Korean Coreference Resolution with Guided Mention Pair Model Using Deep Learning

Cheoneum Park, Kyoung-Ho Choi, Changki Lee, and Soojong Lim

The general method of machine learning has encountered disadvantages in terms of the significant amount of time and effort required for feature extraction and engineering in natural language processing. However, in recent years, these disadvantages have been solved using deep learning. In this paper, we propose a mention pair (MP) model using deep learning, and a system that combines both rule-based and deep learning-based systems using a guided MP as a coreference resolution, which is an information extraction technique. Our experiment results confirm that the proposed deep-learning based coreference resolution system achieves a better level of performance than rule- and statistics-based systems applied separately.

Keywords: Coreference resolution, Guide mention pair, Multi-pass sieve, Mention pair, Deep learning.

Manuscript received Oct. 8, 2015; revised Mar. 27, 2016; accepted June 21, 2016.

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (no. R0101-16-0062, Development of Knowledge Evolutionary WiseQA Platform Technology for Human Knowledge Augmented Services).

Cheoneum Park (parkce3@gmail.com), Kyoung-Ho Choi (gtraccoon@gmail.com), and Changki Lee (corresponding author, leeck@kangwon.ac.kr) are with the Department of Computer Science, Kangwon National University, Chuncheon, Rep. of Korea.

Soojong Lim (isj@etri.re.kr) is with the SW & Content Research Laboratory, ETRI, Daejeon, Rep. of Korea.

I. Introduction

A coreference resolution refers to various words expressed differently with regard to a single entity. In other words, a coreference resolution is used to indicate an entity with links of various words that have the same meaning. For example, in the case of the following sentences, a coreference resolution defines words of the same semantic nature as one entity in a document.

"***Barack Obama*** *was born in Honolulu.*"

"***He*** *lives in the White House.*"

"***The American president*** *will visit Korea next month.*"

In the first sentence, [*Barack Obama*] indicates the same entity as [*He*] in the second sentence and [*The American president*] in the third sentence. Therefore, these words may be referring to each other, and thus we define this as a coreference; a coreference resolution is therefore a method for automatically finding such words, which is possible when the words are coreferent.

A coreference resolution is one of the important tasks that can be adapted for question answering, information extraction, and document abstraction, and has been regularly researched as a method for rule-based [1]–[3] and statistics-based [4], [5] systems through the DARPA Message Understanding Conferences (MUC), the SIGNLL Conference on Computational Natural Language Learning (CoNLL) [6], and other conferences. However, a coreference resolution is managed as a difficult problem because it has the tendency to depend on more semantic problems (that is, in the case of "Korea" and "our country") instead of problems such as the form or grammar of the natural language.

A coreference resolution is classified into two types: rule-based and statistics-based systems. A rule-based system is a method for conducting natural language processing (NLP) using certain defined rules. In this paper, our rule-based system suitably applies the Multi-pass Sieve of Stanford to the Korean language. A statistics-based system is a method for conducting an NLP using statistics or machine learning (ML) methods. In this paper, we apply a coreference resolution based on a mention pair (MP), and for the ML method, we utilize deep learning.

In this way, they both have their own following advantages and disadvantages. The disadvantages of rule-based systems include their dependence on the people who define the rules, a large consumption time for defining complete rules, and maintenance difficulties. Statistics-based systems commonly demonstrate a better level of performance than rule-based systems using ML and other methods, but they require well-designed features. General ML requires human beings to extract features for an optimal combination of features, and a significant amount of time and effort are needed. However, such problems from the use of ML can be overcome thanks to the recent development of deep learning [7], [8].

Deep learning is a model extended from artificial neural networks that was developed by emulating the human brain. Deep learning is constructed with the foundation of multiple levels of hidden layers, where each layer uses a non-linear activation function, which transforms each representation at one level into a representation at a higher and slightly more abstract level. Deep learning makes feature design automation possible, and much higher abstract levels can be continued as the layers become deeper. However, if the neural network layers become deeper, learning requires a longer amount of time because more computational complexity is expected; training also becomes difficult. Recently, the training speed has been able to be reduced through an advanced GPU, and other problems can be solved using pre-training, a dropout, a rectifier linear hidden unit (ReLU), and other methods that have been implemented in deep learning [9]–[11].

In this paper, a coreference resolution is conducted first in a rule-based manner. The combined coreference resolution system of rule-based and statistics-based systems used to conduct a deep-learning based coreference resolution is as follows.

II. Related Works

Traditional research on a coreference resolution can be classified into two types: rule-based [1]–[3] and statistics-based [4], [5] systems. First, Stanford [1] conducted a given task using a multi-pass sieve based on an entity-centric coreference. Such coreference resolution system manages mention-resolved coreferents for any entity, such as entity clusters (that is, an entity). Moreover, among the mentions included in each entity, there is strength in having characters that share each other's attributes. Hence, in [2] it was shown that a coreference resolution is suitable for the Korean language by applying the method of [1], and the rule-based system of this paper follows that of [2].

The coreference resolution system of [4] applied ML to models such as MP and mention ranking. Here, MP is a method that defines the reference relationship of two mentions, and MP is applied to the statistics-based methods of the present paper. Deep learning is used for ML in statistics-based methods, and comparative experiments with SVM were conducted to show that deep learning is more suitable for MPs.

In the case of rule-based systems, they can maintain various mention relationships by sharing the attributes among mentions in an entity because a rule-based system is based on an entity-centric coreference. In contrast, in the case of statistics-based systems, most references comply with the head because the references depend on the feature expression. For this reason, the recall is relatively higher in rule-based systems, but statistics-based systems show a distinctly high rate of precision, as shown in earlier experiments. Therefore, in this paper, we propose a method that uses MP with one sieve and applies a multi-pass sieve based on the entity-centric coreference of rule-based systems. Furthermore, we propose guided learning that uses the results of rule-based systems, and combines rule-based and statistics-based systems with their respective high rates of recall and precision.

III. Coreference Resolution

A coreference resolution is used to solve the relationship of various words expressed for any entity. A coreference resolution must be transitive, recursive, and symmetric. In the case of a document that explains a particular entity, the first proper noun of such entity appears, and the entity is expressed based on the nickname, acronym, and pronoun; the words are included in a coreferent relationship to each other. Accordingly, a coreference resolution is used for clustering by recovering the relationship of these words within a document.

The candidate words to be resolved in a coreference resolution are defined as "mentions." The coreference resolution system proposed in this paper first applies the rule-based system, and subsequently the ML-based MP model. The ML-based MP model uses deep learning, the results of rule-based systems for the ML feature, and entities of the rule-based systems like a continuous pipeline.

Figure 1 shows the coreference resolution system proposed in this paper. This system works by inputting the extracted information from a language analyzer, which means that it

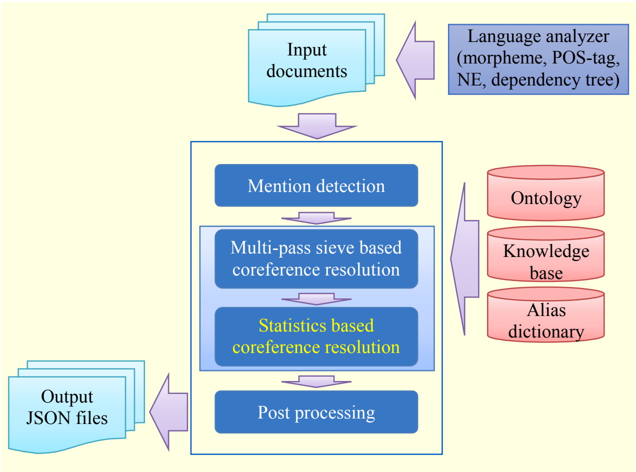

Fig. 1. Coreference resolution system.

Table 1. Multi-pass sieve of coreference resolution.

Sequence		Model Name
Step 1		Mention detection
Step 2	Sieve 0	Impossible coreference
	Sieve 1	Exact string match
	Sieve 2	Precise constructs
	Sieve 3–5	Strict head match A-C
	Sieve 6	Proper head noun match
	Sieve 7	Relaxed head match
	Sieve 8	Pronominal coreference resolution
Step 3		Post processing

extracts the morpheme, named entity (NE), part-of-speech tag (POS-tag), and dependency information from input documents using this analyzer. First, this system extracts all usable mentions that depend on a dependency tree. In addition, the system resolves a coreference resolution using a rule-based, entity-centric, multi-pass sieve. Subsequently, a deep learning-based guided MP model, which uses the results of the rule-based system and features, is applied to the coreference link information based on the results. The deep learning in MP classifies a coreferent or not for two mentions (that is, a pair of mentions). Finally, the end result of the coreference resolution is extracted through post-processing.

IV. Rule-Based Coreference Resolution: Multi-pass Sieves

In this paper, we suitably apply the multi-pass sieve from Stanford, similar to that in [2], to the Korean language; the multi-pass sieve applied in this paper was applied based on mentions extracted through mention detection. The mentions were extracted by targeting every possible noun or noun phrase (NP) based on a dependency tree for a high recall. Subsequently, a coreference resolution was applied to all mentions through the sieves. The sieves are classified into four types: an impossible link rule (sieve 0), morphometric expression (sieve 1) for high precision, semantic expressions (sieves 2 through 7) to increase the recall, and pronoun resolution (sieve 8). Table 1 lists the rule-based multi-pass sieve used in this paper.

1. Mention Detection

Mentions include the NE information, modifiers, and head words present in the mention; we then refer to the information in the running phase of the multi-pass sieve. Accordingly, the performance of the coreference resolution is dependent upon the mention detection. If there are any missed mentions, errors are accumulated in the multi-pass sieve, which causes the overall score to decrease. Therefore, the mention detection must derive the greatest recall. The mention detection used in this paper has the following conditions:

- ***Overall mention based on NPs*** – The mention detection in this paper defines all NPs that appear in the dependency tree as a mention.
- ***Processing mention by word unit*** – We process the mentions based on the word units because the syntax information is based on a dependency tree. We define the last word in the NPs as the head word.
- ***Containing modifier in mention*** – The dependency tree usually contains modifier information for each head word. We can turn modified NPs into mentions using this information.
- ***Atomic of Name Entity*** – In this paper, we define an NE as the least unit of the semantic word in the mentions. Accordingly, for any mention included in a mention whose head word is an NE, our system removes any mentions according to the atomic NE (for example, in the case of [[Renault] Motors: ORGANIZATION], we remove [Renault] because it is a non-NE).
- ***Processing head duplication*** – When our system extracts mentions, if two or more mentions with the same head word are extracted, we only save the longest boundary of a mention.
- ***Stop words*** – We do not extract mentions if the head word of any mention is a stop word (for example, ~걸(*~geol*), ~수 (*~su*), ~중 (*~jung*), ~간 (*~gan*), ~인 채(*~in chea*), and 중간에 (*jung-gan-e*)).
- ***Pronominal classification*** – In this paper, pronouns and

Table 2. Example of mention detection.

- **Sentence of input document**
  - 프랑스의 르노 자동차 그룹은 한국 삼성자동차 인수를 공식 제의할 것이다.
  - "France-ui reuno jadongcha geurup-eun hanguk samsungjadongcha insu-lul gongsik jeuihal kes-i-tta."
  - *Renault Samsung Motors group of France will suggest takeover of Samsung Motors of Korea.*
- **Step 1: Processing mention by word unit -** extracting word of sentence
  - [프랑스], [르노], [자동차], [그룹], [한국], [삼성자동차], [인수]
  - [France], [reuno], [jadongcha], [geurup], [hanguk], [samsungjadongcha], [insu]
  - [*France*], [*Renault*], [*Motors*], [*group*], [*Korean*], [*Samsung Motors*], [*takeover*]
- **Step 2: Containing modifier in mention -** extending modifier
  - [프랑스], [르노], [자동차], [르노 자동차], [그룹], [르노 자동차 그룹], [프랑스의 르노 자동차 그룹], [한국], [삼성자동차], [한국 삼성자동차], [인수], [한국 삼성자동차 인수]
  - [France], [reuno], [jadongcha], [reuno jadongcha], [geurup], [reuno jadongcha geurup], [France-ui reuno jadongcha geurup], [hanguk], [samsungjadongcha], [hanguk samsungjadongcha], [insu], [hanguk samsungjadongcha insu]
  - [*France*], [*Renault*], [*Motors*], [*Renault Motors*], [*group*], [*Renault Samsung Motors group*], [*Renault Samsung Motors group of France*], [*Korean*], [*Samsung Motors*], [*Samsung Motors of Korea*], [*takeover*], [*takeover of Samsung Motors of Korea*]
- **Step 3: Atomic of NE**
  1. Applying NE
  - [프랑스]: LOCATION, [르노], [르노 자동차]: ORGANIZATION, [그룹], [르노 자동차 그룹], [프랑스의 르노 자동차 그룹], [한국]: LOCATION, [삼성자동차]: ORGANIZATION, [한국 삼성자동차], [인수], [한국 삼성자동차 인수]
  - [France]: LOCATION, [reuno], [jadongcha], [reuno jadongcha]: ORGANIZATION, [geurup], [reuno jadongcha geurup], [France-ui reuno jadongcha geurup], [hanguk]: LOCATION, [samsungjadongcha], [hanguk samsungjadongcha]: ORGANIZATION, [insu], [hanguk samsungjadongcha insu]
  - [*France*]: LOCATION, [*Renault*], [*Motors*], [*Renault Motors*]: ORGANIZATION, [*group*], [*Renault Samsung Motors group*], [*Renault Samsung Motors group of France*], [*Korean*]: LOCATION, [*Samsung Motors*]: ORGANIZATION, [*Samsung Motors of Korea*], [*takeover*], [*takeover of Samsung Motors of Korea*]
  2. Result of this step
  - [프랑스], [르노 자동차], [그룹], [르노 자동차 그룹], [프랑스의 르노 자동차 그룹], [한국], [삼성자동차], [인수], [한국 삼성자동차 인수]
  - [France], [reuno jadongcha], [geurup], [reuno jadongcha geurup], [France-ui reuno jadongcha geurup], [hanguk], [samsungjadongcha], [insu], [hanguk samsungjadongcha insu]
  - [*France*], [*Renault Motors*], [*group*], [*Renault Samsung Motors group*], [*Renault Samsung Motors group of France*], [*Korean*], [*Samsung Motors*], [*takeover*], [*takeover of Samsung Motors of Korea*]
- **Step 4: Processing head duplication**
  - [프랑스], [르노 자동차], [프랑스의 르노 자동차 그룹], [한국], [삼성자동차], [한국 삼성자동차 인수]
  - [France], [reuno jadongcha], [France-ui reuno jadongcha geurup], [hanguk], [samsungjadongcha], [hanguk samsungjadongcha insu]
  - [*France*], [*Renault Motors*], [*Renault Samsung Motors group of France*], [*Korean*], [*Samsung Motors*], [*takeover of Samsung Motors of Korea*]
- **Result of Mention Detection**
  - [[프랑스의]$^{1}_{1}$ [르노 자동차]$^{2}_{2}$ 그룹은]$^{0}_{0}$ [[한국]$^{4}_{4}$ [삼성자동차]$^{3}_{3}$ 인수를]$^{5}_{5}$ 공식 제의할 것이다.
  - "[[France-ui]$^{1}_{1}$ [reuno jadongcha]$^{2}_{2}$ geurup-eun]$^{0}_{0}$ [[hanguk]$^{4}_{4}$ [samsungjadongcha]$^{3}_{3}$ insu-lul]$^{5}_{5}$ gongsik jeuihal kes-i-tta."
  - [[*Renault Samsung Motors*]$^{2}_{2}$ *group of* [*France*]$^{1}_{1}$]$^{0}_{0}$ *will suggest* [*takeover of* [*Samsung Motors*]$^{3}_{3}$ *of* [*Korea*]$^{4}_{4}$]$^{5}_{5}$.

determiner phrases must be processed in the Pronominal Coreference Resolution Sieve. Thus, we classify general NPs as nominal mentions and pronouns, and determiner phrases as pronominal mentions. A nominal mention makes an entity through sieves 0 through 7, and in sieve 8, a pronominal mention conducts a pronominal coreference resolution based on the entity.

Table 2 lists the processes for the mention detection that follow each constraint. The table demonstrates the sentence from a document, processing the mention based on the word unit, extending the modifier of the mention, the atomic NE, and processing the duplication of the head word in each sequence step. Finally, the results of mention detection are also demonstrated. Each mention is marked using a square bracket in Table 2, and the final step of the mention detection results indicate the mention index (that is, the subscript) and entity

http://dx.doi.org/10.4218/etrij.16.0115.0896

index (that is, the superscript).

2. Sieve 0: Impossible Coreference

Sieve 0 (that is, Impossible Coreference) is the part that defines the relationship between mentions that were not solved by the coreference resolution during the multi-pass sieve processes. In other words, the mentions defined by the impossible coreference, according to these constraints, do not occur in the coreference resolution with mentions that are relevant to the constraints. The following constraints are shown:

- ***Different last name*** **–** This is defined as an impossible coreference if the last name of two mentions is different for both (for example, 김씨 (*Kim-ssi*) and 이씨 (*Lee-ssi*)).
- ***Different location*** **–** This is defined as an impossible coreference if the location and place information of two mentions are both different.
- ***Different number*** **–** This is defined as an impossible coreference if the information number of two mentions is different for both.
- ***Not i-within-i*** **–** Two mentions are not in an "i-within-i" construct; that is, one cannot be a child NP in the other's NP constituent [12]. However, some words are excluded if the head words are a [name, word, title, designation, appellation, pen name, pseudonym, nom de plume, nickname, or meaning].
- ***Conjunction phrase operation*** **–** Mentions contained in a conjunction phrase conduct a coreference resolution according to the conjunction phrase operation. A conjunction phrase operation is defined as a bitwise operation (that is, an AND operation (AND-op) or OR operation (OR-op)), and all conjunction operations, with the exception of "또는" (*tto-neun*) and "혹은" (*hok-eun*) (that is, or), manage an AND-op. An AND-op is not a coreferent between mentions included in the same conjunction phrase, but an OR-op is. For example, for the sentence "무궁화는 배달계 그리고 단심계, 아사달계로 나뉜다" (*The rose of Sharon is divided into beadalgye, dansimgye, and asadalgye*), we extract the mentions [*무궁화/rose of Sharon*]$^0_0$, [*배달계 그리고 단심계, 아사달계/beadalgye, dansimgye, and asadalgye*]$^1_1$, [*배달계/beadalgye*]$^2_2$, [*단심계/dansimgye*]$^3_3$, [*아사달계/ asadalgye*]$^4_4$. In the sentence, the conjunction is "그리고 (and)." Therefore, the mentions (1 through 4) included in the same conjunction phrase are not coreferent because they are defined as an AND-op.

3. Sieve 1: Exact String Match

This sieve links two mentions into one entity only if they are the same string, including modifiers (for example, [*노란 바나나/yellow banana*]$^2_2$ and [*노란 바나나/yellow banana*]$^2_8$). This sieve is very precise because it has an accuracy of over 90% MUC [2].

4. Sieve 2: Precise Constructs

This sieve turns two mentions into the same entity if any of the following conditions are satisfied:

- **Precision Construct: *Predicate nominative* –** The two mentions (nominal or pronominal) are in a copulative subject-object relationship.
- **Precision Construct: *Role Appositive* –** The two mentions are coreferent if they are in an appositive construction. We can resolve the coreference as follows: the NE label of an antecedent is PERSON, and the NE label (ETRI-Named entity recognition) of a current mention is an OCCUPATION and POSITION (for example, [[*삼도수군통제사/ Historical re-enactment*]$^0_1$ *이순신/Yi Sun-shin*]$^0_0$; in this mention, the two mentions are coreferent).
- **Precision Construct: *Acronym* –** Korean acronyms mix or abbreviate syllables and other elements in the NPs, unlike English acronyms (for example, "Natural Language Processing" is the full name, and "NLP" is its English acronym; in contrast, in Korean, "삼성 자동차" is the full name, and "*삼성차*" is its acronym). We apply the acronym extraction algorithm of [13] to obtain acronyms from the Korean language, and this sieve turns two mentions into the same entity if the two mentions are mapped to each other using the extracted acronym. In this paper, our acronym algorithm uses the noun drop rule, syllable alignment rule, mixed rule, person acronym, and English acronyms [2].

5. Sieves 3–5: Strict Head Match A-C

These sieves link two mentions if the head word of the two mentions is the same, and any of the applicable conditions are satisfied. First, Strict Head Match A as sieve 3 refers to each mention if all conditions are satisfied, and Strict Head Match B as sieve 4 is the exception for only compatible modifiers. Finally, Strict Head Match C as sieve 5 is the exception for word inclusion. We show the following conditions:

- ***Word inclusion*** **–** When two mentions are linked, the antecedent is longer than the current mention. An example of this rule can be seen in the following text:

  "[*나이지리아 남부 아바 마을*]$^6_0$*에서 종족간의 유혈분규로* … [*아바 마을*]$^6_{25}$*에 정통한 정보망을* …"

  "*The bloody affair between species in* [*Nigeria south aba village*]$^6_0$ … *a well-informed person for* [*aba village*]$^6_{25}$ …" Here, it is possible to refer to the two mentions.
- ***Compatible modifiers only*** **–** Two mentions are available for

linking if the modifiers of the current mention are included with the modifiers of the antecedent. An example of this rule can be seen in the following text:

"[*환경에 적응하여 날지 못하는 새*]$^1_0$*는 보통 ... 하지만* [*날지 못하는 새*]$^1_1$*는* ..."

"[*Flightless bird adapted to the environment*]$^1_0$ *is usually ... However,* [*flightless bird*]$^1_1$ *is* ..." Here, it is possible to refer to the two mentions.

"[*어류*]$^0_0$*는 보통 시각에 의존하여 먹이활동을 한다. 하지만* [*눈이 퇴화된 어류*]$^1_1$*는 촉각을 이용하여 먹이활동을 한다.*"

"[*Fish*]$^0_0$ *ordinarily eats prey based on sight. However,* [*degenerated fish eye*]$^1_1$ *eats prey using the sense of touch.*" Here, it is impossible to refer to the two mentions.

- ***Not i-within-i*** – This is similar to sieve 0.

6. Sieve 6: Proper Head Noun Match

This sieve links two mentions if their head word is a proper noun, the mentions have the same head word, and the following constraints are satisfied:

- ***Not i-within-i*** – This is similar to sieve 0.
- ***No location mismatches*** – For this rule, the location information of the two mentions must be equal.
  - The following shows when the current mention matches the location information of the antecedent mention:
    [*대한민국 대통령*]$^0_0$ *이 대중 앞에 나왔을 때는* ... [*한국의 대통령*]$^0_1$ *은* ...
    When [*Korean president*]$^0_0$ *came to the public* ... [*president of Korea*]$^0_1$ *is* ...
  - The following shows when the current mention does not match the location information of the antecedent mention:
    [*프랑스의 르노 자동차그룹*]$^0_0$ *은* ... [*한국의 르노 사*]$^1_1$ *와* ...
    [*Renault S.A*]$^0_0$ *is ... with [Renault Samsung Motors Co., Ltd. of Korea*]$^1_1$ ...
- ***No numeric mismatches*** – The second mention must be equal to the numeric information in the antecedent, for example, [*순례길에 오르는 순례자/pilgrim walked on the Camino de Santiago*]$^0_0$ and [*많은 순례자들/many pilgrims*]$^1_1$ are not coreferents.

7. Sieve 7: Relaxed Head Match

This sieve relaxes the entity head match through heuristics, unlike other head matches, and the conditions are as follows:

- The current mention refers to the antecedent entity if the mention's head matches any word in the entity.
- Two mentions must have the same NE label if their heads are equivalent, for example, the four mentions [*프랑스의 르노 자동차/Renault S.A*], [*르노 삼성 자동차/Renault Samsung Motors*], [*르노사/Renault Co.*], and [*르노/Renault*] have the same NE label as ORGANIZATION, and the mentions indicate the same antecedent entity as "*르노 삼성 자동차/Renault Samsung Motors.*"

8. Sieve 8: Pronominal Coreference Resolution

Korean pronouns take the form of various different pronouns according to the title of honor, and the combination of a determiner and nominal nouns, as the determiner phrase (that is, "*그 동물 /the animal,*" "*이 사람/this person,*" "*저 문/that door,*" and so on) implements the role of the pronoun. Our system gives attributes to mentions, which is the role of pronouns, and the system defines the meaning of a mention. The previous sieves described in this paper prepare the stage for the pronominal coreference by constructing precise entities with shared mention attributes (that is, NE label, head word, and so on). Moreover, we define the animacy, gender, number, and so on, to the attribute of the pronouns using the pronoun dictionary from the Sejong corpus.

The pronominal coreference resolution method of this paper uses the attribute information of defined pronouns (that is, number, person, animacy, NE label, pronoun distance, and so on), a center transition characteristic from the centering theory [14], and the weighting application method from the RAP algorithm [15]. The method for weighting is as follows:

- This condition provides the weight according to the nominative and objective cases of the mention.
- It provides the weight by comparing the attribute of the pronoun mention with the NE label of the antecedent.
- It provides the weight based on the sentence distance between the pronoun mention and antecedent.
- It provides the weight based on the distance of the mention's index between the pronoun mention and antecedent.
- It provides the weight in order to consider the position of the antecedent candidates in a sentence.

The determiner phrase from the pronoun mentions is composed of the grammatical structure used with the determiner and nouns, and the determiner is used with a hypernym higher than the head of the antecedent or the same word. In other words, the determiner phrase mention can use various other features, in addition to pronouns, if the antecedent is expressed by a determiner phrase. We implement a coreference resolution for the mention of the determiner phrase that has an NE label directly, and apply semantic information from the thesaurus in the Sejong corpus to the semantic information of the mention. Such coreference resolution method has the following conditions:

- ***Pronoun String Match*** – This condition provides the weight if the pronouns contain the same text, including determiners.

http://dx.doi.org/10.4218/etrij.16.0115.0896

- ***Pronoun Head String Match*** – This condition provides the weight if the determiner mention and antecedent are the same head word.
- ***Pronoun Head Semantic Match*** – This condition provides the weight if the meanings of the determiner mention and antecedent are similar.
- ***Reflexive and Interrogative*** – The reflexive and interrogative are coreferents to the subject in this sentence.

9. Post Processing

This step discards singleton clusters that do not refer to each other as the same entity, which becomes the case of one mention in one entity.

V. Statistics-Based Coreference Resolution

1. Mention Pair Using Deep Learning

The MP proposed in this paper extracts features by pairing mentions, that is, a candidate antecedent (m_i) and activated mention (m_j), such that the MP determines the possibility of a coreference in the applied ML. The output is a value of $y_{ij} \in [0,1]$. In this paper, deep learning is used for the MP model.

In deep learning, word embedding is generally used for the purpose of a dimension reduction and pre-training in an NLP that has a very high dimension of features. Thus, this paper uses word embedding based on the neural network language model [7]. For example, if the size of the input units is $1 \times V$ and that of the lookup table (LT) is $V \times n$, as shown in Fig. 2, the product of their multiplication has a vector size of $1 \times n$. Here, V is the size of a word dictionary and n is the number of arbitrary dimensionalities (usually 50 dimensions).

Figure 3 shows the feed-forward neural network (FFNN) structure applied in this paper. The input words of two mentions are the head of each noun phrase. In addition, the head is found in the Word Lookup Table, which we use for word embedding. Furthermore, we apply feature embedding, which is analogized with word embedding with regard to the feature set and these two vectors. In sequence, neural networks concatenate each other into a single vector. Subsequently, this vector constitutes a hidden layer by applying the activation function (ReLU) after the vector is multiplied with the weight matrix. This hidden layer informs the output layer as multiplied units of the hidden layer through a weight matrix, and the probability of the output labels is calculated in the output layer using the Softmax function.

In the learning stage, the error rates between the target and output labels extracted by the neural networks are calculated,

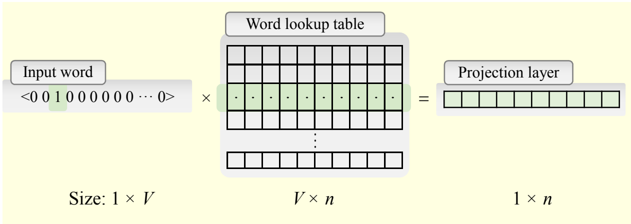

Fig. 2. Dimension reduction using word embedding.

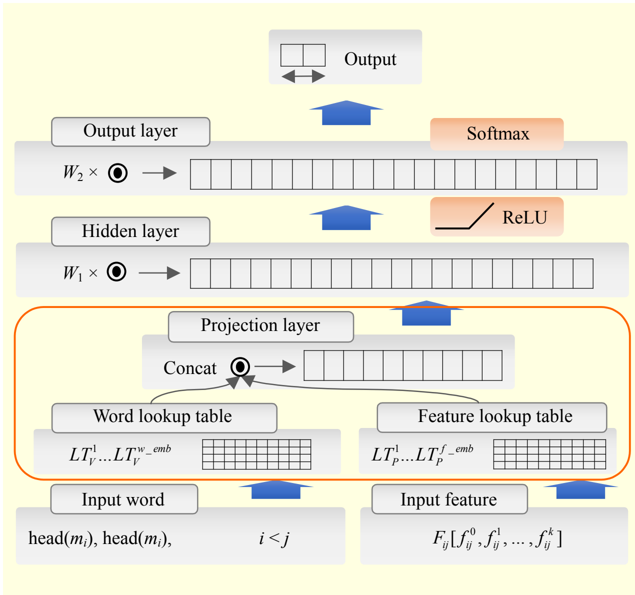

Fig. 3. Feed-forward neural network structure for coreference resolution.

and the weight matrices are updated based on the back-propagated error rates. At this moment, over-fitting is effectively prevented by applying a drop-out. The drop-out sets the state of random units to zero based on the specific probability (normally, 0.5) in the hidden layer.

2. Guided Mention Pair

In this paper, we propose a combination of rule- and statistics-based systems using a guided feature. Guided learning is a method in which the results of other systems are used as the features. In this paper, the results of the rule-based system, that is, the information regarding an entity, is used for the ML features. We extract a significant amount of information from the coreference resolution used from the entities defined during the rule-based multi-pass sieve. Accordingly, we define 36 Korean coreference resolution features that are similar to those in [4], and add the guided feature when the method proposed in this paper is used. Twenty-six feature refer to [5] and remaining 10 features are described in Table 3.

Table 3. Features of coreference resolution in deep learning.

Feature name	Explanation
Left pos 1	POS tag of just before antecedent mention
Right pos 1	POS tag of just after antecedent mention
Left pos 2	POS tag of just before anaphoric mention
Right pos 2	The POS tag of just after anaphoric mention
Same head	C if the mentions' dependency heads have same lemma; else I
Determiner 1	It distinguishes a determiner phrase from an antecedent
Determiner 2	It distinguishes a determiner phrase from a current mention
Deter str match	It distinguishes the same string of a determiner phrase from the both antecedents and a current mention
Both determiners	It distinguishes the determiner phrase from the both antecedents and a current mention
i-with in-i	It distinguishes a possessive relationship from both antecedents and a current mention

```
input: document D
input: coreference model Θ
E = ∅          // set of entities
F = ∅          // set of features
WE = ∅         // set of word embeddings
1.   // all mentions in one document:
2.   M = mentionDetection(D)
3.   // multi-pass sieves
4.   // nominal sieves
5.   E = nominalCoref(M)
6.   // pronominal sieve
7.   E = pronominalCoref(M)
8.   // perform mention pair
9.   for each pair(m_i, m_j) with i < j, m_i, m_j ∈ M do
10.      // map head words to word index
11.      WE = mapWordIndex(m_i, m_j)
12.      // extract features for mention pair
13.      F = extractFeatures(m_i, m_j, E)
14.      if classify(WE, F, Θ) == 1 then
15.          E = merge(m_i, m_j, E)
16.      end-if
17.  end-for
18.  E = postProcessing(E)
19.  output: E
```

Fig. 4. Algorithm combined coreference resolution.

3. Complexity

Figure 4 shows the algorithm used in our coreference resolution system, which is a combination of the rule- and statistics-based systems. This shows each performed step; first, the mention detection (line 2) has a time complexity of $O(N)$ [2] (where N is the number of existing candidate mentions). The multi-pass sieve of lines 5 through 7 shows $O(N^2)$ because it compares the antecedent with the current mention based on existing candidate mentions. When our system applies the mention pair for feature extraction and classification (lines 13 and 15), our system has a time complexity of $O(N^2)$ because it is based on pair (m_i, m_j) for two mentions. Accordingly, the time complexity of our system is shown as $O(N^2)$ for all existing mentions.

VI. Experiment

We start this section with the ETRI corpus, which has been widely used for the evaluation of coreference resolution systems. We continue with the results of experiments analyzing the parameter optimization, compare the results (that is, rule-based systems, SVM, and deep learning-based MP models applied for comparison), and discuss the contribution of each aspect of our approach.

1. Corpus

The ETRI coreference resolution dataset was used to test the Korean coreference resolution, and consists of 153 news domain documents and 767 quiz domain pairs[1]. From these, the training dataset uses 73 news domain documents and 767 quiz domain pairs; 30 news domain documents are used for the validation data, and 50 news domain documents are used for the test data. The ETRI corpus statistics are listed in Table 4. We will release the ETRI corpus at HCLT 2016[2].

2. Evaluation Metrics

We employ the metrics used for the CoNLL-2011 shared task. The CoNLL value (F1 mean value of MUC, B-cube, and Ceaf-e) from [1] is used to measure the performance:

- **MUC** (Vilain and others, 1995): MUC is a link-based metric that measures how many gold and predicted clusters need to be merged to cover the predicted and gold clusters, respectively [16].
- **B^3** (Bagga and Baldwin, 1998): B^3 is a mention-based metric that measures the proportion of overlap between predicted

Table 4. Corpus statistics.

Corpus	Documents	Sentences	Words	Entities	Mentions
ETRI-whole	920	5,429	70,251	6,741	44,251
ETRI-train	840	4,914	59,097	5,780	37,354
ETRI-val	30	196	4,355	365	2,671
ETRI-test	50	319	6,799	596	4,226

1) The quiz documents (that is, quiz domain dataset) feature question-answer pairs consisting of 200 pairs in the Jang-hak quiz, and 567 pairs in WiseQA.
2) The Annual Conference on Human & Cognitive Language Technology.

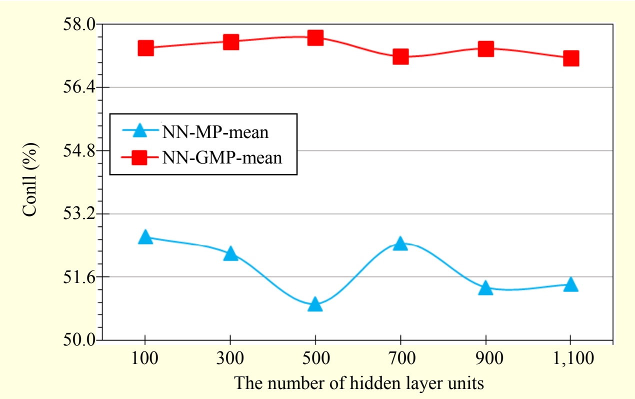

Fig. 5. F1 score on the validation set.

Table 5. NN-MP performance for each the number of hidden layer units in validation set.

	100	300	500	700	900	1,100
Max (%)	52.62	52.19	50.92	52.45	51.34	51.41
Mean (%)	50.66	49.74	49.21	50.60	49.43	50.11
STDEV (% point)	1.62	1.68	1.08	1.13	1.26	0.86

Table 6. NN-GMP performance for each the number of hidden layer units in validation set.

	100	300	500	700	900	1,100
Max (%)	57.41	57.57	57.67	57.19	57.39	57.16
Mean (%)	55.60	55.67	55.61	56.19	55.86	56.40
STDEV (% point)	1.33	1.33	1.04	0.61	0.90	0.64

and gold clusters for a given mention [17].

- **CEAF-e** (Luo, 2005): This entity-based metric enforces a one-to-one alignment between gold and predicted clusters, and measures the best one among the alignments [18].

3. Experimental Results

In this paper, we apply a neural network based MP (NN-MP) and guided mention pair (NN-GMP) using guided features for the coreference resolution, and apply hyper-parameter optimization for each model using the validation set.

Figure 5 shows the CoNLL F1 score for each model in the validation set with respect to the number of hidden units. The experiment was conducted ten times for each parameter because of its tendency to fall into the local minima, and the average and maximum values were then calculated.

In Fig. 5, NN-GMP shows a standard deviation (STDEV) of approximately 0.01 for each number of hidden layer units, and shows the best CoNLL value of approximately 57.67% when the number of hidden layer units is 500 in the validation set. NN-MP shows an STDEV of approximately 0.02 for 100 and 300 hidden layer units, and shows an STDEV of approximately 0.01 for the remaining number of units. In addition, NN-MP shows the optimum CoNLL value of approximately 52.62% when the number of hidden layer units is 100 for the validation set. Tables 5 and 6 (that is, for NN-MP and NN-GMP) show the performance levels of the max, mean, and standard derivations for each number of hidden layer units.

Table 7. Hyper-parameter optimization of the deep learning in the cross validation.

Coreference resolution model	The number of units of the hidden layer	Drop-out	Feature embedding
NN-MP	100	0.5	10
NN-GMP	500	0.5	10

Table 8. Coreference resolution performance each training model.

Coreference resolution models	F1 (mean)	F1 (max)
Rule-based Model	N/A	60.32
SVM-MP	N/A	51.46
NN-MP	55.11	56.07
NN-GMP	60.09	**62.13**

Table 9. Coreference resolution detail performance each training model.

Coreference resolution models	Recall	Precision	F1 (max)
Rule-based Model	**63.74**	57.31	60.32
NN-MP	52.25	**60.78**	56.07
NN-GMP	63.01	61.29	**62.13**

Table 7 lists the hyper-parameters that show the optimum performance under the parameter value applied to the validation set.

Table 8 lists the end results of the coreference resolution based on the hyper-parameters in Table 7 for the test set. In this table, SVM-MP is an MP model based on an SVM that uses the features of Table 3 (with the exception of the guided-link feature). NN-MP is an MP model based on deep learning, and NN-GMP is a model applied using the guided features proposed in this paper.

NN-MP shows a higher level of performance by approximately 4.61% compared with SVM-MP. In addition,

Table 10. Comparison of this work with other coreference resolution approaches on the test set.

Language	Model	MUC			B^3			CEAF-e			CoNLL
		Rec.	Prec.	F1	Rec.	Prec.	F1	Rec.	Prec.	F1	F1
English	Fernandes et al.	65.8	75.9	70.5	51.6	65.2	57.6	50.8	57.3	53.9	60.7
	Chen et al.	63.5	64.0	63.7	66.6	71.5	**69.0**	46.7	46.2	46.4	59.7
	Lee et al.	59.6	60.9	60.3	68.6	73.3	70.9	47.5	46.2	46.9	59.3
	Clack et al.	69.4	76.1	**72.6**	56.0	65.6	60.4	59.4	53.0	56.0	**63.0**
Korean	This work	65.1	62.4	63.7	61.1	57.0	59.0	62.0	58.7	**60.3**	62.1

the NN-GMP proposed in this paper shows approximately a 10.67% higher level of performance compared with SVM-MP, and a 1.81% higher performance level compared with the rule-based model.

Table 9 shows the precision, recall, and F1 value of each model. We can confirm that the recall of rule-based systems is higher than that of NN-MP. On the other hand, NN-MP shows a relatively higher level of precision. Consequently, NN-GMP, which combines the properties of these two methods, improves the performance of the coreference resolution more than the existing methods.

Table 10 compares the results of our system with the following coreference resolution approaches: Fernandes and others (2012) [19], Chen and others (2012) [20], Lee and others (2013) [21], and Model stacking (Clack and others, 2015) [22]. We use a combined coreference resolution system, which is an entity-centric multi-pass sieve algorithm suitable for the Korean language and MP that uses deep learning. Our model shows a CoNLL F1 of approximately 62.1%; it shows a lower performance of approximately 0.9% compared with model stacking, which is a state-of-the-art approach. However, our model shows a higher performance than the first and second grade systems (that is, Fernandes and others, and Chen and others) and the rule-based coreference resolution system of the entity-centric multi-pass sieve (Lee and others). The largest improvement is in the CEAF-e metric. Therefore, we know that our coreference resolution model is suitable for the Korean language because the Korean coreference resolution system proposed in this paper shows a higher level of performance than some English coreference resolution systems and a similar level of performance to the English coreference resolution system.

The time complexity of our system is shown to be $O(N^2)$ for all existing mentions. The agglomerative clustering algorithm (Chen and Ji (2009)) [23] shows a time complexity of $O(N^2)$, where N indicates all existing event mentions. The complexity of Rahimian and Girdzijauskas (2014) [24] is $O(N \times d \times rounds)$, where d is the average node degree and *rounds* is the number of rounds prior to convergence.

VII. Conclusion

In this paper, we proposed an MP model that uses deep learning and a coreference resolution system, which is a combination of rule- and ML-based systems. The experiment results showed a better level of performance in the model proposed in this paper compared with the existing rule-based and MP models, and a similar level of performance to the English coreference resolution system. For future work, we will apply recurrent neural networks to the coreference resolution system, and the mention detection of a coreference resolution.

Reference

[1] H. Lee et al., "Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules," *Comput. Linguistics*, vol. 39, no. 4, 2013, pp. 885–916.

[2] C.-E. Park, K.-H Choi, and C. Lee, "Korean Coreference Resolution using the Multi-pass Sieve," *J. KIISE*, vol. 41, no. 11, Nov. 2014, pp. 992–1005.

[3] A. Haghighi and D. Klein, "Coreference Resolution in a Modular, Entity-centered Model," *Human Language Technol.: Annu. Conf. North American Chapter Assoc. Comput. Linguistics*, Los Angeles, CA, USA, June 1–6, 2010, pp. 385–393.

[4] A. Rahman and V. Ng, "Supervised Models for Coreference Resolution," *Proc. Conf. Empirical Methods Natural Language Process.*, Singapore, Aug. 6–7, 2009, pp. 968–977.

[5] K.-H. Choi, C.-E. Park, and C. Lee, "Coreference Resolution for Korean using Mention Pair with SVM," *KIIES Trans. Comput. Practices*, vol. 21, no. 4, Apr. 2015, pp. 333–337.

[6] S. Pradhan et al., "Conll-2011 Shared Task: Modeling Unrestricted Coreference in Ontonotes," *Proc. Conf. Comput. Natural Language Learning: Shared Task*, Portland, OR, USA, June 23–24, 2011, pp. 1–27.

[7] C. Lee et al., "Korean Dependency Parsing using Deep Learning," *Annu. Conf. Human Cognitive Language Technol.*, Chuncheon, Rep. of Korea, Oct. 7, 2014, pp. 87–91.

[8] R. Collobert et al. "Natural Language Processing (almost) from Scratch," *J. Mach. Learn. Res.*, vol. 12, 2011, pp. 2493–2537.

[9] G.E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Comput.*, vol. 18, no. 7, May 2006, pp. 1527–1554.

[10] G.E. Hinton et al., Improving Neural Networks by Preventing Co-adaptation of Feature Detectors, CoRR, Accessed 2014. https://arxiv.org/abs/1207.0580

[11] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," *Proc. Int. Conf. Artif. Intell. Statistics*, vol. 15, 2011, pp. 315–323.

[12] C. Noam, *Lectures on Government and Binding*, Berlin, Germany: Mouton de Gruyter, 1981.

[13] Y.-C. Yoon, et al., "Construction of Korean Acronym Dictionary by Considering Ways of Making Acronym From Definition," *Proc. KSCS*, 2006, pp. 81–85.

[14] B.J. Grosz, S. Weinstein, and A.K. Joshi, "Centering: A Framework for Modeling the Local Coherence of Discourse," *Comput. Linguistics*, vol. 21, no. 2, 1995, pp. 203–225.

[15] S. Lappin and H.J. Leass, "An Algorithm for Pronominal Anaphora Resolution," *Comput. Linguistics*, vol. 20, no. 4, 1994, pp. 535–561.

[16] M. Vilain et al., "A Model-Theoretic Coreference Scoring Scheme," *MUC6 Proc. Conf. Message understanding*, Columbia, MA, USA, Nov. 1995, pp. 45–52.

[17] A. Bagga and B. Baldwin, "Algorithms for Scoring Coreference Chains," *Int. Conf. Language Resources Evaluation Workshop Linguistics Coref.*, Granada, Spain, May, 1998, pp. 563–566.

[18] X. Luo, "On Coreference Resolution Performance Metrics," *Proc. Conf. Human Language Technol. Empirical Methods Natural Language Process.*, Vancouver, Canada, Oct. 2005, pp. 25–32.

[19] E.R. Fernandes, C.N. Dos Santos, and R.L. Milidiu, "Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution," *Joint Conf. EMNLP CoNLL-Shared Task*, Jeju, Rep. of Korea, July 12–4, 2012, pp. 41–48.

[20] C. Chen and V. Ng, "Combining the Best of Two Worlds: A Hybrid Approach to Multilingual Coreference Resolution," *Joint Conf. EMNLP CoNLL-Shared Task*, Jeju, Rep. of Korea, July 2–4, 2012, pp. 56–63.

[21] H. Lee et al., "Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules," *Comput. Linguistics*, vol. 39, no. 4, Nov. 2013, pp. 885–916.

[22] K. Clark and C.D. Manning, "Entity-Centric Coreference Resolution with Model Stacking," *Proc. Annu. Meeting Assoc. Comput. Linguistics Int. Joint Conf. Natural Language Process.*, Beijing, China, July 27–31, 2015, pp. 1405–1415.

[23] Z. Chen, H. Ji, and R. Haralick, "Event Coreference Resolution: Algorithm, Feature Impact and Evaluation," *Proc. Events Emerging Text Types*, Stroudsburg, PA, USA, 2009, pp. 1–8.

[24] F. Rahimian, S. Girdzijauskas, and S. Haridi, "Parallel Community Detection for Cross-Document Coreference," *Web Intell. Intell. Agent Technol*, Warsaw, Poland, Aug. 11–14, 2014, pp. 46–53.

Cheoneum Park received his BS and MS degrees in computer science from Kangwon National University, Chuncheon, Rep. of Korea in 2014 and 2016, respectively, where he is currently a PhD student. His research interests are in natural language processing, information extraction, question answering, machine learning, and deep learning.

Kyoung-Ho Choi received his BS and MS degrees in computer science from Kangwon National University in 2016, where is currently working toward his PhD. His research interests are in natural language processing, sentiment analysis, and deep learning.

Changki Lee received his BS in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea in 1999. He received his MS degree and PhD in computer engineering from POSTECH, Pohang, Rep. of Korea in 2001 and 2004, respectively. From 2004 to 2012, he was a researcher with the Electronics and Technology Research Institute (ETRI), Daejeon, Rep. of Korea. Since 2012, he has been a professor of Computer Science at Kangwon National University. His research interests include natural language processing, machine learning, and deep learning.

Soojong Lim received his BS in mathematics from Yonsei University, Seoul, Rep. of Korea, in 1997. He also received his MS degree and PhD in computer science from Yonsei University in 1998 and 2014, respectively. He is currently a principal researcher at ETRI. His research interests include natural language processing, machine learning, and question answering.