

## Korean Named Entity Recognition using HMM and CoTraining Model

Euisok Chung, Yi-Gyu Hwang, Myung-Gil Jang  
Electronics and Telecommunications Research Institute  
161, Kajong-Dong, Yusong-Gu, Daejeon, 305-350, KOREA  
 [{eschung,yghwang,mgiang}@etri.re.kr](mailto:{eschung,yghwang,mgiang}@etri.re.kr)

### ABSTRACT

*Named entity recognition is important in sophisticated information service system such as Question Answering and Text Mining since most of the answer type and text mining unit depend on the named entity type. Therefore we focus on named entity recognition model in Korean. Korean named entity recognition is difficult since each word of named entity has not specific features such as the capitalizing feature of English. It has high dependence on the large amounts of hand-labeled data and the named entity dictionary, even though these are tedious and expensive to create. In this paper, we devise HMM based named entity recognizer to consider various context models. Furthermore, we consider weakly supervised learning technique, CoTraining, to combine labeled data and unlabeled data.*

**Keywords** : Korean Named Entity, HMM, Co-Training

### 1. Introduction

Named entity(NE) recognition is important for recent sophisticated information service such as question answering and text mining since it recognizes the words to present the core information in text. In particular, the NE recognizer is important module in the well-known question answering systems such as FALCON, IBM[3][10]. NE recognizer is well suited for the recognition of answer type which can be equal to the NE type or not. Although an answer is not exactly matched to the NE, these two types can be mapping to each other by using WordNet[3].

NE recognition can be explained with two steps, NE detection and NE classification. Whereas NE detection is to catch the named entities in the text, NE classification is to classify NE into person, organization or location. In Korean, NE detection is difficult since each word of name entity has not specific features such as the capitalizing feature of English. It has high dependence on the large amounts of

hand-labeled data and the NE dictionary, even though these are tedious and expensive to create.

In the case of NE classification, NE can be classified with the clues such as inner word and context word. Although these clue words present the feature of NE type, it can be used in detecting the NE since the contained word and context word can be used in determining the boundary of NE. However, the clue words can provoke ambiguity to determine the NE type since various NEs can share the same clue word. Therefore, we devise the statistical model based NE recognizer which can unify the detection and classification.

Furthermore, we consider unlabeled data based statistical learning to extend the initial seed data. The weakly supervised learning technique is Co-Training method. In this paper, we describe the HMM based Korean NE recognition and Co-Training method for HMM based boosting.

### 2. The Problem

NE dictionary is not enough to cover all of the NEs since there are a few types of NEs besides single word. We classify Korean NE into three types. The first is single word type, the second is compound noun type, and the third is noun phrase type. The single word type is usually single noun. The second type, compound noun, is composed of a few words and affix. The third type can have grammatical morphemes besides nouns. The example is described in Figure 1. It describes NEs with PERSON(PER), LOCATION (LOC), and ORGANIZATION(ORG). In each type, second sentence is the result of morphological analysis.

The example shows the diversity of the Korean NE type. In the single word type, if the dictionary has the word, then the NE could be detected easily. However, compound noun type and noun phrase type require a tremendous number of entries in the dictionary. Moreover, in Korean these kinds of NE types are used differently in each people. Therefore, it

should use the clue word to recognize NE which is shown as the compound noun or noun phrase type. This clue word, which is context or inner word of NE, can be used in NE detection and classification, but we found that the clue word can provoke another problem which is the ambiguity; thus different types of NEs can share the same clue word. Which means that the clue word dictionary cannot be the unique solution.

|   |                                   |
|---|-----------------------------------|
| <u>Single word type</u>                               |                                   |
| <ORG>   | <ORG> <PER>                       |
| <PER>   |                                   |
| <ORG> /nc /sn <ORG> <PER> /nc /nc <PER>               | /nc /jc                           |
| <ORG>jeonnam dae<ORG>                                 | <PER>kim - yeong yong<PER> nun    |
| <u>Compound noun type</u>                             |                                   |
| - 11 <LOC>  | <LOC>                             |
| 11/nn /nb <LOC>                                       | /nc /nc /nc /nc <LOC> /jc         |
| 11 il <LOC>seoul hannamdong hayat grandvolum<LOC>     | -eseo                             |
| <u>Noun phrase type</u>                               |                                   |
| <LOC>   | <LOC>                             |
| /nc <LOC>   | /nc /jm /nc /pv /em /nc <LOC> /jc |
| cafe <LOC>syagal - eui nun - naeri - neun maeul <LOC> | -eseo                             |

Figure 1. Named entity example<sup>1</sup>

Consequently, we suggest three approaches for NE recognition. The first is feature dictionary based approach which classify the clue words and generate feature types of the context or inner word of NE. The second is statistical approach which needs named entity tagged corpus and context model to recognize NE. The third is unlabeled data based boosting approach since statistical approach, which needs hand-labeled NE tagged corpus, cannot avoid data sparseness.

### 3. Related Works

Statistical approach in NE recognition can be classified into supervised learning and weakly supervised learning. Supervised learning is based on labeled data. On the other hand, weakly supervised method is the learning approach to combine labeled data and unlabeled data. From the supervised learning point of view, the most representative research is HMM based NE recognition. It builds various

<sup>1</sup> nc(noun), pv(verb), pa(adjective), px(auxiliary verb), co(copula), mag(general adverb), maj(conjunctive adverb), mm(adnoun), xsn(noun -derivational suffix), xsv(verb-derivational suffix), xsm (adjective -derivational suffix), jc(case particle), jx(auxiliary particle), jj(conjunctive particle), jm(adnominal Case Particle), ep(Prefinal Ending)

bigram models of NEs and predicts next NE type with the previous history, lexical item and NE type. Using simple word feature Bikel shows F-measure 90% in English [4]. Zhou's HMM-based chunk tagger approach adopt more detailed feature than Bikel's, and show F-measure 94.3%[5]. In this paper, when we designed feature model, we followed the HMM-based chunk tagger approach considering the property of Korean NE.

Recently there have been many researches in weakly supervised learning technique to combine labeled data and unlabeled data. Co-Training method Blum is most famous approach to boost the initial learning data with unlabeled data[2]. Blum showed that using a large unlabeled data to boost performance of a learning algorithm could be used to classify web pages when only a small set of labeled examples is available [2]. Nigam demonstrated that when learning from labeled and unlabeled data, algorithms explicitly leveraging a natural independent split of the features outperform algorithms that do not[7][8]. Collins and Singershowed that the use of unlabeled data can reduce the requirements for supervision to just 7 simple "seed" rules [9]. In addition to a heuristic based on decision list learning, they also presented a boosting-like framework that builds on ideas from Blum[2].

## 4. Named entity recognizer

In this paper, we propose the Korean NE recognizing methodology. It is based on the feature model, HMM based statistical model and Co-Training based boosting model.

### 4.1 Feature model

NE recognition depends on various clues to distinguish each type. The inside and outside properties of NE could be the clues which can be composed of a few clue class. The class consists of 'character feature', 'named entity dictionary', 'inner word', and 'context word'. It is described in Table 1. (1) 'character feature' shows the digit or Chinese feature in Korean NE. Digit feature is used to recognize MONEY, TIME, and others. (2) 'named entity dictionary' means that we should build NE dictionaries. The dictionaries are composed of DATE, PERSON, LOCATION, and ORGANIZATION. (3) 'inner word' consists of suffix word and constituent word of NE. (4) 'context word' is the word set adjacent to NE. These feature values are constructed manually and are used with history in annotating NE type. As stated above, all the feature classes have ambiguity since one feature word can be another feature type. Therefore, we cannot recognize NE with only feature dictionaries.

|                         | Type         | Feature value  |
|-------------------------|--------------|--|
| character feature       | DIGIT        | OneDigitNumber, TwoDigitNumber, FourDigitNum                       |
|                         | DIGIT&LETTER | ContainsDigitAndPeriod, ContainsDigitAndLetter                     |
|                         | CHINESE      | OneChinese, ThreeChinese, ContainsOneChineseAndLetter              |
|                         | ALPHABET     | ContainsAlphaAndLetter, AllCapitalization                          |
|                         | LETTER       | TreeLetters  |
| named entity dictionary | DATE         | DicDATE  |
|                         | PERSON       | DicPERSON  |
|                         | LOCATION     | DicLOC   |
|                         | ORGANIZATION | DicORG   |
| inner word              | PERCENT      | SuffixPERCENT  |
|                         | MONEY        | SuffixMONEY, SuffixCURRENCY  |
|                         | TIME         | SuffixTIME, PeriodTIME   |
|                         | DATE         | SuffixDATE, WeekDATE, SeasonDATE, PeriodDATE, YearDATE, OthersDATE |
|                         | LOCATION     | DistrictSuffixLOC, SuffixLOC                                       |
| context word            | ORGANIZATION | SuffixORG  |
|                         | PERSON       | PositionPERSON, RelationPERSON, JobPERSON                          |
|                         | LOCATION     | ClueLOC  |
|                         | DATE         | ClueDATE   |
|                         | TIME         | ClueTIME   |
|                         | PHONE        | CluePHONE  |
|                         | ORGANIZATION | ClueORG  |
|                         | PERCENT      | CluePERCENT  |
|                         | MONEY        | ClueMONEY  |
|                         | ADDRESS      | ClueADDRESS  |
|                         | QUANTITY     | ClueQUANTITY   |

Table 1. Feature type

## 4.2 HMM based statistical model

NE recognizer should adopt statistical model since it is impossible to construct the dictionary having all of the NE entries and moreover, clue word dictionary can provoke ambiguity. For instance, proper noun such as person name creates everyday. It is unknown word problem. In the feature model, some feature classes can share the same clue words. Simply, both DATE and TIME can have the DIGIT character feature. Therefore, we adopt statistical model for NE recognition. For the statistical approach, we construct NE tagged documents, design NE context model, and suggest forward-backward view based boosting approach.

We build 300 named entity tagged documents in the newspaper article domains such as economy, performance and travel. The labeled data is tagged by using tagging tools. We attached only NE tags to the text and do not consider morphological information because Korean morphological tag is various to the analyzer. Furthermore, we build statistical information extractor to learn the NE distributional information. The labeled data is used in constructing NE statistical data.

To build NE context model for statistical NE recognizer, we analyze 201 NEs in labeled documents. From the three points of view, word feature, inner word feature, and context

word feature, we analyze NE examples. From the analysis, word feature can be classified into single word and compound word. Inner feature has property of full string and inner word. From the context word feature, we recognized three kinds of features such as root, adjacent morpheme and no context. The result of analysis is described in Table 2. Finally, we can build four types of NE context model such as lexical model, feature model, POS(part of speech) model and root model.

|                      |  |     |        |
|----------------------|--|-----|--------|
| Word feature         | Single Noun                            | 129 | 64.2 % |
|                      | Compound Noun                          | 72  | 35.8 % |
| Inner Word Feature   | full string                            | 118 | 58.7 % |
|                      | Inner word                             | 83  | 41.3 % |
| Context Word Feature | Left root of a word                    | 14  | 7.0 %  |
|                      | Right root of a word                   | 64  | 31.8 % |
|                      | Morpheme left adjacent to name entity  | 38  | 18.9 % |
|                      | Morpheme right adjacent to name entity | 112 | 55.7 % |
|                      | no context                             | 35  | 17.4 % |

Table 2. Named entity analysis

Whereas the rule-based approach needs enormous hand-crafted rules, statistical model has the advantages of simplicity, expansivity and robustness in the named entity model. The most representative approach of statistical model is HMM based approach. To adopt HMM based model, we define HMM state and build NE context model which can cover various NE contextual information.

For HMM based approach, HMM state should be defined. HMM state is the type of NE constituent. Thus, S\_LOC is the state which can be the first lexical item of LOC typed named entity. C\_LOC is the middle state of the NE. E\_LOC is the final morpheme. In the case of U\_LOC, single word is the NE word. For example, location name "Inchon International Airport" can be tagged with "Inchon/S\_LOC International /C\_LOC Airport/ E\_LOC" and another named entity "Seoul" can be labeled with "Seoul/U\_LOC". The HMM state is described in Table 5.

The NE context model is composed of four types such as morp(morphology), root, POS and feature. Through this NE context model, the statistical data is learned from the tagged corpus, and is used in computing probability of predicting HMM state to lexical item. In the case of feature type, "Indiana" has "DicLOC" since it is discovered in LOC dictionary and "professor" is allocated with "Position-PERSON" since it is used with position clue word. The context model and example is presented in table 3.

| Type    | (Indiana) | (Gary) | (Birth) | <-i-> | <i-> | <i+>           | nun |
|---------|-----------|--------|---------|-------|------|----------------|-----|
|         | NE_S      | NE_E   | -       | -     | NE_U | -              |     |
| POS     | Nc        | nc     | nc      | jm    | nc   | nc             | jx  |
| MORP    |           |        |         |       |      |                |     |
| ROOT    | Root      | Root   | Root    | -     | Root | Root           | -   |
| FEATURE | DicLOC    | -      | -       | -     | -    | PositionPERSON | -   |

Table 3. Context model type

| Type | View type | Statistical Model   | Example  |
|------|-----------|---|--|
| MORP | Forward   | $\Pr(s_i   s_{i-1}, m_{i-1})$<br>$\times \Pr(m_i   s_i, m_{i-1})$ | $c(-, NE\_U, eui) / c(-, eui)$<br>$\times c(NE\_U, eui, Stigritz) / c(NE\_U, eui)$                         |
|      | Backward  | $\Pr(s_i   s_{i+1}, m_{i+1})$<br>$\times \Pr(m_i   s_i, m_{i+1})$ | $c(NE\_U, -, professor) / c(-, professor)$<br>$\times c(NE\_U, Stigritz, professor) / c(NE\_U, professor)$ |

Table 4. Forward/Backward model for CoTraining

| NE type      | HMM state                                  |
|--------------|--|
| PERSON       | S_PER, C_PER, E_PER, U_PER                 |
| LOCATION     | S_LOC, C_LOC, E_LOC, U_LOC                 |
| ORGANIZATION | S_ORG, C_ORG, E_ORG, U_ORG                 |
| DATE         | S_DATE, C_DATE, E_DATE, U_DATE             |
| TIME         | S_TIME, C_TIME, E_TIME, U_TIME             |
| PERCENT      | S_PERCENT, C_PERCENT, E_PERCENT, U_PERCENT |
| MONEY        | S_MONEY, C_MONEY, E_MONEY, U_MONEY         |
| QUANTITY     | S_QUANT, C_QUANT, E_QUANT, U_QUANT         |

Table 5. HMM state for NE context model

We designed the NE context model with the divided view types such as forward/backward views which means left-right NE context view. In each view type, the probability is computed with the product of state transitional probability and lexical probability. In forward view, state transitional probability is  $\Pr(S_i | S_{i-1}, m_{i-1})$  which predict  $i^{\text{th}}$  HMM state  $S_i$  with  $(i-1)^{\text{th}}$  state  $S_{i-1}$  and morpheme  $m_{i-1}$ , and lexical probability is  $\Pr(m_i | S_i, m_{i-1})$  which predict  $i^{\text{th}}$  morpheme  $m_i$  with  $i^{\text{th}}$  state  $S_i$  and morpheme  $m_{i-1}$ .

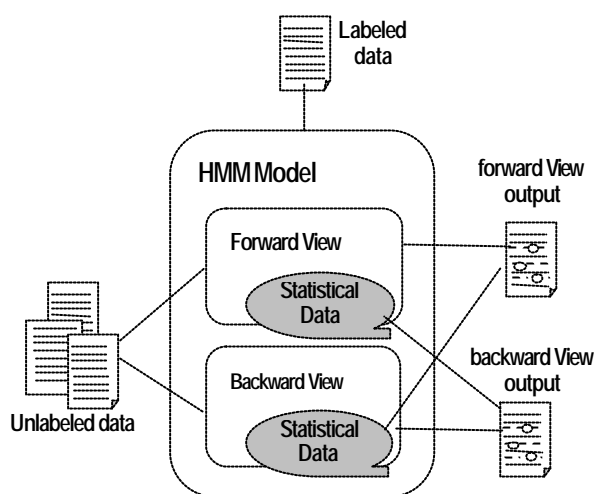
In table 4, the state of morpheme  $m_{i-1}$  “ eui”, the state  $S_i$  is “NE\_U” and next  $S_{i+1}$  is “-“. Which means that state transitional probability can be computed by  $\text{count}(-, NE\_U, eui) / \text{count}(-, eui)$ . In the case of lexical probability, it is computed by  $\text{count}(NE\_U, eui, stiglitz) / \text{count}(NE\_U, eui)$ . The difference of forward view and backward view can be explained with state transitional probability. Whereas forward view computes current state probability with pre-state, backward view computes current state probability with next-state. Thus forward view considers left contextual information. On the other hand, backward view considers right context.

For the statistical approach, we build labeled data for supervised learning and propose four typed NE context

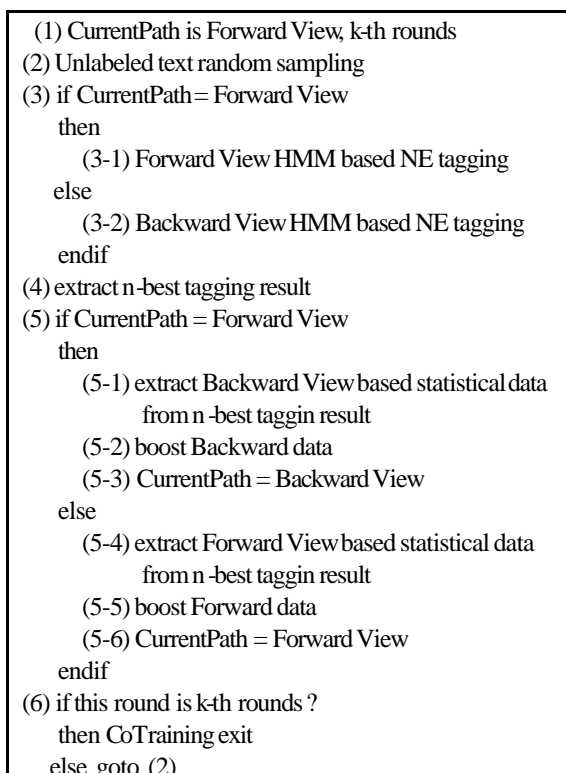
model which considers left-right contextual information. Furthermore, we adopt smoothing model based on the modified Kneser-Ney smoothing technique[6] since HMM based approach needs smoothing technique for better result. After allocating the lexical probability and state transitional probability to the HMM state which is coupled with the morpheme of sentence, the recognition is processed by Viterbi algorithm. Then, NE is recognized in the input sentence.

### 4.3 CoTraining based boosting model

Co-Training method is most famous approach to boost the initial learning data with unlabeled data. In this paper, we propose the method to apply Co-Training method to the HMM based statistical approach. The main idea is to divide view type of the context model into forward view and backward view. Simply the forward view’s output, which is the result of NE recognition to the unlabeled data, is used for input data in the backward view and vice versa. From the iteration of the Co-Training procedure, both views could boost each other, which means that the both statistical data could increase by using unlabeled data. HMM based CoTraining approach is described in Figure 2.



**Figure 2 HMM based CoTraining approach**



**Figure 3. CoTraining Procedure**

The CoTraining algorithm which boosts HMM based NE statistical model is described in Figure 3. Here the procedure of CoTraining is shown for boosting between divided statistical models. In first step, the current view type and the number of times of learning round are determined(1) since CoTraining approach, which is based on feature redundant principle, divides the learning model and reflects one learning result to the other learning model. The next step is

random sampling of unlabeled text data(2).

After random sampling, the NE statistical data should be extracted from the sampling data. At this time the next step depends on the current view type which can be Forward-View or BackwardView. If current learning view is Forward-View, the next step is forward model based NE tagging task(3-1), otherwise, the current learning view is Backward-View, the next step is backward view based task(3-2). After that, from the result of NE tagging, the  $n$ th best tagging results are selected(4) and added to the learning data.

From the first step till now, NE tagging using unlabeled data is processed, and the tagging result prepares for boosting other view type. If CurrentPath is ForwardView(5), backward view data is extracted from the tagging result(5-1) and boost backward view data(5-2), and then CurrentPath change to BackwardView(5-3). Otherwise, CurrentPath is BackwardView(5), forward view data is extracted from the tagging result(5-4) and boost forward view data(5-5), and then CurrentPath change to ForwardView(5-6). Finally, the round is checked whether it is over the pre-defined iteration time or not(6). If it pass the time, the procedure ends, otherwise the random sampling step repeats.

## 5. Experiments

We evaluate named entity recognition with two kinds of experiments. One is the performance of named entity recognition which unified morp model and feature model to learn statistical information. The other experiment is about CoTraining performance.

### 5.1 Name d Entity Recognition Test

We evaluate NE recognition with morp model and feature model since POS statistical data, which is extracted from labeled corpus, has some POS tagging error and root model cannot be implemented due to the difficult to determine what the root is. Therefore, in this paper we suggest the evaluation of the morp model, feature model and morp/feature model considering forward/backward view: (1) morp model based forward view [M/F], (2) morp model based backward view [M/B], (3) morp/feature model based forward view [MF/F], (4) morp/feature model based backward view [MF/B], (5) morp/feature model based forward/backwardview [MF/FB], which combination of forward/backward view is based on forward-backward algorithm. We give 0.93 weight to morp model and give 0.07 weight to feature model. It is optimized from many experiments.

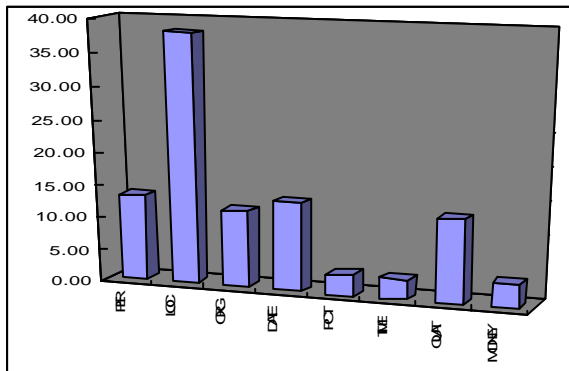
With 300 tagged documents, we train the recognizer with 270 documents which is compose of 90 economy, 90

|       | Num of Test Doc. | NAME ENTITY TYPE |          |              |       |         |       |          |       | TOTAL     |        |           |
|-------|------------------|------------------|----------|--------------|-------|---------|-------|----------|-------|-----------|--------|-----------|
|       |                  | PERSON           | LOCATION | ORGANIZATION | DATE  | PERCENT | TIME  | QUANTITY | MONEY | Precision | Recall | F-measure |
| M/F   | T270             | 0.97F            | 0.82F    | 0.87F        | 0.79F | 0.97F   | 0.74F | 0.71F    | 0.91F | 0.80      | 0.87   | 0.84      |
|       | N30              | 0.33F            | 0.47F    | 0.33F        | 0.63F | 0.71F   | 0.51F | 0.55F    | 0.57F | 0.40      | 0.65   | 0.49      |
|       | N10              | 0.56F            | 0.75F    | 0.50F        | 0.51F | 0.21F   | 0.0F  | 0.0F     | 0.0F  | 0.46      | 0.70   | 0.55      |
| M/B   | T270             | 0.93F            | 0.87F    | 0.85F        | 0.73F | 0.96F   | 0.68F | 0.71F    | 0.77F | 0.80      | 0.87   | 0.83      |
|       | N30              | 0.30F            | 0.47F    | 0.31F        | 0.55F | 0.71F   | 0.32F | 0.50F    | 0.42F | 0.35      | 0.63   | 0.45      |
|       | N10              | 0.46F            | 0.83F    | 0.47F        | 0.53F | 0.74F   | 0.0F  | 0.0F     | 0.33F | 0.51      | 0.75   | 0.61      |
| MF/F  | T270             | 0.78F            | 0.70F    | 0.68F        | 0.43F | 0.28F   | 0.54F | 0.33F    | 0.31F | 0.54      | 0.74   | 0.62      |
|       | N30              | 0.67F            | 0.62F    | 0.57F        | 0.33F | 0.0F    | 0.44F | 0.44F    | 0.28F | 0.47      | 0.64   | 0.54      |
|       | N10              | 0.80F            | 0.85F    | 0.64F        | 0.24F | 0.0F    | 0.0F  | 0.0F     | 0.0F  | 0.54      | 0.76   | 0.63      |
| MF/B  | T270             | 0.79F            | 0.68F    | 0.73F        | 0.61F | 0.94F   | 0.65F | 0.53F    | 0.68F | 0.67      | 0.70   | 0.69      |
|       | N30              | 0.71F            | 0.55F    | 0.54F        | 0.53F | 0.97F   | 0.39F | 0.45F    | 0.37F | 0.53      | 0.58   | 0.55      |
|       | N10              | 0.73F            | 0.71F    | 0.68F        | 0.50F | 1.0F    | 0.0F  | 0.0F     | 0.46F | 0.69      | 0.65   | 0.67      |
| MF/FB | N30              | 0.68F            | 0.55F    | 0.62F        | 0.39F | 0.0F    | 0.48F | 0.27F    | 0.06F | 0.45      | 0.58   | 0.51      |
|       | N10              | 0.81F            | 0.79F    | 0.73F        | 0.38F | 0.0F    | 0.0F  | 0.0F     | 0.0F  | 0.60      | 0.73   | 0.66      |

**Table 6. Named entity recognition evaluation**

performance, and 90 travel articles. Other 30 documents, which is not used in training, are used as test data. Test data has three types such as untrained 10 economical documents (N10), untrained 30 documents(N30), and trained 270 documents(T270).

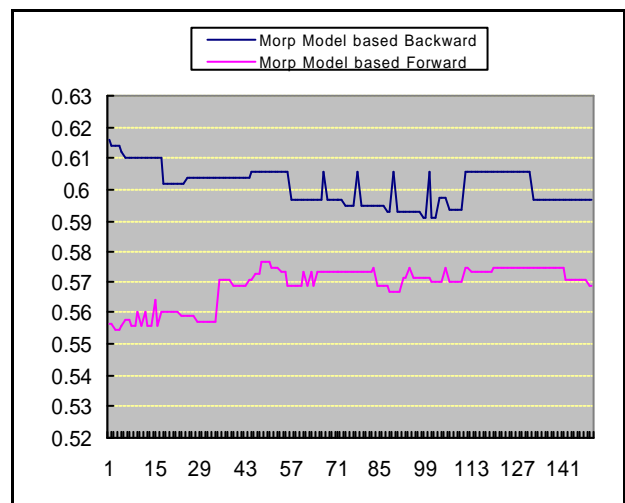
The result of the experiment is described in Table 6. From the result, we find that the best result of economy 10 test data is morph/feature based backward view(MF/B) type with F-measure 0.67 considering all NE types. If we considered only PLO(Person/Location/Organization) type, MF/FB type is the best with F-measure 0.77. The first reason that PLO recognition is better than other types is that the PLO trained data is more abundant. Figure 4 shows that the PLO type is a large number in 300 labeled documents. The second reason is that the statistical model is not appropriate to some kinds of NE types such as DATE, QUANTITY. These type is more appropriate when the pattern based approach is adopted. In the future, we test the NE recognizer with balanced trained data in each NE types.



**Figure 4. Distribution of NE types in 300 NE labeled documents**

## 5.2 CoTraining Test

We evaluate Co-Training for the NE recognition using an unlabeled economy domain newspaper articles(39,480 articles). For the training data, we train the NE recognizer with 270 labeled articles and use 10 evaluate articles as test data in each Co-Training iteration. In each training step, we increase the number of the NE labels from the high ranked NE tagging results in proportion to the training data. We test CoTraining with morph model which is divided into forward view and backward view. After 145 iteration, backward view F-measure decreases from 0.615 to 0.6, but forward view F-measure increases from 0.558 to 0.57.



**Figure 5. CoTraining result**

## 6. Conclusion

In this paper, we suggest HMM based NE recognition and

boosting technique. Through this research, we met some technical problems such as NE context model unification, unbalanced labeled corpus, and boosting degradation. (1) We consider HMM based NE recognition with four types NE context models which are derived from the analysis of NE labeled data. However, we cannot unify all of the models in unique way since the weighted integration of the models do not guarantee the good performance. (2) In using the labeled data, we meet the problem that the recognition of NE types depends highly on the size of learning data. (3) In HMM based CoTraining, the test result shows that high-performance model enhance low-performance model but high-performance model decrease step by step.

Finally, we conclude that (1) unification issue of various context models can be resolved with Maximum Entropy model which can combine diverse forms of contextual information in a principled manner, (2) unbalanced labeled corpus issue may be resolved by gathering contextual information independently from the labeled corpus. It makes it possible to balance learning data in each type, and (3) degradation of the boosting approach is not difficult problem since the boosting step in each round can be controlled with pre-test.

## References

- [1] A. Borthwick. A Japanese named entity recognizer constructed by a non-speaker of Japanese. In Proceedings of the IREX Workshop, pages 187-193, 1999.
- [2] A. Blum and T. Mitchell. Combining labeled and unlabeled data with cotraining. In Proceedings of the 11th Annual Conference on Computational Learning Theory, pages 92-100, 1998.
- [3] A. Ittycheriah, M. Franz, W. Zhu, and A. Ratnaparkhi, IBM's Statistical Question Answering System, In Proceedings of the Text Retrieval Conference TREC-9, 2000.
- [4] D. M. Bikel, S. Miller, R. Schwartz, R. Weischedel, Nymble : a high-performance learning named-finder, In Proceedings of the Fifth Conference on Applied Natural Language Processing, 1997
- [5] G. Zhou, J. Su, Named Entity Recognition using an HMM-based Chunk Tagger, In 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, 2002.
- [6] F. James, Modified Kneser-Ney Smoothing of n-gram Models. Technical Report TR00-07, RIACS, USRA, 2000.
- [7] K. Nigam and R. Ghani. Analyzing the effectiveness and applicability of co-training. In Proceedings of the Ninth International Conference on Information and Knowledge Management, 2000.
- [8] K. Nigam and R. Ghani. Understanding the Behavior of Co-training. In Proceedings of KDD-2000 Workshop on Text Mining, 2000.
- [9] M. Collins and Y. Singer. Unsupervised models for named entity classification. In Empirical Methods in Natural Language Processing and Very Large Corpora, 1999.
- [10] S. Harabagiu, D. Moldovan, M. Pasca, R. Mihalcea, M. Surdeanu, R. Bunescu, R. Girju, V. Rus and P. Morarescu, FALCON: Boosting Knowledge for Answer Engines, In Proceedings of the Text Retrieval Conference TREC-9, 2000.