World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:3, 2016

A Recognition Method for Spatio-Temporal Background in Korean Historical Novels

Seo-Hee Kim, Kee-Won Kim, Seung-Hoon Kim

***Abstract*—**The most important elements of a novel are the characters, events and background. The background represents the time, place and situation that character appears, and conveys event and atmosphere more realistically. If readers have the proper knowledge about background of novels, it may be helpful for understanding the atmosphere of a novel and choosing a novel that readers want to read. In this paper, we are targeting Korean historical novels because spatio-temporal background especially performs an important role in historical novels among the genre of Korean novels. To the best of our knowledge, we could not find previous study that was aimed at Korean novels. In this paper, we build a Korean historical national dictionary. Our dictionary has historical places and temple names of kings over many generations as well as currently existing spatial words or temporal words in Korean history. We also present a method for recognizing spatio-temporal background based on patterns of phrasal words in Korean sentences. Our rules utilize postposition for spatial background recognition and temple names for temporal background recognition. The knowledge of the recognized background can help readers to understand the flow of events and atmosphere, and can use to visualize the elements of novels.

***Keywords*—**Data mining, Korean historical novels, Korean linguistic feature, spatio-temporal background.

I. Introduction

THE most important elements of a novel are the characters, events and background. Among them the background means the time and place that a character thinks and acts, or the situation that a character faces. The basic elements of a background are spatial background and temporal background. Spatial background means the environment in which the event occurred, such as nation, region and geographical feature. Temporal background means the period in which the event occurred or the character acted, such as historical period and seasonal background.

In the previous studies, a statistical method based on machine learning such as Hidden Markov Model, Maximum Entropy Based Model, Support Vector Machines, Conditional Random Field Based Models is used to recognize the spatio-temporal information in documents written in non-Korean languages [1]-[3]. In addition, the dictionaries of time and date are used to recognize the temporal information by using the tagged words [4]. In another studies, the dictionaries of place-names are used to recognize the spatial information in Korean newspaper [5], [6]. But, in cases of previous studies, it is difficult to recognize the information about the historical place-names or era because it analyzes about present spatio-temporal information.

In this paper, we first build a Korean historical national dictionary. We also presented a method for recognizing spatio-temporal background based on patterns of phrasal words in Korean sentences. We use the method based on [7] to recognize spatial background. Our method for temporal background recognition utilizes temple names. We are targeting Korean historical novels because spatio-temporal background especially performs an important role in historical novels among the genre of Korean novels. The knowledge of the recognized background may be helpful for understanding the flow of events and atmosphere and choosing a novel that readers want to read.

II. Method for Recognizing Spatio-Temporal Background in Korean Historical Novels

The pseudo-algorithm for recognizing spatio-temporal background is shown in Fig. 1.

```
algorithm SpatioTemporalBackgroundRecognition(novel) {
 do Preprocessing
 applying Korean_Historical_National_Dictionary
 Spatial_Background = Spatial_Background_Recognition()
 Temporal_Background = Temporal_Background_Recognition()
 print Spatial_Background & Temporal_Background
}

function Spatial_Background_Recognition() {
 applying Postposition_Pattern
 recognizing Spatial_Background
 return(Spatial_Background)
}

fucntion Temporal_Background_Recognition() {
 converting Temple_Name_to_Year
 recognizing Temporal_Background
 return(Temporal_Background)
}
```

Fig. 1 Pseudo-algorithm for recognizing spatio-temporal background in Korean historical novels

Seo-Hee Kim is with the Department of Computer Science, Dankook University, Yongin, Gyeonggi, 16890, Republic of Korea (e-mail: rlatjgml38@naver.com).

Kee-Won Kim is with the College of Convergence Technology, Dankook University, Yongin, Gyeonggi, 16890, Republic of Korea (e-mail: nirkim@dankook.ac.kr).

Seung-Hoon Kim is with the Department of Applied Computer Engineering, Dankook University, Yongin, Gyeonggi, 16890, Republic of Korea, (corresponding author, phone: +82-31-8005-3681; fax: +82-31-8021-7422; e-mail: edina@dankook.ac.kr)

The algorithm performs pre-processing on the input Korean historical novel. After that, the Korean historical national dictionary is applied. It recognizes spatial background by using postposition pattern and frequency. It converts temple name derived from the novel to A.D. year. Then, it recognizes temporal background by using frequency and average calculating.

A. Pre-Processing

Sentences used in novels may be written in a different punctuation marks and have different sentence length depending on the publishers or authors. So, the algorithm separates all sentences by punctuation marks for matching all the sentences in the same format. The used punctuation marks are period, comma, question mark, exclamation mark, and quotation mark.

B. Applying Korean Historical National Dictionary

In this paper, we build a Korean historical national dictionary for recognizing spatio-temporal background in Korean historical novels. Our dictionary has historical places, territory names, temple names and year on the throne of crowned king over many generations in Korean history unlike previous dictionaries only including the currently existing places or temporal words [5], [6].

The algorithm recognizes spatio-temporal background candidates by applying Korean historical national dictionary. It extracts frequency of spatial background candidates in the novel by applying historical places and territory names in Korean historical national dictionary. Also, it extracts the year on the throne of the temple name by applying temple names in the dictionary.

C. Applying Postposition Pattern for Recognizing Spatial Background

Korean language has postposition called “조사 (josa)”. It is attached to word and represent the meaning or grammatical relationship between two words. It uses postpositions "에 (to) " and "에서 (from)" that represent a spatial background. It attaches <TO> tag to the postposition "에 (to)" and <FROM> tag to the postposition "에서 (from)".

Now, the algorithm separates sentences by spacing word for recognizing spatial background on the pre-processed Korean historical novel. If the separated word is attached by <TO> tag or <FROM> tag, it recognizes the word as a spatial background candidate. In this process, the non-spatial background word also can be recognized as a spatial background candidate because of Korean linguistic feature such as "주머니에 (in the pocket) <TO>" and "주머니에서 (from the pocket) <FROM>". Therefore, if a word that appears after <TO> or <FROM> tag composes of "Verb+Ending", it recognizes the word located before the postposition as a spatial background candidate. In addition, it extracts the frequency of the spatial background candidates used in the novel.

D. Recognizing Spatial Background

Our scheme recognizes the typical spatial background for the Korean historical novel according to frequency of spatial background candidates recognized by Korean historical national dictionary and postposition pattern.

If spatial background candidates, c_i, are a total of n, it compares the frequency, f_i, of the c_i used in the novel. The c_i with the highest f_i is recognized as the typical spatial background of the Korean historical novel. It is expressed as (1):

$$spatial\ \ background\ \ = \max_{1 \le i \le n} \{ f_i \} . \quad (1)$$

E. Converting Temple Name to Year for Recognizing Temporal Background

In this paper, the proposed algorithm recognizes temporal background as year. In Korean historical novels, the year is not only expressed by using number, e.g. 2016, but it is also expressed by using temple name, e.g. "세종 7년 (King Sejong 7 years)". The temple name is a name given after king’s death. A number attached behind the temple name of king represents the period after he ascended the throne. So the algorithm separates sentences based on spacing word on the pre-processed Korean historical novel. And it recognizes the separated word shown in Table I as temporal background candidates.

TABLE I
THE FORMS THAT RECOGNIZED TO TEMPORAL BACKGROUND CANDIDATES

Form	Tag	Examples
Number+년 (year)	<A.D.>	1997년(year), 689년(year), etc.
Temple Name +Number+년 (year)	<TEMPLE>	세종7년(King Sejong seven years), 보장왕27년 (King Bojang twenty seven years), etc.
(Number)	<NUM>	(1872), (668), etc.

As one can see in Table I, there are three ways of expressing temporal background by using tags in Korean historical novels. So, a candidate with <TEMPLE> tag cannot be compared with the one with the other tags. The proposed algorithm converts temple name with year to A.D. year. The A.D. year of the temple name j's year on the throne, K_j^1, is expressed as $AD(K_j^1)$. For a candidate with <TEMPLE> tag, K_j^t, it representation to A.D. year is expressed as (2):

$$AD(K_j^t) = AD(K_j^1) + t - 1 \quad if\ \ t > 0 . \quad (2)$$

For example, "세종 7년 (King Sejong seven years)", temple name j is "세종 (King Sejong)" and the year on the throne, $AD(K_j^1)$, is 1418. According the equation, the final value of the $AD(K_j^7)$ will be 1424.

F. Recognizing Temporal Background

The algorithm sorts the year and the converted year from temple name in ascending order and extracts the highest frequency range. And the average of years in that range is computed by the algorithm. It recognizes the average year as the typical temporal background of the Korean historical novel.

III. Experimental Results

A. Procedure of Experiment

To evaluate the accuracy of our proposed method, we experimented with twelve Korean historical novels. The file format of the novels is UTF-8. It removes the contents not in the body such as title, bibliographical information, table of contents, author's words, and addendum for precise experiment.

We show an example of experiment on the Korean historical novel "고구려를 위하여 1 (For Goguryeo 1)" in Fig. 2. As a result, spatial background of " 고구려를 위하여 1 (For Goguryeo 1)" is "고구려 (Goguryeo)" and temporal background is "About 666".

B. Result of Experiment

Table II is result of spatio-temporal background recognition about Korean historical novels by applying our proposed method. The novels used in the experiment are twelve.

As the result of spatial background recognition shows that ten out of twelve Korean historical novels were matched. It showed accuracy of 83.33%. On the other hand, the result of temporal background recognition shows that eight out of twelve novels were matched. It showed accuracy of 66.67%.

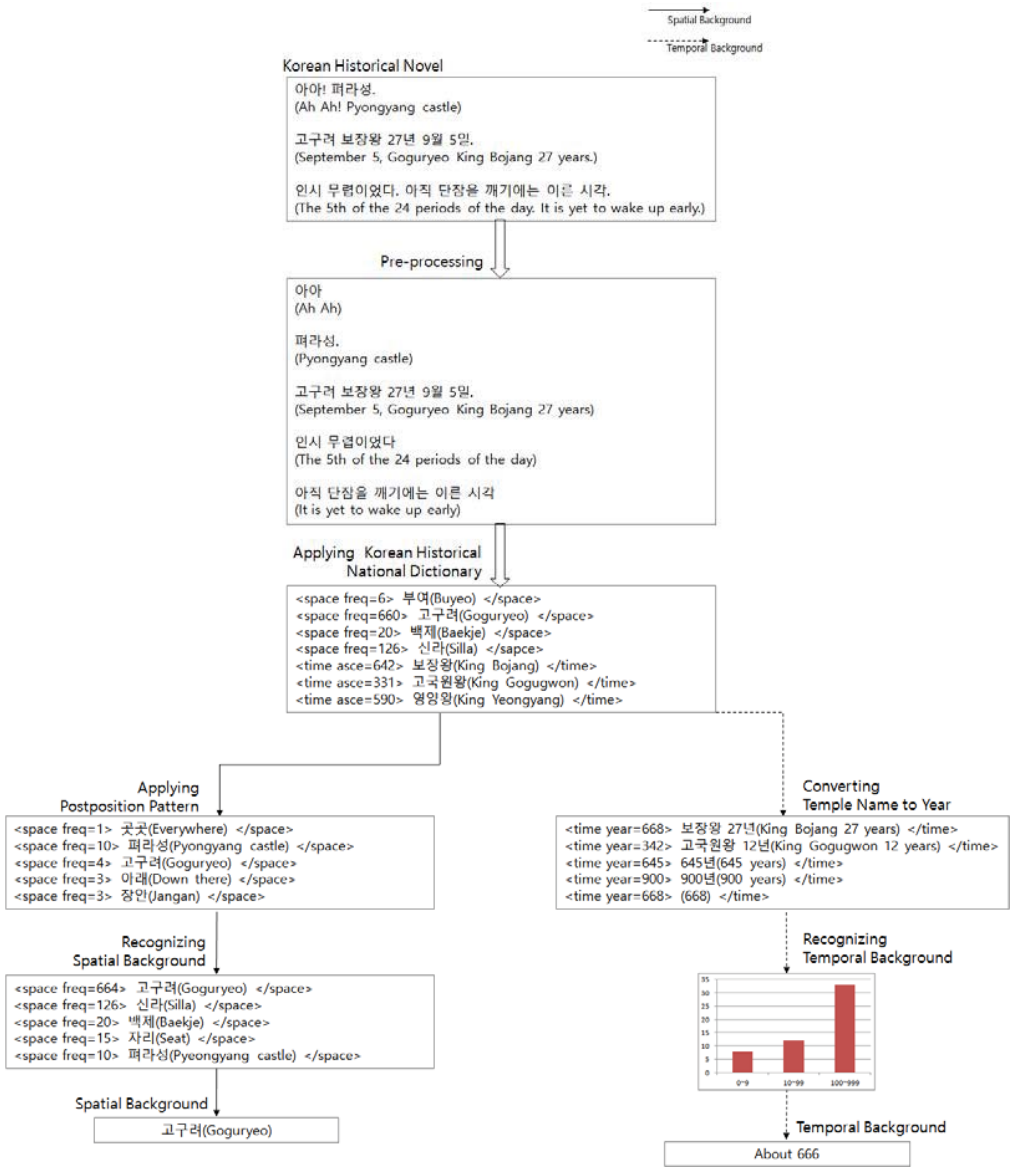

Fig. 2 Example of experiment on "고구려를 위하여 1 (For Goguryeo 1)"

TABLE II
RESULT OF SPATIO-TEMPORAL BACKGROUND RECOGNITION ABOUT KOREAN HISTORICAL NOVELS

Title	Recognized Spatial Background	Spatial Background Matching Result	Recognized Temporal Background	Temporal Background Matching Result
고구려 무사 1 (A Warrior of Goguryeo 1)	고구려 (Goguryeo)	T	About 523	T
고구려를 위하여 1 (For Goguryeo 1)	고구려 (Goguryeo)	T	About 666	T
대한독립기 1 (Independence Notes of Korea 1)	조선 (Joseon)	F	About 1923	T
대한독립기 2 (Independence Notes of Korea 2)	한반도 (Korean Peninsula)	T	About 1994	F
대한독립기 3 (Independence Notes of Korea 3)	한반도 (Korean Peninsula)	T	About 1928	T
대한독립기 4 (Independence Notes of Korea 4)	한반도 (Korean Peninsula)	T	About 2032	F
대한독립기 5 (Independence Notes of Korea 5)	대한제국 (Korean Empire)	F	About 1935	T
동학농민운동 (Donghak Peasant Revolution)	조선 (Joseon)	T	About 1881	T
미실 (Lady Mishil)	신라 (Silla)	T	About 561	T
백범일지 (Baekbeom's Daily Record)	조선 (Joseon)	T	About 15	F
추사 김정희 (1) (Chusa Kim Jung-Hee (1))	조선 (Joseon)	T	About 1635	T
홍길동전 (Hong Kil-Dong)	조선 (Joseon)	T	-	F

IV. CONCLUSION

In this paper, we built a Korean historical national dictionary and proposed a method for recognizing spatio-temporal background about Korean historical novels. The method consists of applying postposition pattern and converting temple name. To evaluate the accuracy of our proposed method, we were experiment with twelve Korean historical novels. As the result of spatial background recognition showed accuracy of 83.33%, on the other hand, the result of temporal background recognition showed accuracy of 66.67%.

In the future work, we will study about other expressions for recognizing spatio-temporal background and increase accuracy for Korean historical novels.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2015.

REFERENCES

[1] K. G. Srinivasagan, S. Suganthi and N. Jeyashenbagavalli, "An Automated System for Tamil Named Entity Recognition Using Hybrid Approach," *International Conference on Intelligent Computing Applications (ICICA)*, 2014, pp. 435-439.

[2] G. Prasad and K. K. Fousiya, "Named Entity Recognition Approaches: A Study applied to English and Hindi Language," *International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, 2015, pp. 1-4.

[3] Y. J. Zhang and T. Zhang, "Me-Based Chinese Person Name and Location Name Recognition Model," *International Conference on Machine Learning and Cybernetics*, vol. 6, 2007, pp. 3441-3447.

[4] B. T. Todorovic, S. R. Rancic, I. M. Markovic, E. H. Mulalic and V. M. Ilic, "Named Entity Recognition and Classification using Context Hidden Markov Model," *9th Symposium on Neural Network Applications in Electrical Engineering*, 2008, pp. 43-46.

[5] H. S. Jang, K. C. Jung, J. K. Lee and K. H. Park, "Recognition of Korean Place Names on the Internet by Using the Rules of Dictionary Use," *Korean Society for Internet Information*, vol. 6, no. 1, 2005, pp. 397-400.

[6] K. H. Lee, J. H. Lee, M. S. Choi and G. C. Kim, "Study on Named Entity Recognition in Korean Text," *Proceedings of the 12th Annual Conference on Human and Cognitive Language Technology*, 2000, pp. 292-299.

[7] S. H. Kim and S. H. Kim, "A Recognition Method for Korean Spatial Background in Historical Novels", *Journal of Information Technology services*, submitted.