot V_{w''_v}
\end{aligned}
$$

式 (5) では周辺単語の語義の分散表現を学習するモデルから得られる語義の分散表現と辞書に含まれるペアによって計算されるモデルとなっている。一般的に考えれば、その語義が正しいかどうかを判定する材料として、入力語義 s_{tk} と周辺単語 w_{t+j} の類似度を用いるのが普通である。我々はそこに対して辞書から得られる関連語のペアを語義の分散表現を学習するモデルと同様に適用することで同様に良いモデルが学習できるのではないかと考えた。β_r が掛かっている項が辞書を考慮するための項である。この項では周辺単語と対象語義の辞書内に含まれるペアとの類似度を考慮する項を表す。これによって、語義曖昧性解消を学習するモデルに対しても正しい語義の選択を学習させることが可能となる。

これをまとめると我々が提案するモデルは語義曖昧性解消を行うモデルと語義の分散表現を学習するモデルの 2 つを相互に学習しつつ、辞書資源を活用することでそれぞれのモデルがより良い学習ができるように改善を加えたものとなる。

3.4 その他

提案手法では語義の分散表現と語義曖昧性解消をバランスよく学習する必要がある。これは、不正確なモデルが他方のモデルに影響を及ぼすためある。そこで、分散表現の学習に際して入力単語に対してすべての周辺単語を予測するのではなく、周辺からサンプリングした 1 単語に対して学習を行い、その後語義曖昧性解消モデルを学習するようにする。

また、語義の分散表現を学習する際に、同義である単語の分散表現を統一して学習することで、より正しく語義の分散表現を得られるようにする。これは“car”や“automobile”といった単語に存在する共通の意味の語義の分散表現を一つにまとめるものである。

4 実験設定

提案手法を評価するため、語義曖昧性解消モデルと語義の分散表現についてそれぞれ評価を行う。語義曖昧性解消では粒度の細かいタスク [1] と粒度の粗いタスク [7] の 2 つの英語語義曖昧性解消タスクで評価する。語義の分散表現の評価にはコンテキスト付き単語類似度のデータセットである SCWS[8] を用いる。

英語での実験に際して語義と強く共起するペアを Basile ら [9] が用いている方法で抽出し、重み上位 15 単語を使用することとした。モデルの学習には 2010 年 4 月にダンプされた英語 Wikipedia のコーパス[2]を用いる。企業名など WordNet に存在しない語義の影響を除くため固有名詞を学習から省いたものを使用する。式 (1) 中の τ は 1.0 から 0.1 まで線形的に下がるようにした。辞書を考慮する項である β_d と β_r はそれぞれ 0.20, 1.00 に設定した。語義の分散表現を学習するモデルの次元は 300 に設定し、Negative Sampling の値は 5、分散表現の学習に用いる窓幅の大きさ c は 5、語義曖昧性解消の際に考慮する窓幅の大きさ c' は 20 とした。また、W_{pos} としてペアの集合からサンプルする数は 4 に設定した。

5 実験と考察

5.1 語義ベクトルの評価

語義の分散表現について SCWS を用いて比較を行った。表 1 に各手法の AvgSimC と MaxSimC を記す。表 1 を見ると AvgSimC においては他の手法とほぼ同等の精度を記録していることがわかる。この結果から我々の手法が正しく語義の分散表現を学習できていることが確認できる。

次に学習された語義の分散表現を定性的に確認する。表 2 を確認すると各語義の近傍に対して関連語が集中しているということが見て取れ、語義とその関連語の関係が正しく学習できていることがわかる。

表 1: SCWS での評価

Method	MaxSimC	AvgSimC
Word2Vec	66.1	
MUSE [5]	67.9	68.7
Chen et al. [3]	-	68.9
SV4D	61.0	68.7

[2]http://nlp.stanford.edu/data/WestburyLab.wikicorp.201004.txt.bz2

表 2: 語義ベクトルに対する近傍の単語, 語義

Word	Sense	Nearest Neighbour
bank	bank.n.01	bank.v.02, riverbank.n.01, west_bank, river, west_bank.n.01
	depository_financial_institution.n.01	bank.v.05, financial_institution.n.01, financial_institution
star	star.n.01	stars, star, galaxy.n.03, binary_star, nebula, galaxies
	ace.n.03	superstar, superstars, adept.s.01, star.v.02, prodigy

5.2 語義曖昧性解消タスクでの評価

語義曖昧性解消のモデルに関して各タスクを用い評価を行う。粒度の細かい語義曖昧性解消では、その難易度が高いことから語義の頻度を用いて補正を行う場合が多いため、提案手法に関しても同様の補正を行い評価する。

表 3 を確認すると粒度によらず、高い精度での語義曖昧性解消ができていることが確認できる。特に粒度の細かい語義曖昧性解消においては、知識ベースによる既存手法を大きく上回る精度を記録しており、教師ありのシステムとも肉薄する精度を記録している。粒度の粗い語義曖昧性解消では既存の教師なしシステムである Chen らの手法を上回る精度を記録しており、我々の手法が語義の頻度がない場合でも有効であることを示している。

一方で語義の頻度を用いていない粗い語義曖昧性解消タスクでは MFS の精度を超えることができていない。名詞に関してのみ調査すると MFS が 77.4 なのに対し提案手法が 83.8 と大きく上回っていることから、名詞以外の品詞での精度が低いことがわかる。これは、名詞以外の品詞については分散表現上の距離以外の特徴量が語義曖昧性解消に必要であることを示唆するものである。

表 3: 語義曖昧性解消タスクでの評価 (F 値)

Type	Method	Fine-grained	Coarse-grained
Supervised	MFS	64.8	78.9
	IMS	68.4	-
	IMS$_{-s+emb}$	69.6	-
Knowledge-based	Babelfy	65.5	-
	UKB	67.3	-
	Chen et al.	-	75.8
	SV4D	69.5	76.4

6 結論

本論文では新たな知識ベース語義曖昧性解消を提案し英語の語義曖昧性解消タスクで有効性を確認した。我々の手法は一般的な語義曖昧性解消手法に比べ適用が容易な一方で固有名詞等への対応ができていない。今後はこれらの問題を解決し、より汎用的なツールを作成することが課題となる。また、日本語への適用と有効性の確認も行いたいと考えている。

謝辞

本研究は、平成 27〜31 年科学研究費補助金基盤 (B) 課題番号 15H03216、及び平成 29〜31 年科学研究費補助金挑戦的研究 (萌芽) 課題番号 17K18481 の助成を受けています。

参考文献

[1] Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 99–110, 2017.

[2] Oier Lopez de Lacalle and Eneko Agirre. A methodology for word sense disambiguation at 90% based on large-scale crowdsourcing. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pp. 61–70, 2015.

[3] Xinxiong Chen, Zhiyuan Liu, and Maosong Sun. A unified model for word sense representation and disambiguation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1025–1035, 2014.

[4] 桾澤優希, 山本和英. かな漢字換言を通した日本語語義曖昧性解消の分析. 自然言語処理, Vol. 25, No. 3, pp. 255–293, 2018.

[5] Guang-He Lee and Yun-Nung Chen. MUSE: Modularizing unsupervised sense embeddings. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 327–337, 2017.

[6] Julien Tissier, Christopher Gravier, and Amaury Habrard. Dict2vec : Learning word embeddings using lexical dictionaries. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 254–263, 2017.

[7] Roberto Navigli, Kenneth C. Litkowski, and Orin Hargraves. Semeval-2007 task 07: Coarse-grained english all-words task. In *Proceedings of the Fourth International Workshop on Semantic Evaluations*, pp. 30–35, 2007.

[8] Eric Huang, Richard Socher, Christopher Manning, and Andrew Ng. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 873–882, 2012.

[9] Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. An enhanced lesk word sense disambiguation algorithm through a distributional semantic model. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, pp. 1591–1600, 2014.