

Distant Supervisionにおける 事前学習によるノイズ削減法

須藤 広大 能地 宏 松本 裕治
奈良先端科学技術大学院大学 情報科学研究科
{sudo.kodai.rx2, noji, matsu}@naist.is.jp

訂正記事について

表題の研究について、実験設定について誤りがあったため、本稿に訂正記事を掲載する。Distant Supervisionにおける事前学習によるノイズ削減法という表題の実験について、一部の手法において開発データが含まれている点など、実験の設定について誤りがあったため、Distant Supervisionにおける選択学習によるノイズ削減法という表題にて、正しい実験設定、結果について、元論文の4.5,6章についての訂正記事を本稿にて報告する。

正誤表

1,2,3章
(誤) 事前学習
(正) 選択学習
(誤) バグ
(正) バグ

4.5,6章
元論文の4.5,6章の訂正記事を以下に記載する。

4 選択学習によるノイズ削減法

本論文では、前述した問題点を解決するために、Riedelら [6] によって、

「もし何らかの意味関係を持つ二つのエンティティが事実として存在するならば、それらのエンティティを含むような文は、少なくとも一つの文はその意味関係を表す」

と置かれていた仮定を、

「もし何らかの意味関係を持つ二つのエンティティが事実として存在するならば、それらのエンティティを含むような文は、少なくとも一つの文はその意味関係を表す。ただし、例外が存在する。」

と新たに仮定を置き換え、学習データにはバグレベルでノイズデータが混入することを仮定する。上記の仮定をふまえて、信頼度の高いデータのみを選択して、学習に用いることで、学習時のノイズ削減を実現し、関係抽出器の性能向上を目指す。本論文において、信頼度の高いデータとは、正解の関係ラベルに対して、モデルによる確信度が高いデータと定義する。

本論文では、通常の学習を行った後に、選択学習を行うことで、ノイズ削減を目指す。三種類のノイズ削減法を提案し、関係抽出器の性能が従来手法より優れていることを検証する。

4.1 段階学習を用いたノイズ削減法

本手法のアイデアは、通常の学習により得た重みパラメータ θ を用いて、信頼度の高いデータのみを用いて学習することで、ノイズ削減を行う手法である。本論文では、エポックごとに学習するサンプル数を増やしていく手法を段階学習という名称で定義する。

段階学習の目的関数は、以下のように定義する。

$$J(\theta) = \sum_{i=1}^n \log p(\mathbf{y}_i | \mathbf{b}_i, \theta) \quad (1)$$

n は、初期値に一定のサンプル数の割合を設定しており、エポックごとに一定数の割合で増やす。初期値に設定する一定のサンプル数の割合を $init_rate$ 、増やしていく割合を $pace_rate$ と定義し、 n はエポックごとに以下の式の更新を行う。

$$n = \begin{cases} (init_rate + (pace_rate * \text{エポック数})) * T & \text{if } (n < T) \\ T & \text{(otherwise)} \end{cases}$$

また、選択するデータはクロスエントロピーの値が低いサンプルから順に n 個を選択し、シャッフルした後、学習に用いる。

4.2 Self-Paced Learning を用いたノイズ削減法

本手法のアイデアは、Kumar らが提案した Self-Paced Learning [3] (SPL) をノイズ削減法として用いることである。SPL は、Bengio らが提案したカリキュラム学習 [1] の一種であり、エポックごとに、動的にデータを学習する順番とサンプル数を決める手法である。本論文では、選択学習を実現するために、SPL を用いる。

SPL では、目的関数 \mathbf{E} を以下のように定義する。

$$\min_{\mathbf{w}} \mathbf{E}(\theta, \mathbf{v}; \lambda) = \sum_{i=1}^T v_i J(\theta) - \lambda \sum_{i=1}^T v_i \quad (2)$$

上記の式の v_i は 選択パラメータ $\mathbf{v} = [v_i]_1^T$ の要素であり、1 であればサンプルを選択して、0 であればサンプルを選択しないことを表す。更新率 λ は、学習速度をコントロールするパラメータである。

本論文での SPL の目的は、負の L1 正則化項である $-\|\mathbf{v}\|_1 = -\sum_{i=1}^T v_i$ を加えたトレーニングロス最小化することで、学習サンプルを選択する選択パラメータ \mathbf{v} と重みパラメータ θ を交互に学習することで、信頼度の高いサンプルから順に学習を行う選択学習の役割を実現する。

4.3 ミニバッチ内選択学習を用いたノイズ削減法

本手法は、ミニバッチを用いた学習時に、ミニバッチ内の上位 n 個を学習に用いる手法である。段階学習と違う点は、常に一定数を学習に用いており、全データ内の上位 n 個を学習に用いるのではなくて、ミニバッチ内の上位 n 個を学習に用いることで、過学習しないことを考慮した学習を行う。以下に、 n の定義を行う。

$$n = [\text{batch_rate} * \text{バッチサイズ}]$$

n は、必ず自然数となり、ミニバッチ内で学習に用いるデータの個数を表す。

5. 実験

本論文では、以下の三種類の実験を行った。また、初期値として、通常の学習により最適化した重みパラメータ θ を用いる。

1. 段階学習を用いた実験
2. SPL を用いた実験
3. ミニバッチ内選択学習を用いた実験

5.1 実験設定

5.1.1 データセット

本研究で用いたデータセットは Riedel ら [6] によって作成されたデータセットを用いる。このデータセットは Hoffman ら [2] や Zeng ら [7] 等によって幅広く使われているデータセットである。知識ベースに Freebase、ドキュメントコーパスに NYTimes が用いられている。Freebase は訓練データ、評価データの二つに分かれており、訓練データを用いて、2005-2006 年の NYTimes のコーパスにラベル付けしたものを訓練コーパスとして、評価データを用いて、2007 年の NYTimes のコーパスにラベル付けしたものを評価コーパスとしている。訓練コーパスで使用されている文は 522,611 文あり、実体のペアは 281,270 ペアあり、18,252 の事実が存在する。評価コーパスで使用されている文は 172,448 文あり、実体のペアは 96,678 ペアあり、1950 の事実が存在する。また、本実験では、コーパスは全て小文字処理とトークナイズ処理が行われているデータであり、Yankai ら [4] が実験に用いたデータを使用する。¹

5.1.2 評価手法

Mintz ら [5] や Zeng ら [7] によって、用いられている評価手法で評価を行う。テストコーパスから予測した事実の上位 2000 件、つまり、テストコーパス \mathcal{D} を入力としたときのモデルの予測 \hat{y} の中で、予測確率が高いもの上位 2000 件が、評価データベースの Freebase に存在するかどうかで検証を行う。本論文では、適合率-再現率曲線による評価と AUC 値での二種類による評価での検証を行う。AUC 値は、適合率-再現率曲線の線分の下での面積の値であり、適合率-再現率曲線は、高い位置にあれば良い結果となる。また、この評価手法は、人手での評価のように時間がかからず、アノテーションも必要がないため、適切な精度の評価をコストなしに行う事が出来る。

¹<https://github.com/thunlp/NRE/tree/master/data>

表 1: ハイパーパラメータ一覧

パラメータ	値
ウィンドウ幅 w	3
フィルター数 H	230
単語ベクトル次元 d_w	50
位置ベクトル次元 d_p	5
ドロップアウト率 p	0.5
バッチサイズ $Batch$	160
学習率 η	0.003
最適化手法 Opt	SGD

表 2: 開発データと評価データに対する、AUC 値による段階学習の比較結果

モデル	初期値 <i>init_rate</i>	ベース値 <i>pace_rate</i>	AUC 値 (TEST)	AUC 値 (DEV)
PCNN	-	-	23.85	43.42
+ 段階学習	0.1	0.05	23.72	42.67
	0.3	0.05	24.47	42.59
	0.5	0.01	25.09	42.90
		0.05	23.41	43.61
	0.7	0.01	25.17	43.40
		0.05	23.84	43.38
1.0	-0.0005	-0.003	24.03	43.54
			24.38	43.36

5.1.3 実装詳細

モデルの実装には、TensorFlow² を用いて、CPU(Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz) の環境下で実行した。本論文の実験は、配布されている訓練データを 2:1 の割合で分割して、前者を学習データ、後者を開発データとする。Zeng ら [7] による PCNN での通常の学習結果をベースラインとする。ベースラインの実験はエポック数 70 に設定し、提案手法の実験はエポック数 50 に設定した。このとき、提案手法の実験には事前学習として、通常の学習を 20 回行った。ベースライン、提案手法の実験は開発データに対しての AUC 値が最も高いエポック時のパラメータを用いて、テストデータでの評価実験を行った。

使用したハイパーパラメータの一覧は表 1 に示す。各ハイパーパラメータは、Zeng ら [7] と Yankai ら [4] の実験設定を参考に選択した。また、重みパラメータ行列は、それぞれ [-0.2, 0.2] から正規分布に従ってサンプリングした値を設定し、モデルに用いた単語分散表現の初期値には、CBOW を用いて NYTimes コーパスを学習した Yankai ら [4] が実験に用いた単語分散表現を使用した。

5.2 実験結果と考察

開発データと評価データに対する、従来手法と提案手法の AUC 値の比較を表 5 に示す。それぞれの提案手法で開発データに対して、最も高い AUC 値を持つパラメータを用いた評価データに対する実験結果を比較する。手法によらず、ノイズ削減法を用いることで、Distant Supervision を用いた関係抽出のようなノイズが存在するデータでの学習時に従来手法と比較して有効であることを示すことができ、従来手法よりも高い性能の関係抽出器を実現した。従来手法では、文レベルのノイズ削減しか行われていなかったが、本タスクで用いられる Riedel らのデータセット [6] において、バグレベルのノイズ削減を行うことで、間違った関

²<https://www.tensorflow.org/>

表 3: 開発データと評価データに対する, AUC 値による SPL の比較結果

モデル	更新率 λ	更新速度率 μ	AUC 値 (TEST)	AUC 値 (DEV)
PCNN	-	-	23.85	43.42
+ SPL	0.3	$\times 1.5$ $\times 2.0$	26.16 26.17	42.76 42.86
	1.0	$+ 0.5$ $+ 1.0$	25.84 25.48	43.44 43.27

表 4: 開発データと評価データに対する, AUC 値によるミニバッチ内選択学習の比較結果

モデル	データの割合 <i>batch_rate</i>	AUC 値 (TEST)	AUC 値 (DEV)
PCNN	-	23.85	43.42
+ ミニバッチ内選択学習	0.2	24.85	43.83
	0.4	24.10	44.03
	0.6	25.02	43.53
	0.8	23.70	43.94

係ラベルが付与されているバッグが学習の妨げになっていることが立証できた。

段階学習, SPL の実験結果の表 2, 3 によると, 評価データに対して, AUC 値が高くて, 開発データに対して, AUC 値が低いこともあり, 逆もまた同様であると考えられる。これは, 学習データと開発データが同じコーパスから生成されていることで, エンティティが三つ以上含まれているような文に対しては, 違うエンティティのペアを指しているにもかかわらず, 中身の文集合が同じになるような分割になっていることが考えられる。そのため, 学習データと開発データに対して過学習してしまい, 評価データに対しては, 性能が下がってしまうといったことが考えられる。

ミニバッチ内選択学習の表 4 によると, 段階学習や SPL では, 全データの中で, 上位に来るサンプルを学習しているが, ミニバッチ内選択学習では, ミニバッチ単位で上位に来るサンプルを学習しているため, 過学習せずに, 開発データ, 評価データに対して, AUC 値が高くなっていると考えられる。

6 おわりに

本研究では, Distant Supervision における, 選択学習におけるノイズ削減法を提案した。通常の学習と比較して, データセットの中で信頼度の高いデータのみを用いる選択学習が予測時により高い性能が得られることがわかった。今後の課題として, どのように選択して学習するか, 信頼度の高いデータの抽出方法の工夫, また, 他のタスクにおいて, データ拡張と組み合わせた半教師あり学習手法としての応用の検証などが考えられる。

参考文献

- [1] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 41–48, New York, NY, USA, 2009. ACM.
- [2] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In

表 5: 開発データと評価データに対する, AUC 値による提案手法の比較結果

モデル	AUC 値 (TEST)	AUC 値 (DEV)
PCNN	23.85	43.42
段階学習	23.41	43.61
SPL	25.84	43.44
ミニバッチ内選択学習	24.10	44.03

Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11, pages 541–550, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

- [3] M. P. Kumar, B. Packer, and D. Koller. Self-paced learning for latent variable models. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, NIPS'10, pages 1189–1197, USA, 2010. Curran Associates Inc.
- [4] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*, 2016.
- [5] M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 1003–1011, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [6] S. Riedel, L. Yao, and A. McCallum. Modeling relations and their mentions without labeled text. In *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part III*, ECML PKDD'10, pages 148–163, Berlin, Heidelberg, 2010. Springer-Verlag.
- [7] D. Zeng, K. Liu, Y. Chen, and J. Zhao. Distant supervision for relation extraction via piecewise convolutional neural networks. In L. Mrquez, C. Callison-Burch, J. Su, D. Pighin, and Y. Marton, editors, *EMNLP*, pages 1753–1762. The Association for Computational Linguistics, 2015.