

Article

Low-Cost Implementation of a Named Entity Recognition System for Voice-Activated Human-Appliance Interfaces in a Smart Home

Geonwoo Park and Harksoo Kim *

Program of Computer and Communications Engineering, Kangwon National University, Chuncheon-si 24341, Korea; parkku01@kangwon.ac.kr

* Correspondence: nlpdrkim@kangwon.ac.kr; Tel.: +82-33-250-6388; Fax: +82-33-259-5677

Received: 18 January 2018; Accepted: 11 February 2018; Published: 12 February 2018

Abstract: When we develop voice-activated human-appliance interface systems in smart homes, named entity recognition (NER) is an essential tool for extracting execution targets from natural language commands. Previous studies on NER systems generally include supervised machine-learning methods that require a substantial amount of human-annotated training corpus. In the smart home environment, categories of named entities should be defined according to voice-activated devices (e.g., food names for refrigerators and song titles for music players). The previous machine-learning methods make it difficult to change categories of named entities because a large amount of the training corpus should be newly constructed by hand. To address this problem, we present a semi-supervised NER system to minimize the time-consuming and labor-intensive task of constructing the training corpus. Our system uses distant supervision methods with two kinds of auto-labeling processes: auto-labeling based on heuristic rules for single-class named entity corpus generation and auto-labeling based on a pre-trained single-class NER model for multi-class named entity corpus generation. Then, our system improves NER accuracy by using a bagging-based active learning method. In our experiments that included a generic domain that featured 11 named entity classes and a context-specific domain about baseball that featured 21 named entity classes, our system demonstrated good performances in both domains, with F1-measures of 0.777 and 0.958, respectively. Since our system was built from a relatively small human-annotated training corpus, we believe it is a viable alternative to current NER systems in smart home environments.

Keywords: human-appliance interface system; named entity recognition; bagging-based active learning; distant supervision; low-cost implementation

1. Introduction

In the near future, smart homes will offer social networking to their residents or their appliances. Some information appliances will interact with their residents by using natural language commands. To correctly catch users' intentions, the information appliances should extract target objects from users' natural language commands which are in the form of short text messages. As shown in Figure 1, to perform the natural language command, "Play Yesterday and call Gildong Hong", information appliances should extract "Yesterday" and "Gildong Hong" from the command. Then, they should identify "Yesterday" and "Gildong Hong" with the semantic categories, "SONG" and "PERSON". In natural language processing, this task is called named entity recognition (NER).



Figure 1. Example of a natural language command.

Named entities (NEs) are informative elements that refer to proper names, such as the names of people, locations, and organizations. Named entity recognition (NER) is a subtask of information extraction that identifies NEs in texts and classifies them into predefined classes, such as PERSON, LOCATION, and ORGANIZATION. As shown in Figure 2, NE classes are contextually relevant.

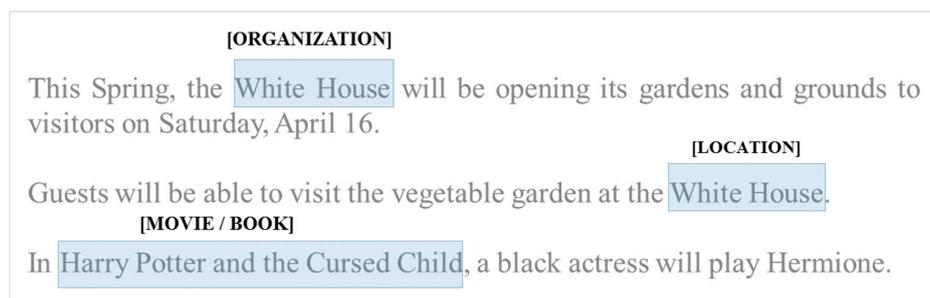


Figure 2. Example of named entity recognition.

In the first sentence, “White House” refers to the name of an organization and is used as the name of location in the second sentence. Additionally, NE classes are defined according to their domains. Referring back to Figure 2, if the third sentence belongs to a movie article, then “Harry Potter and the Cursed Child” may be classified as the title of a movie, and if the sentence is a part of book review, the term may be classified as the title of a book. Therefore, a substantial amount of domain-specific training corpus is needed to implement an NER that is based on machine learning; however, constructing such a corpus requires annotations and tags, which is a time-consuming and labor-intensive task that makes it difficult to promptly implement NER systems according to change of information appliances in smart home. To address this problem, we developed an NER system that utilizes an NE dictionary and a raw corpus (i.e., a set of sentences that are weakly labeled and are not annotated with any tags).

2. Previous Works

Previous NER systems are divided into two types: systems based on symbolic rules (rule-based systems) and systems based on machine learning (ML-based systems). Rule-based systems use regular-expression patterns and NE dictionaries [1,2]. If an NE dictionary is sufficiently large and patterns are generated by referring to a large corpus, the performances of rule-based systems may be satisfactory. However, the initial implementation for rule-based systems is high and rule-based systems are unfeasible when there are too many rules to manage. To address these limitations, ML-based systems have been implemented that primarily utilize supervised learning models to collect statistical

information from a large annotated corpus and determine NE classes based on this information [3–10]. Recently, ML-based systems that implement well-known supervised learning models have been developed to improve the accuracy of NER systems. These models include: Decision Trees (DT) [4], Maximum Entropy Models (MEM) [5], Conditional Random Fields (CRF) [6,7], structural Support Vector Machines (SVM) [8], and recent neural network models based on Long-Short Term Memory (LSTM) with a CRF layer [11–13].

The ML-based systems are a more feasible alternative to rule-based systems, but their performances depend on the size of NE tagged training corpus. To address this problem, some active-learning models are proposed [14,15]. These models showed that manual labeling cost can be reduced without or with only a little degrading of the performances. However, they still need human annotation efforts for constructing the initial training corpus. To resolve this problem, we propose a semi-supervised NER system using active learning [16] based on bagging (bootstrap aggregating) [17] with distant supervision [18]. Unlike existing ML-systems, our system does not require a substantial amount of NE tagged training corpus and instead, only requires a NE dictionary that contains NEs and their classes. By using a distant supervision learning process that is based on the NE dictionary, our system is capable of automatically annotating a raw corpus with NE classes. Furthermore, we use a bagging-based active learning process to refine noises in the corpus (i.e., words annotated with incorrect NE classes), which refines the NER accuracy of our system. A preliminary discussion of our model was presented in [19] as a short paper. The preliminary model did not consider that most of entry words in a NE dictionary can have multiple NE classes during the distant supervision phase. In other words, most of the location names can be used as organization names. In the previous work, we compulsory assigned a NE class to each entry word in a NE dictionary. This increased the noise in an initial training corpus that is automatically constructed by distant supervision. To resolve this problem, a new proposed model performs a learning phase for single-class NEs and a separate learning phase for multi-class NEs. In addition, the preliminary model was evaluated by using a generic data set, but the new model is evaluated by using two different data sets (a generic domain and a context specific domain) in order to prove domain portability. A brief description of these processes is explained below.

3. Named Entity Recognition Using Two-Phase Bagging-Based Active Learning

This section is divided into subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

3.1. System Architecture

As shown in Figure 3, our system uses distant supervision to automatically annotate a large amount of raw corpus with NE classes by matching word sequences against single-class NEs from the NE dictionary.

The bagging-based active learning component includes a learning phase for single-class NE learning and a separate learning phase for multi-class NE learning. During the single-class NE learning phase, our system selects “noise sentences”, which come from the weakly labeled training corpus based on disagreement scores between bagging models trained by the weakly labeled training corpus (i.e., the training corpus annotated with single-class NEs). Then, the noise sentences are manually revised according to an active learning method. This refinement process is repeated until some terminal conditions are satisfied and the result is a single class NER model. In the multi-class NE learning phase, our system first extracts sentences including multi-class NEs from the raw training corpus by matching word sequences against multi-class NEs in a NE dictionary. Second, our system automatically annotates the extracted sentences with NE classes by using the single-class NER model corpus. Then, our system performs the same bagging-based active learning with the single-class NE learning phase, except that the sentences annotated by the single-class NER model are used as the weakly labeled training corpus.

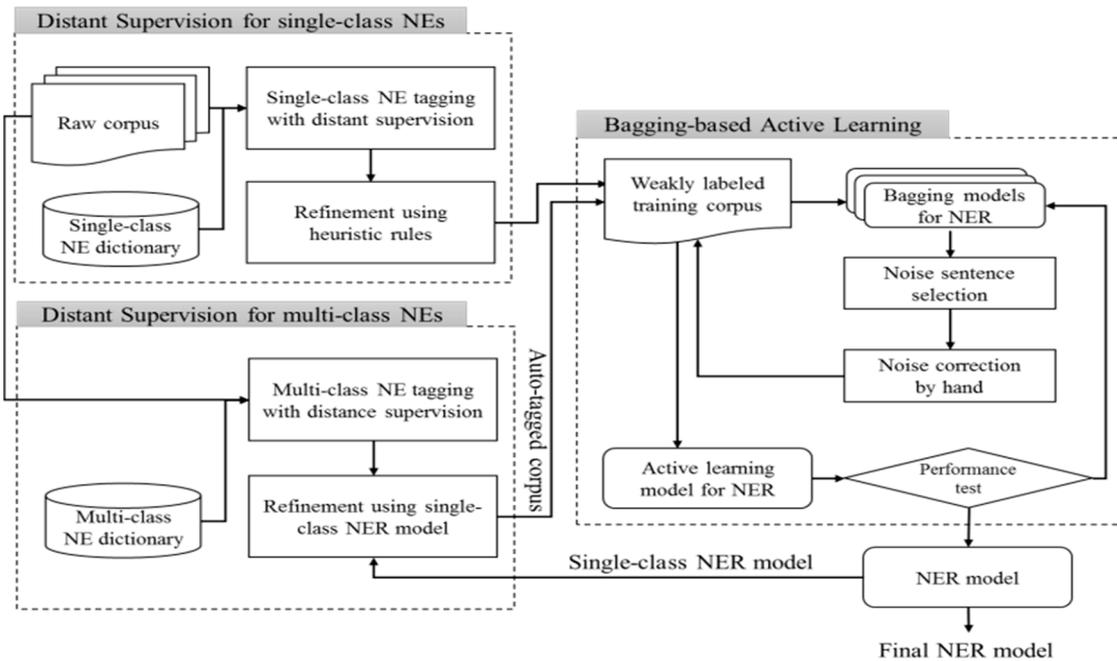


Figure 3. Overall architecture of the proposed model.

3.2. Constructing Weakly Labeled Corpus Using Distant Supervision

The first step to construct a weakly labeled corpus using distant supervision is to generate a NE dictionary. Next, a raw corpus is gathered from any collection of chosen documents. A weakly labeled training corpus is then constructed by matching sentences from the raw corpus against single class NEs in the NE dictionary. Using heuristics, incorrect labels in the training corpus are then removed in the following manner:

- Remove labels of words with declined or conjugated endings because endings of NEs are generally nouns.
- Remove labels of high-frequency words in the weakly labeled training corpus because NEs are not common words (Zipf’s law) [20].

Heuristics are language-specific and can be modified accordingly. Figure 4 illustrates snippets of Korean sentences that are weakly labeled by distant supervision using heuristic rules.



Figure 4. Example of weakly labeled sentences.

After the weakly labeled training corpus is constructed, we utilize a bagging-based active learning algorithm to improve the accuracy of our system (see Algorithm 1).

Algorithm 1. Bagging-Based Active Learning

1. Generate n bagging corpus from training corpus by sampling with replacement.
2. Train n NER models using n bagging corpus, respectively.
3. Check disagreement scores between outputs of n NER models by using the whole training corpus as test data.
4. Select m sentences with high disagreement scores.
5. Revise incorrect labels in m sentences by hand.
6. Update the training corpus with the revised sentences. Train an NER model using the updated training corpus.

Check accuracy of the NER model by using gold-labeled validation corpus. If accuracy improvement is converged, terminate the learning process. Otherwise, go to step (1).

The size of the bagging corpus is experimentally set to 10% to 20% of the training corpus. Sentences with high disagreement scores are corrected manually. Disagreement scores are the number of NER models that return different NER results and are calculated in the following manner: if a sentence has two NEs and the first NE is annotated with unique labels by three NER models and the second NE is annotated with unique labels by five NER models, a disagreement score of this sentence is five.

A variety of machine learning models can be used to execute the bagging-based active learning algorithm. For our system, we chose to implement CRFs, introduced by Lafferty [21], because they typically have high performances in the sequence labeling process that assigns categorical labels to each member of a sequence of observed values. The NER models annotate input sentences according to a Begin-Inner-Outer (BIO) tagging scheme. For example, “Obama lives in White House” is labeled as “Obama/B_PER lives in White/B_LOC House/I”, where “B_PER” means the beginning of person’s name, and “B_LOC” means the beginning of location’s name, and “I” means the inner of an NE. Table 1 shows input features of the NER models. As shown in Table 1, the input features are designed for Korean sentences, but we believe that language change will not be a difficult task because the features are based on shallow NLP (natural language processing) knowledge.

Table 1. List of input features for NER.

Feature Name	Explanation
LEX	The current eojeol (Korean spacing unit)
FW_2_Lex, BW_2_Lex	First two eomjeols (Korean syllable) and last two eomjeols in the preceding, current, and next eojeols
FW_2_Tags, BW_2_Tags	NE categories matching against FW_2_Lex and BW_2_Lex in the preceding, current, and next eojeols
FW_3_Lex, BW_3_Lex	First three eomjeols and last three eomjeols in the preceding, current, and next eojeols
FW_3_Tags, BW_3_Tags	NE categories matching against FW_3_Lex and BW_3_Lex in the preceding, current, and next eojeols
BIEF (BE, BF, IE, IF)	BE: a tag meaning “the current eojeol is exactly matched against an entry in an NE dictionary” BF: a tag meaning “the current eojeol is partially matched against first few eomjeols in an entry in an NE dictionary” IE: a tag meaning “the current eojeol is included in an entry in an NE dictionary” IF: a tag meaning “the current eojeol is partially matched against last few eomjeols in an entry in an NE dictionary”
POS_Bigram	POS (part-of-speech) bi-grams of the preceding, current, and next eojeols
LEX-POS_Unigram	“Morpheme/POS” uni-grams of the preceding, current, and next eojeols

3.3. Multi-Class NE Learning Phase: Constructing a Final NER System from Single-Class NE Tagged Corpus

After constructing a single-class NE tagged corpus (single-class NER model), our system constructs a multi-class NE tagged corpus by first extracting sentences that include multi-class NEs from the raw training corpus by using the distant supervision method detailed in Section 3.1, but the refinement process based on heuristic rules is excluded. Then, our system automatically annotates the extracted sentences by using the single-class NE tagged corpus. Finally, our system then merges the single-class NE tagged corpus and the multi-class NE tagged corpus and performs the bagging-based active learning with the single-class NE learning phase that is detailed in Section 3.3 using the merged corpus as the full training corpus. The final output is an NER system that is capable of annotating both single-class NEs and multi-class NEs. Figure 5 illustrates the multi-class NE learning phase.

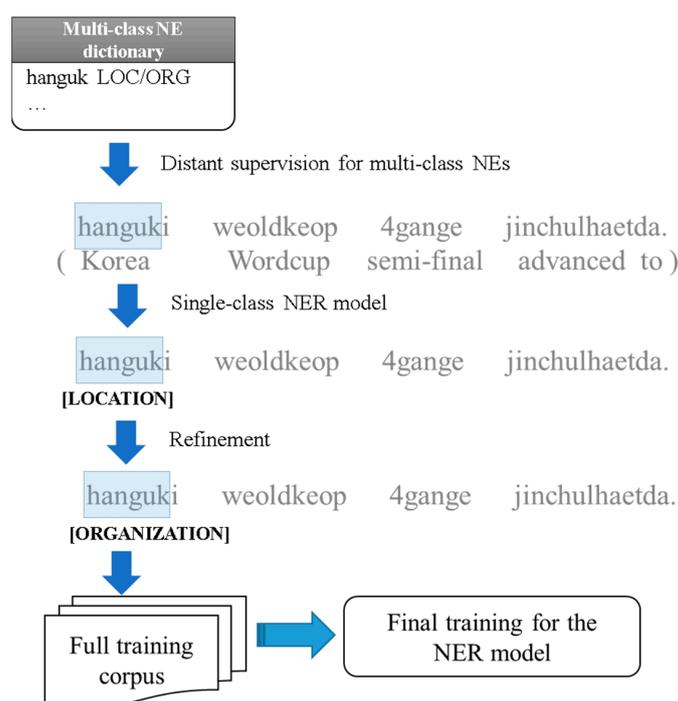


Figure 5. Example of multi-class NE learning.

4. Evaluation

4.1. Data Sets and Experimental Settings

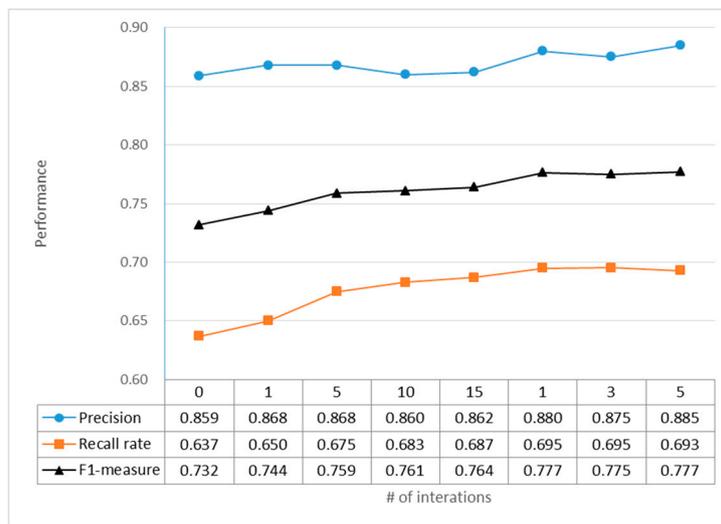
For our experiments, we constructed two different domains to evaluate our system. The first is a generic domain that included the following 11 NE classes that was constructed from the Korean version of Wikipedia [22,23]: PERSON, LOCATION, ORGANIZATION, CELESTIAL_BODY, EVENT, FACILITY, GAME, LANGUAGE, LAW, PERSON_FICTION, and STUDY_FIELD. Next, we collected a raw corpus from random Wikipedia abstracts. This generic corpus consisted of 55,000 sentences, with 54,000 sentences being used to train and the remaining 1000 sentences being used for evaluation. The second domain was a context-specific domain that involved the sport of baseball. We found that people frequently seek information from news articles through smart speakers like Amazon (<http://www.amazon.com>) Echo and Naver (<http://www.naver.com>) Clova. In particular, they often asked the smart speakers about sports records. Based on this observation, we assumed that smart home residents will like to seek information from news articles. We chose the baseball domain in order to evaluate efficiency and usefulness of our system according to domain change. For this domain, we used the following 21 NE classes: BATTER, PITCHER, MANAGER, ANNOUNCER, COMMENTATOR, CHEERLEADER, OPENING_DAY_PITCHER, OWNER, PRESIDENT, UMPIRE, ETC_PERSON, TEAM,

ASSOCIATION, BROADCASTING, COMPANY, SCHOOL, LEAGUE, STADIUM, NATION, CITY, and STATE. The baseball corpus was collected from online articles that pertained to baseball sporting news. This corpus consisted of 162,000 sentences with 161,500 sentences being used to train and the remaining 500 being used for evaluation.

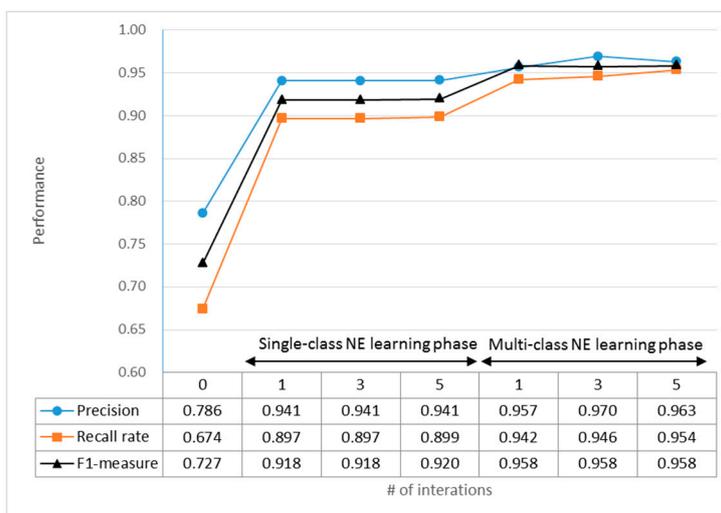
Both corpora were trained using our distant supervision and bagging-based active learning phases described in Section 3. Both testing corpora were randomly selected from the collected corpora and were manually annotated with gold labels to indicate the correct NE classes. The manual annotation was done by five graduate students who had knowledge of natural language processing, and, for consistency, was post-processed by a student doing a doctoral course. The number of bagging models, n , was set to 10 and the threshold values of disagreement scores ranged from 8 to 10.

4.2. Experimental Results

Our first experiment evaluated the performance of our system for each domain. As indicated in Figure 6a,b, our system gradually increased the F1-measure for each domain.



(a) Performance change according to iterations in the general domain.

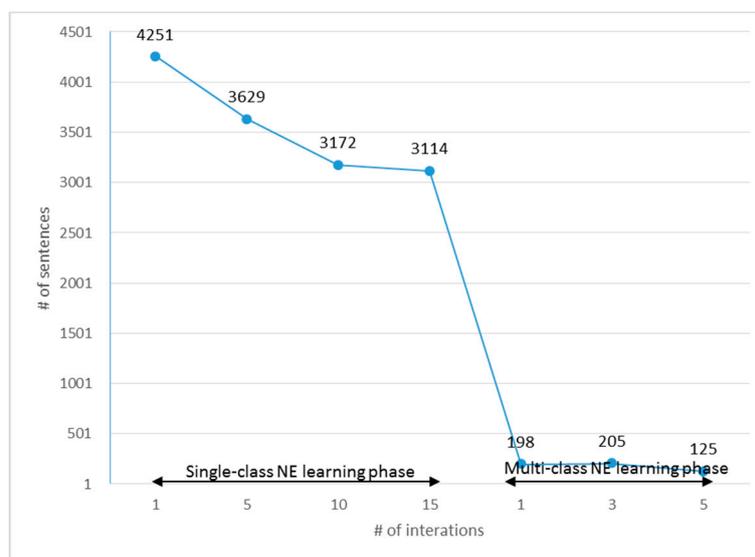


(b) Performance change according to iterations in the baseball domain.

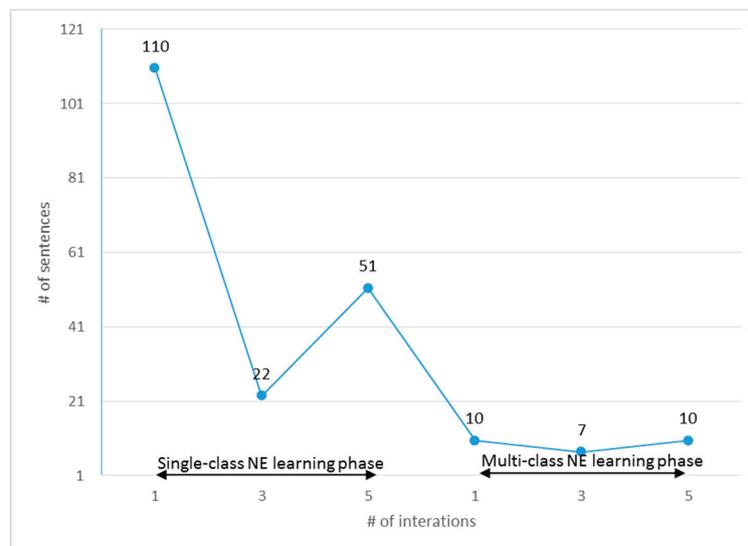
Figure 6. Performance change according to iterations.

Of particular interest is the performance increase per iteration in the baseball domain being greater than that in the generic domain, which can be explained in the following manner. First, NEs in the baseball domain have less ambiguity. For example, the name of an organization found in the generic domain can also be names of location, such as the White House; conversely, in the baseball domain, the team name (name of an organization) is different from the name of the stadium (location), such as the Atlanta Braves team that plays at Turner Field. Second, the baseball dictionary covers most NEs in the baseball corpus, so there are fewer NEs in the baseball corpus that are not in the baseball dictionary than the number of NEs in the generic corpus that are not in the generic dictionary. Consequently, NEs for the baseball domain have more accurate boundaries compared to NEs in the generic domain.

The second experiment was to evaluate the efficiency of our system according to the number of manual tagged sentences during bagging-based active learning. As shown in Figure 7a,b, the number of manually tagged training corpus needed was reduced for almost every subsequent iteration.



(a) The number of manual tagged sentences according to iterations in the general domain.



(b) The number of manual tagged sentences according to iterations in the baseball domain.

Figure 7. The number of manual tagged sentences according to iterations.

The last experiment we conducted was to compare our system with the previous systems. Figure 8 shows the performance differences between our system and the previous systems.

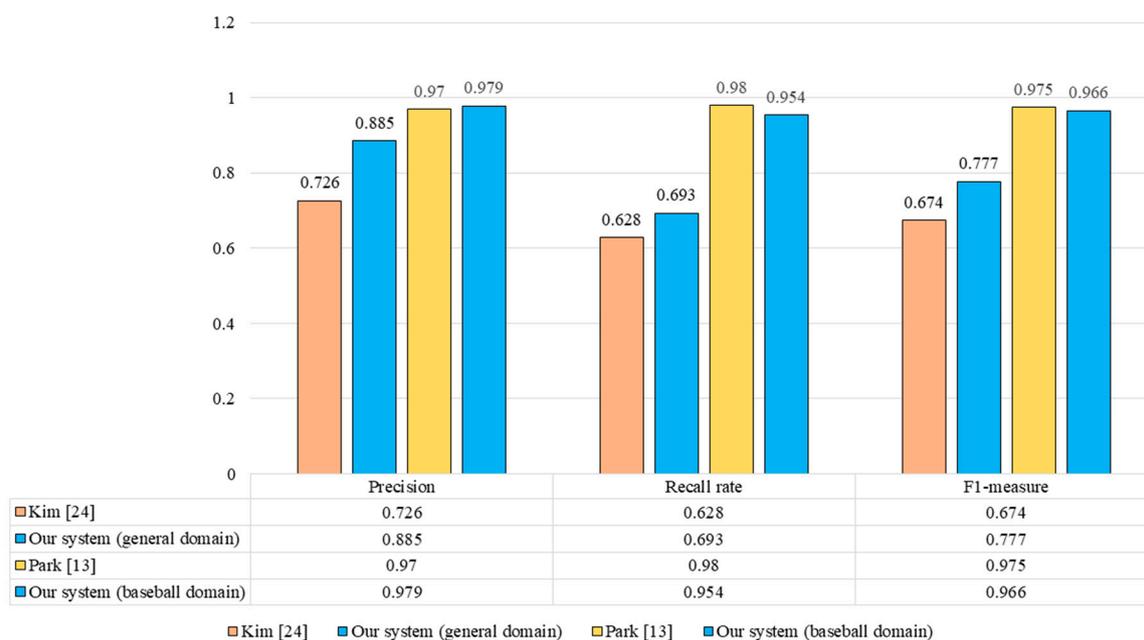


Figure 8. Performance comparison of NER models.

In Figure 8, Kim's system [24] is an NER model based on CRFs, and Park's system [13] is an NER model based on a BiLSTM-CRF (Bidirectional LSTM with a CRF layer). Kim's system was trained by large amounts of NE tagged corpus that are automatically constructed from the Korean version of Wikipedia according to his/her method in [24]. Park's system was trained by the baseball corpus in which we grouped 21 NE classes to three NE classes named PERSON, LOCATION, and ORGANIZATION. As shown in Figure 8, our system outperformed Kim's system in the same test environments (i.e., the same training and test data). This fact shows that the proposed data construction method is more effective than Kim's method [24]. In addition, our system achieved competitive performances with Park's system in the same test environments. The performance differences were caused by the underlying machine-learning models, CRFs and BiLSTM-CRF. This fact shows that our system can achieve higher performances if we adopt BiLSTM-CRF as an underlying machine learning model.

4.3. Limitations

There were instances where our system failed to return the correct NEs. This occurred when incorrect entries in the NE dictionary caused wrong annotations in the weakly labeling training corpus. For example, the generic NE dictionary included some noise entries (i.e., the construction accuracy of the NE dictionary is the average micro F1-measure of 0.955). Additionally, some NEs that were not in the NE dictionary did not participate in the training process. These issues can be addressed by refining the NE dictionary more thoroughly.

5. Conclusions

Our semi-supervised NER model was developed using distant supervision and bagging-based active learning. Our system effectively generates a weakly labeled training corpus to create single-class and multi-class NER models and refines these models to improve NER accuracy. Based on our experimental results, our system performed generally well, especially for the baseball domain

(context-specific domain). Additionally, our system did not require a substantial amount of manually entered training corpus. The value added by our system is a reduced effort to manually constructing training data and thus, our system may be considered a viable and feasible alternative to ML-based NER systems for information appliances in smart home.

In the future, we will concentrate on reducing the ambiguities of weak labeling that is caused by the distance supervision method. Also, we will change the underlying machine-learning model CRFs into the latest deep-learning model in order to increase overall performances. In addition, finally, we are working on efficient ways of refining NE dictionaries.

Acknowledgments: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2016R1A2B4007732). It was also supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2017M3C4A7068188).

Author Contributions: Harksoo Kim conceived and designed the experiments; Geonwoo Park performed the experiments; Geonwoo Park analyzed the data; Harksoo Kim wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Noh, T.; Lee, S. Extraction and Classification of Proper Nouns by Rule-based Machine Learning. In Proceedings of the KIISE Korea Computer Congress, Gyeongju, Korea, 29 June–1 July 2000. (In Korean)
2. Seon, C.; Kim, H.; Seo, J. Translation Assistance System Based on Selective Weighting and Cluster-Based Searching Methods. *Int. J. Artif. Intell. Tools* **2012**, *21*. [[CrossRef](#)]
3. Hwang, Y.; Lee, H.; Chung, E.; Yun, B.; Park, S. Korean Named Entity Recognition Based on Supervised Learning Using Named Entity Construction Principle. In Proceedings of the HCLT, Choengju, Korea, 11–12 October 2002. (In Korean)
4. Sekine, S.; Grishman, R.; Shinnou, H. A Decision Tree Method for Finding and Classifying Names in Japanese Texts. In Proceedings of the 6th Workshop on Very Large Corpora, Montreal, QC, Canada, 15–16 August 1998.
5. Brothwick, A.; Sterling, J.; Agichtein, E.; Grishman, R. NYU: Description of the MENE Named Entity System as Used in MUC-7. In Proceedings of the Seventh Message Understanding Conference, Fairfax, VA, USA, 29 April–1 May 1997.
6. Cohen, W.W.; Sarawagi, S. Exploiting Dictionaries in Named Entity Extraction: Combining Semi-Markov Extraction Processes and Data Integration Methods. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004.
7. Lee, C.; Hwang, Y.G.; Oh, H.J.; Lim, S.; Heo, J.; Lee, C.; Kim, H.; Wang, J.; Jang, M. Fine-Grained Named Entity Recognition Using Conditional Random Fields for Question Answering. In Proceedings of the HCLT, Pohang, Korea, 13–14 October 2006. (In Korean)
8. Lee, C.; Jang, M. Named Entity Recognition with Structural SVMs and Pegasos Algorithm. *Korean J. Cogn. Sci.* **2010**, *21*, 655–667. (In Korean)
9. Seon, C.; Kim, H.; Seo, J. Efficient Appointment Information Extraction from Short Messages in Mobile Devices with Limited Hardware Resources. *Pattern Recognit. Lett.* **2011**, *32*, 127–133. [[CrossRef](#)]
10. Seon, C.; Yoo, J.; Kim, H.; Kim, J.; Seo, J. Lightweight Named Entity Extraction for Korean Short Message Service Text. *KSII Trans. Internet Inf. Syst.* **2011**, *5*, 560–574. [[CrossRef](#)]
11. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural Architectures for Named Entity Recognition. In Proceedings of the NAACL-HLT, San Diego, CA, USA, 12–17 June 2016.
12. Kwon, S.; Ko, Y.; Seo, J. A Robust Named-Entity Recognition System Using Syllable Bigram Embedding with Eojeol Prefix Information. In Proceedings of the CIKM, Singapore, 6–10 November 2017.
13. Park, G.; Lee, H.; Kim, H. Named Entity Recognition Model Based on Neural Networks Using Parts of Speech Probability and Gazetteer Features. *Adv. Sci. Lett.* **2017**, *23*, 9530–9533. [[CrossRef](#)]
14. Shen, D.; Zhang, J.; Su, J.; Zhou, G.; Tan, C.L. Multi-Criteria-based Active Learning for Named Entity Recognition. In Proceedings of the ACL, Barcelona, Spain, 21–26 July 2004.
15. Law, F.; Schütze, H. Stopping Criteria for Active Learning of Named Entity Recognition. In Proceedings of the COLING, Manchester, UK, 18–22 August 2008.

16. Cohn, D.A.; Ghahramani, Z.; Jordan, M.I. Active Learning with Statistical Models. *J. Artif. Intell. Res.* **1996**, *4*, 705–712.
17. Ha, K.; Cho, S.; MacLachlan, D. Response Models based on Bagging Neural Networks. *J. Interact. Mark.* **2005**, *19*, 17–30. [[CrossRef](#)]
18. Mintz, M.; Bills, S.; Snow, R.; Jurafsky, D. Distant Supervision for Relation Extraction without Labeled Data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Suntec, Singapore, 2–7 August 2009.
19. Lee, S.; Song, Y.; Choi, M.; Kim, H. Bagging-Based Active Learning Model for Named Entity Recognition with Distant Supervision. In Proceedings of the BigComp, HongKong, China, 18–20 January 2016.
20. Zipf, G.K. *The Psychobiology of Language*; The MIT Press: Cambridge, MA, USA, 1935.
21. Lafferty, J.; McCallum, A.; Pereira, F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In Proceedings of the ICML, Williamstown, MA, USA, 28 June–1 July 2001.
22. Song, Y.; Kim, H. Semi-automatic Construction of a Named Entity Dictionary Based on Active Learning. *Comput. Sci. Appl.* **2015**, *330*, 65–70.
23. Song, Y.; Jeong, S.; Kim, H. Semi-automatic Construction of a Named Entity Dictionary for Entity-Based Sentiment Analysis in Social Media. *Multimed. Tools Appl.* **2017**, *76*, 11319–11329. [[CrossRef](#)]
24. Kim, Y. Automatic Training Corpus Generation Method of Named Entity Recognition Using Big Data. Master's Thesis, Sogang University, Seoul, Korea, 2015.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).