Building Patterns for Biomedical Event Extraction

Hong-Woo Chun[1]
chun@is.s.u-tokyo.ac.jp

Tomoko Ohta[1,2]
okap@is.s.u-tokyo.ac.jp

Jin-Dong Kim[1,2]
jdkim@is.s.u-tokyo.ac.jp

Jun'ichi Tsujii[1,2]
tsujii@is.s.u-tokyo.ac.jp

[1] Tsujii Laboratory, Room 615, 7th Building of Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-0033, Japan
[2] CREST, Japan Science and Technology agency, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

Keywords: biomedical relationship, information extraction, event extraction

1 Method

Generally, *Event Extraction* is to identify any instance of a particular class of events in a natural language text, to extract the relevant arguments of the event, and to represent the extracted information into a structured form.[1]Let us define *Event* on the binary relation between two entities for special event verbs which are predefined by biologists. Here, *Entity* means biomedical entities such as proteins, genes, cells, tissues, etc. According to the definition of event, our event extraction system considers only such sentences which contain at least one event verb and two entities.

The training consists of two procedures (Figure 1). First, the preprocessor involves chunking, named entity tagging, dependency relation tagging and sentence normalization with special items for building patterns. Special items are entities, event verbs, non-event verbs, prepositions, relatives, conjunctions and symbols. Second, all possible candidate events are extracted from the training corpus and the corresponding patterns are also generated. At this time, we utilize the following assumptions: one event can be described by one or more patterns in the whole documents and one pattern also can be generated by one or more events. Therefore, the event and the pattern information has reciprocal relation. We use the event score (Equation 1) to measure the reliability of extracted events and the pattern score (Equation 2) to measure the reliability of extracted patterns. The scores are iteratively updated in a co-updating method. Updating the event score causes reranking of candidate events and the iteration is continued until the ranking of events is no longer changed. The result of the training is a set of generated patterns and their scores. The events in training corpus are also extracted as the by-product of the training.

$$EventScore(E) = [\alpha Cooccurrence + \beta Dependency\ Relation] \times [Average\ of\ PatternScore(P^E)] \quad (1)$$

$$PatternScore(P) = \frac{\sum_{EventScore(E_i^P) > \delta} freq(E_i^P)}{\sum_{E_j^P} freq(E_j^P)} \quad (2)$$

[1]following the definition from MUC-7

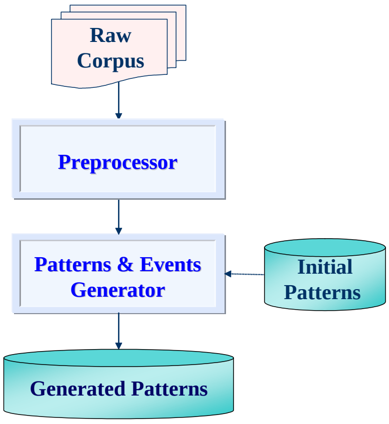

Figure 1: System architecture

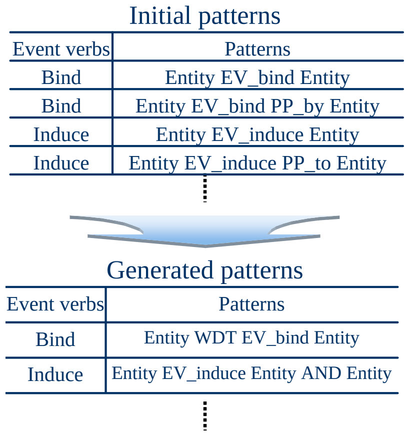

Figure 2: Generation of patterns

2 Experimental Results

For the experiment, we considered 157 event verbs that were chosen by biologists and constructed 433 initial patterns considering prepositions information corresponding to the event verbs(Figure 2). After that, we generated patterns by using a training corpus. Most part of the GENIA corpus[2]containing 18,545 sentences were used for the training purpose except 200 sentences that were reserved for testing. Table 1 shows the number of the generated patterns which are generated as the size of training corpus increases. Among these patterns, we used only patterns whose score was over a threshold[3].

To evaluate performance of the generated patterns, we tried to extract events by using the patterns from the test set of 200 sentences. At this time, 99 sentences that do not have at least one event verb and two entities were simply filtered out. A biologist checked all the remaining 101 sentences and revealed that there were 75 events. Figure 3 shows the precision and recall curves according to the size of the training corpus. The results are encouraging. By generating the patterns, we got progress significantly in terms of recall without loss of precision. As the improvement may not be satisfactory, it was achieved at the minimal cost: the proposed method requires no manually annotated corpus.

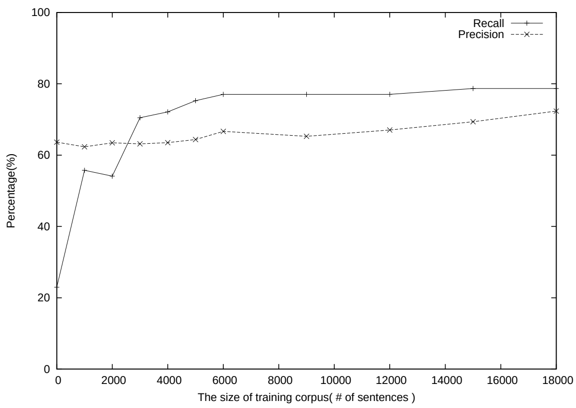

Figure 3: Precision and recall curves according to the size of training corpus

Table 1: The number of generated patterns according to the size of traing corpus

Training(sentences)	# of Patterns	# of Patterns(≥ 0.3)
0	433	433
1000	669	667
2000	978	943
3000	1174	1137
6000	1790	1721
9000	2528	2424
12000	3245	3099
15000	3945	3763
18000	4700	4464

[2] http://www-tsujii.is.s.u-tokyo.ac.jp/genia/topics/Corpus/3.0/GENIA3.0p.intro.html

[3] The threshold was determined empirically to be 0.3.