

A Machine Learning Approach to Personal Pronoun Resolution in Turkish

Savaş Yıldırım, Yılmaz Kılıçaslan

Istanbul Bilgi University and Trakya University
Istanbul, Turkey
savasy@cs.bilgi.edu.tr, yilmazk@trakya.edu.tr

Abstract

In this paper, we present a machine learning based approach for estimating antecedents of anaphorically used personal pronouns in Turkish sentences using a decision tree classification technique coupled with the ensemble learning method. The technique learns from an annotated corpus, which has been compiled mostly from various popular child stories.

Introduction

Up until today, all computational studies on anaphora resolution in Turkish have used knowledge-based methods. Tin et al. (1998) described an approach to the problem of pronominal anaphora in Turkish. Their BABY-SIT was a computational medium based on situation theory (Barwise and Perry 1983), which was intended to deal with anaphora resolution in Turkish along with many other aspects of the language. Two other computational models for anaphora resolution in Turkish were proposed by Yıldırım, Kılıçaslan, and Aykaç (2004) and Tüfekçi and Kılıçaslan (2005). While the former is based on Grosz et al.'s (1995) Centering Theory, the latter is an implementation of a version of Hobbs' (1986) naïve algorithm adapted for Turkish.

The approach presented in this paper resolves personal pronouns in Turkish texts. Using the decision tree approach yields a human-readable model from which we can interpret and understand the solution. We use the J48 induction system along the InforSense data mining software, which combines integrative analytics, knowledge management, large-scale data analysis and visualization. We adopt a corpus-based learning approach to personal pronoun resolution. This technique requires a training corpus and a testing corpus that have been annotated with necessary features and relationships between personal pronouns and their antecedents. The training examples are used as inputs to a learning algorithm to train or build a classifier which decides whether or not a pronoun refers to a referring expression, given a description of a possible referring expression and a personal pronoun. The model is then tested against a set of test examples which is distinct

from the training set. Afterwards, the classifier technique is reinforced with one of the most popular ensemble learning methods, boosting (Schapire 2002). Our presentation culminates with an evaluation of the results.

Feature Vector

Knowledge-based studies have led us to select relevant attributes at this step. Some studies reveal that the grammatical function plays an important role for anaphora resolution. Brennan et al. (1987) suggest the following ranking list: subject > object > object2 > other subcategorized object > adjunct. Turan (1996) emphasizes the significance of some case values, such as accusative and nominative, for the problem of anaphora resolution in Turkish. She also underlines the effect of overtness of pronouns on this process. In addition to these attributes, we also include the animacy type in our list. Another important factor is the distance between a pronoun and a candidate.

Building a classifier

All experiments were performed using InforSense. We created a decision tree using the Weka - J48 classifier with the following settings:

- Reduced Error Pruning
- Minimum number of instance: 2
- Validation method: Cross Validation
- Number of fold for validation: 10

According to cross validation, the overall error rate is 17.41% and the accuracy rate is 82.59% for the first default setting. For our model, the recall, precision and f-measure scores are 0.59, 0.68 and 0.63, respectively. In order to get to better performance results, we have incorporated the boosting ensemble method into our model.

Boosting

Combining several classifiers increases accuracy a great deal. Ensemble learning methods (Dietterich, 2000) allow for a combination of models to be built from a base learner. This base learner is first used for training several models over re-sampled or re-weighted training instances. The models are then combined into a single classifier that makes predictions by taking a vote of its constituent models.

After applying the boosting method to our model, we got the confusion matrix shown in Table 1.

	Predicted as No	Predicted as Yes
Actual Class: No	1450	127
Actual Class: Yes	163	385

Table 1: Confusion Matrix

We had better results compared to the previous ones: 0.70 (recall), 0.75 (precision), and 0.72 (f-measure). According to cross validation, the overall error rate is 13.64% and the accuracy rate is 86.36%.

Evaluation

Connolly et al. (1997) propose a machine learning model with an F-measure of 57.2. Ng and Cardie (2002) presented a noun phrase coreference system. Their best f-measure value is 70.4 on the MUC-6 coreference data set. Soon et al. (2001) presented a resolution of noun phrase anaphora in unrestricted texts with the value of f-measure being 62.6 on the MUC-6 data set. Aone and Bennet (1995) applied a learning method to Japanese texts. Their f-measure result is 77.42.

In fact, one cannot go into decisive conclusions about the question of whether or not our results are satisfactory compared to the results given above. The fact that the studies concern distinct corpora from different languages would make any such comparison at best hardly reliable. However, we can at least suggest that the proposed approach is reasonably acceptable for personal pronoun resolution in Turkish.

Conclusion

The machine learning approach discussed in this paper is based on a decision tree classification technique, which learns from an annotated corpus. It is intended to identify a unique antecedent for each personal pronoun. Using reduced error pruning, keeping the minimum number of instances as 2, and using cross validation with 10 folds, we trained a model with an f-measure of 0.63. In order to increase the accuracy rate we used boosting, an ensemble learning method. Having used this method, our f-measure value rose up to 0.72. When examining the big picture which involves other studies as well, this seems to be a promising score for the problem of anaphora resolution in Turkish.

References

- Aone, Chinatsu and Bennett, Scott W. 1995. Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies. In Proceedings of the 33th Annual Meeting of the Association for Computational Linguistics, 122-129. ACL.
- Barwise, J. and Perry, J. 1983. *Situation and Attitudes*, MIT Press, Cambridge, MA.
- Brennan, Susan; Friedman, Marilyn Walker; and Carl Pollard. 1987. A Centering Approach to Pronouns in Proceedings of the 25th Annual Meeting Of ACL, Stanford.
- Connolly, Dennis; Burger, John D. and Day, David S. 1997. A machine learning approach to anaphoric reference. New Methods in Language Processing, 133-144.
- Dietterich, Tomas G. 2000. Ensemble methods in machine learning. In First International Workshop on Multiple Classifier Systems, Lecture notes in Computer Science, 1-15 , New York.
- Grosz, Barbara; Joshi, Aravind; and Weinstein, Scott. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. Computational Linguistics.
- Hobbs, Jerry. 1986. Resolving Pronoun References. Readings in Natural Language Processing. California: Morgan Kaufman Publishers Inc.
- Ng, Vincent and Cardie, Claire. 2002. Improving machine learning approaches to coreference resolution. In proceedings of the 40rd Annual Meeting of the Association for Computational Linguistics, 104-111.
- Schapire, Robert E. 2002. The boosting approach to machine learning: an overview. In Proceedings of the MSRI Workshop on Nonlinear Estimation and Classification.
- Soon, Wee Meng; Yong, Daniel Chung Yong and Ng, Hwee Tou. 2001. A Machine Learning Approach to Coreference Resolution of Noun Phrases. In Computational Linguistics 27(4), 521-544.
- Tin, Erkan and Varol Akman. 1998. Situated Processing of Prenominal Anaphora. Bilkent University, Ankara.
- Turan, Ümit Deniz. 1996. Null Vs. Overt Subjects in Turkish Discourse: A Centering Analysis. Ph.D. Dissetation. University pf Pennsylvania, Philadelphia.
- Tüfekçi, Pınar and Kılıçaslan, Yılmaz 2005: A Computational Model for Resolving Prenominal Anaphora in Turkish Using Hobbs' Naïve Algorithm. WEC (5) 2005: 13-17
- Yıldırım, Savaş; Kılıçaslan, Yılmaz; and Aykaç, R. Erman.2004. A Computational Model for Anaphora Resolution in Turkish via Centering Theory: an Initial Approach. International Conference on Computational Intelligence. 124-128