

sbpark@nova.snu.ac.kr {btzhang,ytkim}@cse.snu.ac.kr

Word Sense Disambiguation From Unlabelled Data

Seong-Bae Park⁰ Byoung-Tak Zhang Yung Taek Kim

School of Computer Science and Engineering, Seoul National University

- 가
가
가
33.6%
가
[5].
(WSD) Tadopalli
(committee) , Liere
(Active Learning) [6].
가 (classifier)
[2,5,6,9]. WSD 가
2 가
WSD 가
1. 가 가
2. 가 90%
3. 가
[3].
가 Atsushi
가 [1].
가 [5].
[2]. 가 (TUF)
가 ●
가 ●
가
가 , Hwee Lee
가 (1).
(PARENT, NMODWORD, ADNWORD)

가 , (GFUNC) 1 11 가
 가 , 가 .
 .
 , Quinlan C4.5
 release 8 [7].
 C4.5 , C4.5
 가 , C4.5
 PARENT, ADNWORD, NMODWORD
 가

GFUNC	W
PARENT	W
SUBJECT	PARENT 가
OBJECT	PARENT 가
NMODWORD	W
ADNWORD	W
ADNSUBJ	ADNWORD 가
ADNOBJ	ADNWORD 가

1 : w

3. , (selective
 sampling) 1 W 가

AdaBoost.M1[4]
 L

M , M ,
 $C_j(1 \leq j \leq M)$ x_t ,
 W_t 가 C_j ,
 x_t s_t 가 ,
 α_t ϵ_t 가 , ϵ_t 가
 α_t 가 ,
 α_t 가 θ

가 x_t 가 α_t 가 θ
 x_t 가 s_t 가 x_t θ

3.1 1 가 w
 가 x x s_t

$$s_t = \arg \max_{s \in SS(w)} \sum_{j: C_j(x)=s} W(j)$$

Given unlabeled example set $D = \{x_1, \dots, x_T\}$
 And labeled example set L

Initialize $W_1(j) = 1 / M$

Resample $S_j^{(0)}$ from L for each classifier C_j where $|S_j^{(0)}|=|L|$ as done in Bagging.

Train each base classifier $C_j(1 \leq j \leq M)$ from $S_j^{(0)}$.

For $t = 1, \dots, T$:

1. Determine the sense of $x_t \in D$ from each C_j .
 $S = \langle s_1, \dots, s_M \rangle$
2. Find the most likely sense s_t from S using distribution W .
3. Set $\alpha_t = (1 - \epsilon_t) / \epsilon_t$ where

$$\epsilon_t = \frac{\text{No. of classifiers whose output is not } s_t}{M}$$

4. If α_t is larger than a certainty threshold θ , then update W_t .

$$W_{t+1}(j) = \frac{W_t(j)}{Z_t} \times \begin{cases} \alpha_t & \text{if } s_j = s_t \\ 1 & \text{otherwise} \end{cases}$$

where Z_t is a normalization constant.

5. Otherwise every classifier C_j is restructured from $S_j^{(t)}$.
 $S_j^{(t)} = S_j^{(t-1)} + (x_t, s_t)$

Output the final classifier:

$$C(x) = \arg \max_{s \in SS(x)} \sum_{t: C_t(x)=s} W_t(j)$$

1 :

SS(w) w 가
 t 가 W_t 가
 W_T 가

3.2

가
 가
 가
 W_1 W_2 WordNet
 $Sim(W_1, W_2)$
 W_1 W_2 $Sim(W_1, W_2)$

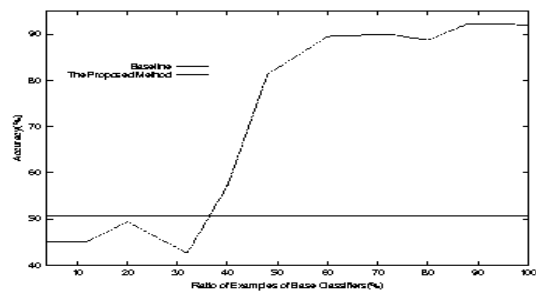
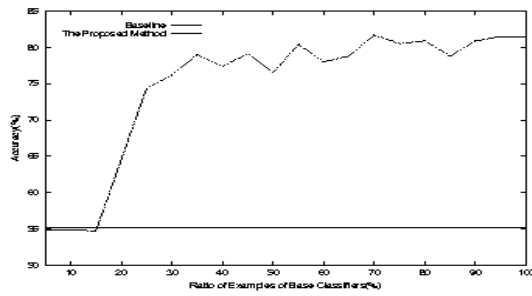
4.

4.1

KORTERM KAIST
 KAIST 가

3
 가 가
 2
 가
 3
 가

	876
	796



2 : Y X
10-fold cross validation

		Pear	6.2%
	4	Ship	55.2%
		Times	13.7%
		Stomach	24.9%
	3	Person	46.2%
		Minute	50.8%
		Indignation	3.0%

3 :

4.2

L , L + 1
4 3
10-fold cross validation

4 가 가

33.9%

2 가

2 X-

가

2(a)

가 가 , 가 , 35%
35%

		C4.5	
	81.5 ± 7.7%	82.3 ± 5.9%	55.2%
	92.3 ± 7.7%	94.3 ± 5.7%	50.8%
	86.9%	88.3%	53.0%

2 :

5.

가

가 5%

Utgoff [8].

6.

[1] Atsushi F., Kentaro I., Takenobu T., and Hozumi T. "To What Extent Does Case Contribute to Verb Sense Disambiguation?" In *Proceedings of COLING-96*, pp. 59-64, 1996.

[2] Atsushi F., Kentaro I., Takenobu T., and Hozumi T., "Selective Sampling of Effective Example Sentence Sets for Word Sense Disambiguation," *Computational Linguistics*, Vol. 24, No. 4, pp. 573-597, 1998.

[3] Cho J. and Kim G., "Korean Verb Sense Disambiguation Using Distributional Information From Corpora," In *Proceedings of Natural Language Processing Pacific Rim Symposium (NLPRS)*, pp. 691-696, 1995.

[4] Freund Y. and Schapire R., "Experiments with a New Boosting Algorithm," In *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148-156, 1996.

[5] Hwee T. and Lee H., "Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exmaplar-Based Approach," In *Proceedings of the 34th Annual Meeting of the ACL*, pp. 40-47, 1996.

[6] Liere R. and Tadepalli P., "Active Learning with Committees for Text Categorization," In *Proceedings of AAAI-97*, pp. 591-596, 1997.

[7] Quinlan R., *C4.5: Programs For Machine Learning*, Morgan Kaufmann Publishers, 1993.

[8] Utgoff P., Berkman N., and Clouse J., "Decision Tree Induction Based on Efficient Tree Restructuring," *Machine Learning*, Vol. 29, pp. 5-44, 1997.

[9] Yarowsky, D. "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods," In *Proceedings of the 33rd Annual Meeting of the ACL*, pp. 189--196, 1995.