ICSI-CRF: The Generation of References to the Main Subject and Named Entities using Conditional Random Fields

Benoit Favre and **Bernd Bohnet**
International Computer Science Institute
1947 Center Street. Suite 600
Berkeley, CA 94704, USA
{favre|bohnet}@icsi.berkeley.edu

Abstract

In this paper, we describe our contribution to the Generation Challenge 2009 for the tasks of generating Referring Expressions to the Main Subject References (MSR) and Named Entities Generation (NEG). To generate the referring expressions, we employ the Conditional Random Fields (CRF) learning technique due to the fact that the selection of an expression depends on the selection of the previous references. CRFs fit very well to this task since they are designed for the labeling of sequences. For the MSR task, our system has a String Accuracy of 0.68 and a REG08-Type Accuracy of 0.76 and for the NEG task a String Accuracy of 0.79 and REG08-Type Accuracy of 0.83.

1 Introduction

The GREC Generation Challenge 2009 consists of two tasks. The first task is to generate appropriate references to an entity due to a given context which is longer than a sentence. In the GREC-MSR task, data sets are provided of possible referring expressions which have to be selected. In the first shared task on same topic (Belz and Varges, 2007), the main task was to select the referring expression type correctly. In the GREC-MSR 2009 task, the main task is to select the actual word string correctly, and the main evaluation criterion is String Accuracy.

The GREC-NEG task is about the generation of references to all person entities in a context longer than a sentence. The NEG data also provides sets of possible referring expressions to each entity ("he"),

groups of multiple entities ("they") and nested references ("his father").

2 System Description

Our approach relies in mapping each input expression for a given reference to a class label. We use the attributes of the REFEX tags as basic labels so that, for instance, a REFEX with attributes REG08-TYPE="pronoun" CASE="nominative" is mapped to the label "nominative_pronoun". In order to decrease the number of potential textual units for a predicted label, we derive extra label information from the text itself. For instance a qualifier "first_name" or "family_name" is added to the expressions relative to a person. Similarly, types of pronouns (he, him, his, who, whose, whom, emphasis) are specified in the class label, which is very useful for the NEG task. Only the person labels have been refined this way. While we experimented with a few approaches to remove the remaining ambiguity (same label for different text), they generally did not perform better than a random selection. We opted for a deterministic generation with the last element in the list of possibilities given a class label.

For prediction of attributes, our system uses Conditional Random Fields, as proposed by (Lafferty et al., 2001). We use chain CRFs to estimate the probability of a sequence of labels ($Y = Y_1 \ldots Y_n$) given a sequence of observations ($X = X_1 \ldots X_m$).

$$P(Y|X) \propto exp\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\lambda_i f_i(Y_{j-1}, Y_j, X)\right) \quad (1)$$

Here, $f_i(\cdot)$ are decision functions that depend on

Evaluation Metric	MSR						NEG					
	R1	**R2**	**S1**	**S2**	**S2R**	**S2O**	**R1**	**R2**	**S1**	**S2**	**S2R**	**S2O**
REG08 Type Accuracy	0.36	1.00	0.74	**0.75**	0.75	0.75	0.40	1.00	0.83	**0.83**	0.83	0.83
String Accuracy	0.12	0.82	0.62	**0.67**	0.66	0.75	0.12	0.70	0.52	**0.79**	0.79	0.80
Mean Edit Distance	2.52	0.31	0.95	0.85	0.87	0.72	2.38	0.61	1.07	0.53	0.52	0.49
Mean Norm. Edit Dist.	0.79	0.09	0.31	0.28	0.28	0.24	0.84	0.22	0.43	0.19	0.20	0.19
BLEU 1	0.19	0.88	0.65	0.69	0.68	0.74	0.17	0.79	0.64	0.81	0.81	0.83
BLEU 2	0.14	0.76	0.55	0.60	0.59	0.71	0.18	0.75	0.69	0.83	0.83	0.85
BLEU 3	0.10	0.69	0.51	0.56	0.55	0.70	0.18	0.73	0.71	0.83	0.84	0.86

Table 1: Results for the GREC MSR and NEG tasks. Are displayed: a random[2] output (R1), a random output when the attributes are guessed correctly (R2), the CRF system predicting basic attributes (S1), the CRF system predicting refined attributes (S2), CRF-predicted attributes with random selection of text (S2R) and CRF-predicted attributes with oracle selection of text (S2O).

the examples and a clique of boundaries close to Y_j, and λ_i is the weight of f_i estimated on training data. For our experiments, we use the CRF++ toolkit,[1] which allows binary decision functions dependent on the current label and the previous label.

All features are used for both MSR and NEG tasks, where applicable:

- word unigram and bigram before and after the reference
- morphology of the previous and next words (-ed, -ing, -s)
- punctuation type, before and after (comma, parenthesis, period, nothing)
- SYNFUNC, SYNCAT and SEMCAT
- whether or not the previous reference is about the same entity as the current one
- number of occurrence of the entity since the beginning of the text (quantized 1,2,3,4+)
- number of occurrence of the entity since the last change of entity (quantized)
- beginning of paragraph indicator

In the MSR case, this list is augmented with the features of the two previous references. In the NEG case, we use the features of the previous reference and those of the previous occurrence of the same entity.

3 Results and Conclusion

Table 1 shows the results for the GREC MSR and NEG tasks.[2] We observe that for both tasks, our system exceeds the performance of a random[3] selection (columns R1 vs. S2). In the MSR task, guessing correctly the attributes seems more important than in the NEG task, as suggested by the difference in string accuracy when randomly selecting the references with the right attributes (columns R2). Generating more specific attributes from the text is especially important for the NEG task (columns S1 vs. S2). This was expected because we only refined the attributes for person entities. We also observe that a deterministic disambiguation of the references with the same attributes is not distinguishable from a random selection (columns S2 vs. S2R). However it seems that selecting the right text, as in the oracle experiment, would hardly help in the NEG task while the gap is larger for the MSR task. This shows that refined classes work well for person entities but more refinements are needed for other types (city, mountain, river...).

References

J. Lafferty, A. McCallum, and F. Pereira. 2001. *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.* In Proc. of ICML, pages 282-289

A. Belz and S. Varges. 2007 *Generation of Repeated References to Discourse Entities.* In Proceedings of the 11th European Workshop on Natural Language Generation (ENLG07), pages 9-16.

---

[1] http://crfpp.sourceforge.net/

[2] Our system is available http://www.icsi.berkeley.edu/~favre/grec/

[3] All random experiments are averaged over 100 runs.