OSU-2: Generating Referring Expressions with a Maximum Entropy Classifier

Emily Jamison
Department of Linguistics
The Ohio State University
Columbus, OH 43210, USA
jamison@ling.osu.edu

Dennis Mehay
Department of Linguistics
The Ohio State University
Columbus, OH 43210, USA
mehay@ling.osu.edu

Abstract

Selection of natural-sounding referring expressions is useful in text generation and information summarization (Kan et al., 2001). We use discourse-level feature predicates in a maximum entropy classifier (Berger et al., 1996) with binary and n-class classification to select referring expressions from a list. We find that while mention-type n-class classification produces higher accuracy of type, binary classification of individual referring expressions helps to avoid use of awkward referring expressions.

1 Introduction

Referring expression generation is the task of inserting noun phrases that refer to a mentioned extra-linguistic entity into a text. REG is helpful for tasks such as text generation and information summarization (Kan et al., 2001).

2 Task Description

The Referring Expressions Generation Challenge (Belz and Gatt, 2008) includes a task based on the GREC corpus, a collection of introductory texts from Wikipedia that includes articles about cities, countries, rivers, people, and mountains. In this corpus, the main topic of each text (MSR) has been replaced with a list of possible referring expressions (REs). The objective of the task is to identify the most appropriate referring expression from the list for each mention of the MSR, given the surrounding text and annotated syntactic and semantic information.

3 Predicates

We created 13 predicates, in addition to the six predicates available with the corpus. All predicates can be used with the binary classification method; only non-RE-level predicates can be used with the n-class classification method. Predicates describe: string similarity of the RE and the title of the article, the mention's order in the article, distance between previous mention and current mention, and detection of a contrastive discourse entity in the text.[1]

4 Maximum Entropy Classifier

We defined indicator feature functions for a number of contextual predicates, each describing a pairing of some potential property of the syntactico-semantic and discourse context of a RE (a 'predicate') and a label. These feature functions f_i were used to train a maximum entropy classifier (Berger et al., 1996) (Le, 2004) that assigns a probability to a RE re given context c_x as follows:

$$p(re \mid c_x) = Z(c_x) \exp \sum_{i=1}^{n} \lambda_i f_i(c_x, re)$$

where $Z(c_x)$ is a normalizing sum and the λ_i are the parameters (feature weights) learned. Two classification systems were used: binary and n-class. With the binary method, the classifier estimates the likelihood of a possible referring expression's correct insertion into the text, and inserts the RE with the highest 'yes' probability. With the n-class method,

---

[1]More details at http://www.ling.ohio-state.edu/~jamison

Predicates Used	Single	Combinations
GREC predicates	40.40%	50.91%
all predicates	50.30%	58.54%
no contrasting entities	50.30%	59.30%
all non-RE-level preds	44.82%	51.07%

Table 1: Results with binary classification.

Predicates Used	Single	Combinations
all non-RE-level preds	61.13%	62.50%

Table 2: Results with n-class classification.

the mention is classified according to type of referring expression (proper name, common noun, pronoun, empty) and a RE of the proper type is chosen.

A predicate combinator was implemented to create pairs of predicates for the classifier.

5 Results

Our results are shown in tables 1 and 2; table 3 shows further per-category results. N-class classification has a higher type accuracy than the binary method(single: 61.13% versus 44.82%). Added predicates made a notable difference (single, original predicates: 40.40%; with added predicates: 50.30%). However, the predicates that detected contrasting discourse entities proved not to be helpful (combinations: 59.30% declined to 58.54%). Finally, the predicate combinator improved all results (binary, all predicates: 50.30% to 58.54%).

6 Discussion

The n-class method does not evaluate characteristics of each individual referring expression. However, the accuracy measure is designed to judge appropriateness of a referring expression based only on whether its type is correct. A typical high-accuracy n-class result is shown in example 1.

System	City	Ctry	Mnt	River	Pple
b-all	53.54	57.61	49.58	75.00	65.85
b-nonRE	51.52	53.26	45.83	40.00	57.07
n-nonRE	53.54	63.04	61.67	65.00	67.32

Table 3: Challenge-submitted results by category.

Example 1: Albania
The Republic of Albania itself is a Balkan country in Southeastern Europe. **Which itself** borders Montenegro to the north, the Serbian province of Kosovo to the northeast, the Republic of Macedonia in the east, and Greece in the south.

In example 1, both mentions are matched with an RE that is the proper type (proper name and pronoun, respectively), yet the result is undesireable.

A different example typical of the binary classification method is shown in example 2.

Example 2: Alfred Nobel
Alfred Nobel was a Swedish chemist, engineer, innovator, armaments manufacturer and the inventor of dynamite. [...] In **his** last will, **Alfred Nobel** used **his** enormous fortune to institute the Nobel Prizes.

In example 2, the use of predicates specific to each RE besides the type causes use of the RE "Alfred Nobel" as a subject, and the RE "his" as a possessive pronoun. The text, if mildly repetitive, is still comprehensible.

7 Conclusion

In this study, we used discourse-level predicates and binary and n-class maximum entropy classifiers to select referring expressions. Eventually, we plan to combine these two approaches, first selecting all REs of the appropriate type and then ranking them.

References

Anya Belz and Albert Gatt. 2008. *REG Challenge 2008: Participants Pack.* http://www.nltg.brighton.ac.uk/research/reg08/.

A. L. Berger, S. D. Pietra, and V. D. Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistcs*, 22(1):39–71.

Min-Yen Kan, Kathleen R. McKeown, and Judith L. Klavans. 2001. Applying natural language generation to indicative summarization. *EWNLG '01: Proceedings of the 8th European workshop on Natural Language Generation.*

Zhang Le. 2004. *Maximum Entropy Modeling Toolkit for Python and C++.* http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html.