Adaptively entropy-based weighting classifiers in combination using Dempster-Shafer theory for word sense disambiguation

Huynh V.-N., Nguyen T.T., Le C.A.

Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan;
College of Technology, Vietnam National University, 144 Xuan Thuy, Cau Giay District, Hanoi, Viet Nam

Abstract: In this paper we introduce an evidential reasoning based framework for weighted combination of classifiers for word sense disambiguation (WSD). Within this framework, we propose a new way of defining adaptively weights of individual classifiers based on ambiguity measures associated with their decisions with respect to each particular pattern under classification, where the ambiguity measure is defined by Shannon's entropy. We then apply the discounting-and-combination scheme in Dempster-Shafer theory of evidence to derive a consensus decision for the classification task at hand. Experimentally, we conduct two scenarios of combining classifiers with the discussed method of weighting. In the first scenario, each individual classifier corresponds to a well-known learning algorithm and all of them use the same representation of context regarding the target word to be disambiguated, while in the second scenario the same learning algorithm applied to individual classifiers but each of them uses a distinct representation of the target word. These experimental scenarios are tested on English lexical samples of Senseval-2 and Senseval-3 resulting in an improvement in overall accuracy. © 2009 Elsevier Ltd. All rights reserved.

Author Keywords: Classifier combination; Computational linguistics; Dempster's rule of combination; Entropy; Word sense disambiguation

Index Keywords: Ambiguity measure; Classification ,; Classification tasks; Classifier combination; Combination of classifiers; Combining classifiers; Dempster's rule of combination; Dempster-Shafer theory; Dempster-shafer theory of evidence; Evidential reasoning; Individual classifiers; Shannon's entropy; Word Sense Disambiguation; Computational linguistics; Decision theory; Entropy; Formal logic; Learning algorithms; Natural language processing systems; Classifiers

Document Type: Article

Authors with affiliations:

• Huynh, V.-N., Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan

• Nguyen, T.T., College of Technology, Vietnam National University, 144 Xuan Thuy, Cau Giay District, Hanoi, Viet Nam

• Le, C.A., College of Technology, Vietnam National University, 144 Xuan Thuy, Cau Giay District, Hanoi, Viet Nam

References:

• (2006) Word Sense Disambiguation: Algorithms and Applications, , Agirre E., and Edmonds P. (Eds), Springer, Dordrecht, The Netherlands

• Al-Ani, A., Deriche, M., A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence (2002) Journal of Artificial Intelligence Research, 17, pp. 333-361

• Bell, D., Guan, J.W., Bi, Y., On combining classifiers mass functions for text categorization (2005) IEEE Transactions on Knowledge and Data Engineering, 17 (10), pp. 1307-1319

• Bloehdorn, S., Andreas, H., Text classification by boosting weak learners based on terms and concepts (2004) Proceedings of the Fourth IEEE International Conference on Data Mining, pp. 331-334

• Chang, C.C., Lin, C.J., (2001) LIBSVM: A Library for Support Vector Machines, , http://www.csie.ntu.edu.tw/cjlin/libsvm

• Clough, P., Stevenson, M., Cross-language information retrieval using Euro WordNet and word sense disambiguation (2004) Proceedings of Advances in Information Retrieval, 26th European Conference on IR Research (ECIR), pp. 327-337. , Sunderland, UK, pp

• Dempster, A.P., Upper and lower probabilities induced by a multi-valued mapping (1967) Annals of Mathematics and Statistics, 38, pp. 325-339

• Denoeux, T., A k-nearest neighbor classification rule based on Dempster-Shafer theory (1995) IEEE Transactions on Systems, Man and Cybernetics, 25 (5), pp. 804-813

• Denoeux, T., A neural network classifier based on Dempster-Shafer theory (2000) IEEE Transactions on Systems, Man and Cybernetics A, 30 (2), pp. 131-150

• Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Zien, J.Y., Semtag and seeker: Bootstrapping the semantic web via automated semantic annotation (2003) Proceedings of the Twelfth International Conference on World Wide Web, pp. 178-186

• Escudero, G., Màrquez, L., Rigau, G., Boosting applied to word sense disambiguation (2000) Proceedings of the 11th European Conference on Machine Learning, pp. 129-141

• Florian, R., Yarowsky, D., Modeling consensus: Classifier combination for word sense disambiguation (2002) Proceedings of EMNLP, pp. 25-32

• Grozea, C., Finding optimal parameter settings for high performance word sense disambiguation (2004) Proceedings of ACL/SIGLEX Senseval-3, pp. 125-128. , Barcelona, Spain, July

• Hoste, V., Hendrickx, I., Daelemans, W., van den Bosch, A., Parameter optimization for machine-learning of word sense disambiguation (2002) Natural Language Engineering, 8 (3), pp. 311-325

• Ide, N., Véronis, J., Introduction to the special issue on word sense disambiguation: the state of the art (1998) Computational Linguistics, 24, pp. 1-40

• Kilgarriff, A., English lexical sample task description (2001) Proceedings of Senseval-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems, pp. 17-20. , Toulouse, France, pp

• Kilgarriff, A., Rosenzweig, J., Framework and results for English SENSEVAL (2000) Computers and the Humanities, 36, pp.

15-48

- Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., On combining classifiers (1998) IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (3), pp. 226-239

- Klein, D., Toutanova, K., Tolga Ilhan, H., Kamvar, S.D., Manning, C.D., Combining heterogeneous classifiers for word-sense disambiguation (2002) ACL WSD Workshop, pp. 74-80

- Leacock, C., Chodorow, M., Miller, G., Using corpus statistics and WordNet relations for sense identification (1998) Computational Linguistics, 24 (1), pp. 147-165

- Lee, Y.K., Ng, H.T., An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation (2002) Proceedings of EMNLP, pp. 41-48

- Le, C.A., Huynh, V.-N., Shimazu, A., 2005. An evidential reasoning approach to weighted combination of classifiers for word sense disambiguation. In: Perner, P., Imiya, A. (Eds.), MLDM 2005, LNCS 3587. Springer-Verlag, pp. 516-525Le, C.A., Huynh, V.-N., Shimazu, A., Nakamori, Y., Combining classifiers for word sense disambiguation based on Dempster-Shafer theory and OWA operators (2007) Data & Knowledge Engineering, 63 (2), pp. 381-396

- Le, A.C., Shimazu, A., Huynh, V.-N., Nguyen, L.M., Semi-supervised learning integrated with classifier combination for word sense disambiguation (2008) Computer Speech and Language, 22 (4), pp. 330-345

- Leroy, G., Rindflesch, T.C., Effects of information and machine learning algorithms on word sense disambiguation with small datasets (2005) International Journal of Medical Informatics, 74 (7-8), pp. 573-585

- Lin, H.-T., Lin, C.-J., Weng, R.C., A note on Platts probabilistic outputs for support vector machines (2007) Machine Learning, 68, pp. 267-276

- Melamed, I.D., Resnik, P., Tagger evaluation given hierarchical tag sets (2000) Computers and the Humanities, 34 (1-2), pp. 79-84

- Mihalcea, R., Chklovski, T., Killgariff, A., The Senseval-3 English lexical sample task (2004) Proceedings of ACL/SIGLEX Senseval-3, pp. 25-28. , Barcelona, Spain, July

- Montoyo, A., Suarez, A., Rigau, G., Palomar, M., Combining knowledge- and corpus-based word-sense-disambiguation methods (2005) Journal of Artificial Intelligence Research, 23, pp. 299-330

- Mooney, R.J., Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning (1996) Proceedings of the EMNLP, pp. 82-91

- Ng, H.T., Lee, H.B., Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach (1996) Proceedings of the 34th Annual Meeting of the ACL, pp. 40-47

- Pedersen, T., A simple approach to building ensembles of Naive Bayesian classifiers for word sense disambiguation (2000) Proceedings of the North American Chapter of the ACL, pp. 63-69

- Platt, J., Probabilistic outputs for support vector machines and comparison to regularized likelihood methods (2000) Advances in Large Margin Classifiers, , Smola A., Bartlett P., Schölkopf B., and Schuurmans D. (Eds), MIT Press, Cambridge

- Rogova, G., Combining the results of several neural network classifiers (1994) Neural Networks, 7 (5), pp. 777-781

- Sanderson, M., Word sense disambiguation and information retrieval (1994) Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 142-151. , Dublin, Ireland, pp

- Shafer, G., (1976) A Mathematical Theory of Evidence, , Princeton University Press, Princeton

- Tsuruoka, Y., 2006. A Simple C++ Library for Maximum Entropy Classification. Vossen, P., Rigau, G., Alegria, I., Agirre, E., Farwell, D., Fuentes, M., Meaningful results for information retrieval in the MEANING project (2006) Proceedings of Third International WordNet Conference, , Jeju Island, Korea

- Wang, X.J., Matsumoto, Y., Trajectory based word sense disambiguation (2004) Proceedings of the 20th International

Conference on Computational Linguistics, pp. 903-909. , Geneva, August

• Xu, L., Krzyzak, A., Suen, C.Y., Several methods for combining multiple classifiers and their applications in handwritten character recognition (1992) IEEE Transactions on Systems, Man and Cybernetics, 22, pp. 418-435

• Yarowsky, D., Cucerzan, S., Florian, R., Schafer, C., Wicentowski, R., The Johns Hopkins SENSEVAL2 system descriptions (2001) Proceedings of SENSEVAL2, pp. 163-166

• Zadeh, L.A., Reviews of books: a mathematical theory of evidence (1984) The AI Magazine, 5, pp. 81-83