Journal of Intelligent & Fuzzy Systems 24 (2013) 251–259
DOI:10.3233/IFS-2012-0551
IOS Press

Web intelligence for tourism using railway data by a simplified fuzzy reasoning method

Takashi Hasuike[a,*] and Takumi Ichimura[b]
[a]*Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, Japan*
[b]*Faculty of Management and Information Systems, Prefectural University of Hiroshima, Hiroshima, Japan*

Abstract. This paper proposes an approach to analyze the tourism information and data derived from the Web, particularly seat availability data of bullet trains in Japan, and to discover some useful knowledge for the tourism. For the fast development of information and communication technologies, the relation between the web data and tourism is inseparable. However, the Web data include various types of information such as numerical, linguistic, and graded data. Furthermore, the expert tourism planner's subjectivity is also an important factor to develop new favorable plans. A simplified fuzzy reasoning method, which is one of the useful approaches in Data mining, is introduced in order to deal with these data mathematically. The analysis of the tourism data and the knowledge discovery are performed using actual data of bullet trains in Japan.

Keywords: Tourism information, Web intelligence, Fuzzy reasoning method, Data mining

1. Introduction

The fast development of information and communication technologies (ICTs) and the global spread of the Internet and the electronic business have changed industry structures around the world. Then, devices using theses ICTs such as personal computers and mobile phones have become increasingly accessible to large populations, and the means by which consumers search for tourism information has shifted dramatically over the years. Therefore, the accelerating and synergistic interaction between ICTs and tourism has affected various sectors of economy and brought fundamental changes by creating new opportunities to tourism companies in marketing, tourism destination planning and advertising [2, 3, 6, 9, 11]. Recently, some researchers have studied the relation between Web data and tourism (most recently, [1, 4, 5, 7, 15]).

For tourism planners, the Internet provides ways to globally sell their products to potential travelers at any time and to develop tour plans compatible with tourist's needs from Web data. These suppliers can remotely control their servers to display information on products or services at a high speed. For travelers, the Internet allows them to communicate directly with tourism suppliers to request information, and to search and purchase products or services at any time and any place online [10]. Thus, tourists and suppliers can collect and deliver various information and data using the Internet. As a practical example, websites of hotel reservation encourage users to express opinions on tourism services like hotels by posting feature ratings and textual reviews. These numerical ratings are often used by recommender systems to recommend highly rated hotels, assisting users in making decisions. The general approach is to

*Corresponding author. Takashi Hasuike, Graduate School of Information Science and Technology, Osaka University, 2-1 Yamadaoka, Suita, 565-0871 Osaka, Japan. E-mails: thasuike@ist.osaka-u.ac.jp (Takashi Hasuike) and ichimura@ieee.org (Takumi Ichimura).

compute only the accurate numerical information given by users to provide a ranking value of these hotels and their features.

On the other hand, for tourism planners, it is important to understand the flow of tourists visiting a sightseeing place as well as the availability ratio of accommodations. In this case, transportation data such as railways, airlines, and so on, are also important, because transportation data are closely related to the availability ratio of accommodations. In Japan, it is not difficult to collect such transportation data from railway and airline companies. However, as a website of Japan Railway (JR) company [8] shown in Fig. 1, the availability information is provided as separate classes; "full", "a few", and "the others", but we cannot understand how many seats are vacant in the case of "a few". Therefore, even if tourism planners cannot collect the detailed data of transportations, they often must develop favorable tour plans.

Thus, the current ICTs can also collect various data sets including not only numerical values but also language, evaluation, picture, etc. The theoretical methods to discover knowledge from these various data sets are known to Data mining, and several useful approaches have been proposed, for instance Regression analysis, Bayesian theory, Neural Network, Clustering. They have widely used in many fields such as marketing and medical analysis as well as tourism. Furthermore, in the case of vague or imprecise knowledge, a better approach may be to use linguistic assessments instead of numerical values. As one useful approach for the vague or imprecise knowledge, a fuzzy theory was introduced by Zadeh [17]. The fuzzy theory makes it possible to numerically define qualitative information based on the concept of linguistic variable. The fuzzy theory has been used successfully in many problems. Particularly, a fuzzy reasoning method is the most important approach to extract and decide effective rules under fuzziness mathematically, and it has been widely used in many fields such as control theory, data mining, and decision making. It is one of rule-based approaches with consequent fuzzy rules. Mamdani's fuzzy reasoning [13], functional fuzzy reasoning including Takagi-Sugeno (TS) fuzzy reasoning [14], and simplified fuzzy reasoning are called the direct approach, and are in widespread use for fuzzy control and fuzzy expert systems. Most recently, the fuzzy reasoning method is applied to researches of Web text classification [12] and Semantic Web data [18]. In this paper, we apply the simplified fuzzy reasoning method to data sets of tourism, because this approach is one of the simplest and most useful fuzzy reasoning methods in terms of practical usage. The remainder of this paper is organized as follows. In Section 2, we briefly introduce a simplified fuzzy reasoning method. In Section 3, we introduce tourism data derived from the Web, particularly occupied rate of reserved seats in bullet trains called Japanese Super Express or Shin-Kansen in Japan. Furthermore, we set membership functions to each input factor in order to quantify categorical data and ambiguity of border lines among the categories, and to use the simplified fuzzy reasoning method as an analysis tool. In Section 4, we compare our proposed approach with a linear regression model using the first method of quantification which is one of standard statistical analysis approaches for the categorical data. We analyze the dataset from 2010 to 2011 to predict the future trend of tourists and discover specific knowledge from the dataset. Finally, in Section 5, we conclude this paper.

2. Simplified fuzzy reasoning method

In this section, we introduce a mathematical formulation of simplified fuzzy reasoning method. Until now, many researchers have proposed various fuzzy inference and reasoning methods based on or extending Mamdani's study [13] or Takagi and Sugeno's study [14]. In this paper, as a mathematically simple approach of consequent parts in these studies, we introduce a simplified fuzzy reasoning method whose consequent parts are given as constant real values. In this method, m rule modules are given as follows:

$$\begin{aligned}\text{Rule-}i:\ & if\ x_1\ is\ A_1^i,\ x_2\ is\ A_2^i, \dots, x_n\ is\ A_n^i,\\ & then\ y_1\ is\ w_1^i,\ y_2\ is\ w_2^i, \dots, y_k\ is\ w_k^i,\ (i=1,2,\dots,m)\end{aligned} \tag{1}$$

where $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ is the input data set, and $\boldsymbol{y} = (y_1, y_2, \dots, y_k)$ is the consequent data set, respectively. Then, w_1^i is the real value of output for the consequent part in the ith fuzzy rule. A_j^i is the fuzzy set for the jth input data of the Rule-i. The degree of the antecedent part to the jth input data of Rule-i is obtained as $h_j^i = A_j^i(x_j)$. In the simplified fuzzy reasoning method using the multiple calculation of weights h^i, the inference result y_s, $(s = 1, 2, \dots, k)$ is calculated as follows:

11 月 23 日 東京 → 京都

空席案内						
列車名	発時刻	着時刻				
のぞみ　7号<全席禁煙>	06:50	09:11	○	＊	○	＊
のぞみ　２０３号	07:00	09:21	△	○	△	○
ひかり　４６１号	07:03	09:48	×	○	○	○
のぞみ　9号<全席禁煙>	07:10	09:28	○	＊	○	＊
のぞみ　３０７号	07:13	09:34	○	○	○	○
のぞみ　１１号<全席禁煙>	07:30	09:51	×	＊	○	＊
ひかり　５０３号	07:33	10:15	×	△	△	○
のぞみ　３１１号	07:47	10:08	○	○	○	○
のぞみ　１３号<全席禁煙>	07:50	10:11	○	＊	○	＊
のぞみ　２０７号<全席禁煙>	08:00	10:21	○	＊	○	＊
ひかり　４６３号	08:03	10:48	×	○	○	○
のぞみ　１５号<全席禁煙>	08:10	10:28	○	＊	○	＊

【凡例】
○：空席があります。　△：空席が残りわずかです。　×：満席です。
ー：該当する席はありません。または予約できません。
＊：喫煙席はありませんが、喫煙ルームが設置されています。

前日を検索　1時間前を検索　1時間後を検索　翌日を検索
条件を変えて再検索
戻る

23 November, 2011 Tokyo -> Kyoto						
Unoccupied seats guide			Reserved car		Green car (Speical seat)	
Train Number	Departure time	Arrival time	Non smoking	Smoking	Non smoking	Smoking
NOZOMI No.7	6:50	9:11	○	*	○	*
NOZOMI No.203	7:00	9:21	△	○	△	○
HIKARI No.461	7:03	9:48	×	○	○	○
NOZOMI No.9	7:10	9:28	○	*	○	*
NOZOMI No.307	7:13	9:34	○	○	○	○
NOZOMI No.11	7:30	9:51	×	*	○	*
HIKARI No.503	7:33	10:15	×	△	△	○
NOZOMI No.311	7:47	10:08	○	○	○	○
NOZOMI No.13	7:50	10:11	○	*	○	*
NOZOMI No.207	8:00	10:21	○	*	○	*
HIKARI No.463	8:03	10:48	×	○	○	○
NOZOMI No.15	8:10	10:28	○	*	○	*

Fig. 1. Availability information of vacant seats from JR CYBER STATION on 23 November, 2011. The above figure is the original website, and the below is translated the above figure to English. State labels of seat availability “×”, “△”, and “◯” are set. Label “×” is the state of “full” occupancy, label “△” is the state of “a few” vacant seats, and label “◯” is the state of the others. Label “∗” means that smoking cars are not assigned in the bullet train.

$$y_s = \frac{\sum_{i=1}^{m} h^i w_s^i}{\sum_{i=1}^{m} h^i}, (s = 1, 2, \ldots, k),$$
$$h^i = A_1^i(x_1)A_2^i(x_2)\cdots A_n^i(x_n) \qquad (2)$$

From the fuzzy reasoning method, we not only obtain effective rules from expert's subjectivity but also discover unexpected knowledge from specific data which is inapplicable to given rules. In order to deal with the simplified fuzzy reasoning method, we must determine membership functions of fuzzy sets A_j^i. The following two approaches to determine membership functions are mainly used: (i) Setting membership functions according to expert's subjectivity, and (ii) Setting membership functions directly using learning algorithm based on input and consequent data. In this paper, in terms of using tourism planner's subjectivity, membership functions in the simplified fuzzy reasoning method are assumed to be initially determined according to expert tourism planner's subjectivity.

3. Analysis of Web data for tourism using the fuzzy reasoning method: Dataset of bullet train in Japan

We apply the simplified fuzzy reasoning method to tourism data from the Web, and extract some useful and specific features by analyzing this tourism data. To simplify the following discussion, we focus on tourism data at Kyoto in this analysis, and deal with dataset of bullet train in Japan derived from JR CYBER STATION. Kyoto is one of the most famous and historical sightseeing cities in the world, and many tourists visit to Kyoto not only all over the world but also from many places in Japan. Most of tourists visiting Kyoto in Japan, particularly from not only Metropolitan area including Tokyo but Kyushu area, takes the bullet train, because there is no airport in Kyoto and the quickest access to Kyoto is to use bullet trains. Surely, some tourists visit to Kyoto by bus, but the rate of tourists to use buses is much smaller than bullet trains. Therefore, in this paper, we deal with bullet train data of Kyoto as important tourism data.

Before analyzing railway data, we must set input and consequent data for the fuzzy reasoning method. The main object is to predict vacant seat data of bullet train, because tourism planners will perform the discount of an accommodation price and new tourism campaigns according to this prediction. Therefore, in this paper, we set input and consequent data in the next subsections.

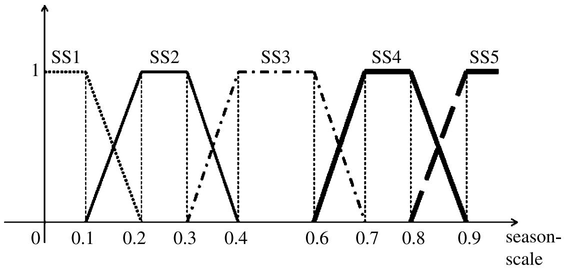

Fig. 2. Membership functions of season-scale index to the input data. Membership functions represent "small" (SS1), "a little small" (SS2), "medium" (SS3), "a little large" (SS4), and "large" (SS5) index from left, respectively.

3.1. Input data

As input data to analyze the tourism data of Kyoto using the fuzzy reasoning method, we consider the following three factors.

3.1.1. Season-Scale (SS) index

We set membership functions shown in Fig. 2 considering two factors: (a) season factors such as on-season or off-season, weekends or not based on the data, and (b) event scale factors such as size, category, historical background, and number of participants of the event derived from statistical data. First we collect periods of high season and low season of bullet train from the website of JR company, and event scale factors from the Web or statistics published by ministries and organizers of event. Then, by using the statistical analysis such as the regression or experts' comprehensive judgment, set a value between 0 and 1 as a season-scale (SS) index. The maximum value will be obtained from one of the high seasons or largest events in Japan. In this paper, we assume that the value of SS index is initially given using the statistical theory. After the step, we set membership functions shown in Fig. 2 with "large", "a little large", "medium", "a little small", or "small" to the value of SS index.

In this membership function, we assume that the center value of season-scale 0.5 is also located in the center of membership function of season-scale index SS3, and the range whose membership value is 1 is twice as the others, because season-scale index SS3 is applied to an ordinary season, neither high season nor low season. Then, we also assume that the other membership functions except for the season-scale index SS3 are located at even intervals. In general and basic fuzzy reasoning methods, it is often to locate all membership functions

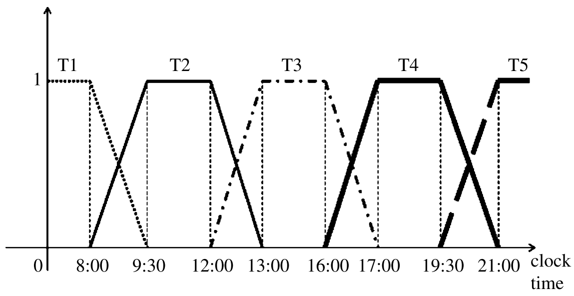

Fig. 3. Membership functions for the departure time in a day. Membership functions represent "early morning" (T1), "morning" (T2), "afternoon" (T3), "evening" (T4), or "late-evening" (T5) from left, respectively.

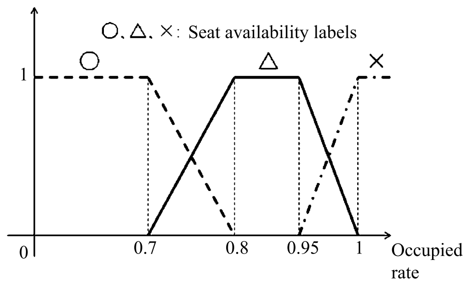

Fig. 4. Membership functions for states of seat availability in a bullet train. The dashed line represents the state of "full" occupancy, the solid line represents the state of "a few" vacant seats, and the dotted line represents the state of the others.

at even intervals, and hence, these assumptions are not singular at all.

3.1.2. *Departure time*

Membership functions, which are applied to represent the degree of membership whether each departure time of input data is in "early morning", "morning", "afternoon", "evening", or "late-evening", are shown in Fig. 3.

The clock time for many Japanese people to distinguish T1 from T2 will be between 8 : 30 to 9 : 00, because almost all schools and companies start between 8 : 30 to 9 : 00. Therefore, we set membership functions of T1 and T2 as Fig. 3. In a way similar to this assumption, according to a typical Japanese daily lifestyle, we set the other membership functions of T3, T4, and T5 as Fig. 3.

3.1.3. *Occupied rate of reserved seats in a bullet train from seven to four days before*

Membership functions which consist of "full", "a few", and "the others" on the occupied rate are shown in Fig. 4.

From our field research of bullet train's seat availability, the lines to distinguish label "◯" from "Δ" and "Δ" from "× " are different every day, but occupied rate to label "Δ" is included in about [0.75, 0.975] from our field research of ticket-vending machines on some days. Therefore, we set membership function as Fig. 4.

3.2. *Consequent data*

As a consequent data of the fuzzy reasoning method, we introduce an occupied rate of reserved seats in a bullet train one day before. By predicting the occupied rate of seats in a bullet train one day before using various

Table 1
Some rules from the dataset to occupied rate of seats in bullet trains

Rule	SS	Time	7 day	6 day	5 day	4 day	1 day
Rule-1	SS4, SS5	T1, T2	Δ	Δ	×	×	×
Rule-2	SS4, SS5	T1, T2	×	×	Δ	Δ	×
Rule-3	SS4, SS5	T1, T2	Δ	Δ	Δ	×	×
Rule-4	SS4, SS5	T1, T2	◯	◯	Δ	Δ	Δ
Rule-5	SS5	T1	×	Δ	Δ	Δ	Δ
Rule-6	SS5	T1	Δ	Δ	Δ	Δ	×
Rule-7	SS4, SS5	T1, T2	◯	◯	◯	Δ	Δ
Rule-8	SS5	T2	◯	Δ	Δ	Δ	Δ
Rule-9	SS4	T1	×	×	◯	◯	×
Rule-10	SS4	T1	Δ	×	◯	◯	Δ

data from seven to four days before, the tourism planners will grasp more detailed information of the total number of tourist visiting the sightseeing place, and preliminarily perform some actions to attract more tourists such as the discount of an accommodation price and new events at souvenir stores.

4. Analysis of dataset from 2010 to 2011 and knowledge discovery

Using these input and consequent data in Section 3 and applying the simplified fuzzy reasoning method in Section 2, we analyze the datasets to occupied rate of seats in bullet trains from Tokyo to Kyoto in 2010 and 2011. First, from the dataset in 2010 and expert's subjectivity, the following partial rules are provided, particularly for the large and small large SS index, and in the early morning and morning.

In Table 1, "SS4, SS5" in the SS index means that the SS index becomes "a little large" or "large", and "T1, T2" in Time means that the departure time becomes "early morning" or "morning". In this paper, we use

Table 2
Level set for the classification using the simplified fuzzy reasoning method

State labels of seat availability	Classification
"×"	$0.975 \leq y$
"Δ"	$0.75 \leq y < 0.975$
"○"	$0.75 < y$

Table 3
Real values of consequent part for seat availability labels "×" randomly set by uniform distribution [0.975, 1.000]

Rule	SS	Time	7 day	6 day	5 day	4 day	Value
Rule-1.1	SS4	T1	Δ	Δ	×	×	0.986
Rule-1.2	SS5	T1	Δ	Δ	×	×	0.995
Rule-1.3	SS4	T2	Δ	Δ	×	×	0.981
Rule-1.4	SS5	T2	Δ	Δ	×	×	0.989
Rule-2.1	SS4	T1	×	×	Δ	Δ	0.995
Rule-2.2	SS5	T1	×	×	Δ	Δ	0.999
Rule-2.3	SS4	T2	×	×	Δ	Δ	0.981
Rule-2.4	SS5	T2	×	×	Δ	Δ	1.000
Rule-3.1	SS4	T1	Δ	Δ	Δ	×	0.989
Rule-3.2	SS5	T1	Δ	Δ	Δ	×	0.999
Rule-3.3	SS4	T2	Δ	Δ	Δ	×	0.993
Rule-3.4	SS5	T2	Δ	Δ	Δ	×	0.998
Rule-6	SS5	T1	Δ	Δ	Δ	Δ	0.991
Rule-9	SS4	T1	×	×	○	○	0.997

Table 4
Real values of consequent part for seat availability labels "Δ" randomly set by uniform distribution [0.75, 0.975]

Rule	SS	Time	7 day	6 day	5 day	4 day	Value
Rule-4.1	SS4	T1	○	○	Δ	Δ	0.868
Rule-4.2	SS5	T1	○	○	Δ	Δ	0.961
Rule-4.3	SS4	T2	○	○	Δ	Δ	0.814
Rule-4.4	SS5	T2	○	○	Δ	Δ	0.839
Rule-5	SS5	T1	×	Δ	Δ	Δ	0.944
Rule-7.1	SS4	T1	○	○	○	Δ	0.806
Rule-7.2	SS5	T1	○	○	○	Δ	0.892
Rule-7.3	SS4	T2	○	○	○	Δ	0.776
Rule-7.4	SS5	T2	○	○	○	Δ	0.945
Rule-8	SS5	T2	○	Δ	Δ	Δ	0.962
Rule-10	SS4	T1	Δ	×	○	○	0.765

a simplified fuzzy reasoning method, and hence, we must set real values for the consequent part. As the classification of "×", "Δ", and "○" for the consequent value y using the simplified fuzzy reasoning method, we introduce the level set as shown in Table 2.

As another analytical approach to represent the advantage of the proposed approach, we compare the fuzzy reasoning method with a linear regression model using the first method of quantification (Quantification I) proposed by Hayashi in 1950 s (in detail, see [16]). The Quantification I is a method to predict the quantitative external criterion or its variable on the basis of the information concerning the qualitative attributes of each subject and to analyze the influence of each attribute to the criterion variable [16]. Therefore, the Quantification I is one of the most standard statistical analysis approaches for the categorical data such as Table 1. In general, consequent values of the simplified fuzzy reasoning method and external criterion scores of external criterion in the Quantification I are real, and we assume that consequent values of seat availability labels "× " and "Δ" are randomly set as Tables 3 and 4 from the random simulation based on the uniform distribution [0.975, 1.000] and [0.75, 0.975] based on the classification of Table 2, respectively.

Table 5
Some data of occupied rate in bullet trains on 17 July in 2011

Train No.	Departure time	7 day	6 day	5 day	4 day
7	6:50	×	×	○	Δ
15	8:10	Δ	×	Δ	Δ
215	9:00	○	Δ	○	Δ
225	11:00	○	○	Δ	Δ

From rules, classification of seat availability labels and real values of consequent part shown in Tables 1–4, we consider the following two opposite cases.

4.1. *Case of the big event held in Kyoto*

We first analyze the dataset of vacant rates in bullet trains on 17 July in 2011. On 17 July in 2011, Gion Festival was held in Kyoto. This festival is one of the biggest and most famous festivals in Japan. The value of SS index and scale is set as 0.88 in this paper, and the sample data of vacant rates in bullet trains on 17 July in 2011 are given as follows:

The simplified fuzzy reasoning method needs to set real values for input data of seat availability labels "×", "Δ", and "○" in Table 5. In this paper, we introduce the following random simulation considering the membership functions shown in Fig. 4 and the level set of classification in Table 2.

(Random simulation)

Step 1: All real values of "×" are set by a uniform distribution [0.975, 1.0].
Step 2: All real values of "Δ" are randomly set by a uniform distribution [0.75, 0.975].
Step 3: With respect to each real value of "○", in general, the occupied rate to state label of seat availability "○" is closely related to that

Table 6
Dataset of each seat availability label, consequent values of the proposed approach and external criterion scores of the Quantification I (QI)

Train no	Departure time	7 day	6 day	5 day	4 day	Proposed approach	QI	Actual data
7	6:50	0.990	0.990	0.740	0.821	0.989(×)	1.008(×)	×
15	8:10	0.775	0.976	0.918	0.941	0.990(×)	0.931(Δ)	×
215	9:00	0.729	0.773	0.741	0.780	0.920(Δ)	0.946(Δ)	Δ
225	11:00	0.748	0.718	0.761	0.953	0.878(Δ)	0.883(Δ)	Δ

of immediate label "Δ". For instance, in the case of Fig. 4, we consider that the occupied rate to label "○" is high and close to 0.75 where the satisfaction value of membership function to label "○" is the same as that of membership function to label "Δ". Then, the next day's occupied rate is probably higher than 0.75, that is, the next day's unoccupied state label probably becomes "Δ". Therefore, in this numerical example, each real value of "○" is randomly set by a uniform distribution [0.70, 0.75].

From the random simulation, a dataset of real values for seat availability labels "×", "Δ", and "○" is set as Table 6. We analyze this dataset using the simplified fuzzy reasoning method and the Quantification I, and obtain consequent values of the proposed approach and external criterion scores of the Quantification I (QI) shown in Table 6.

We obtain each consequent value of occupied rates using the simplified fuzzy reasoning method in Section 2 based on membership functions in Section 3, and set an appropriate label "×", "Δ", or "○" according to the level set for the classification shown in Table 2. We also obtain each external criterion score using the Quantification I based on the linear regression in Hayashi's study (Tanaka, 1979). From these calculation results with actual data on 17 July in 2011, the external criterion score of Train Number 215 is different from the actual data. On the other hand, all consequent values from the simplified fuzzy reasoning method have exactly the same properties with the actual data. This means that the simplified fuzzy reasoning method can be adequately available to analyze the Web data for the tourism. Particularly, the tourism data includes some ambiguous, linguistic, and experts' information. The fuzzy logic transforms linguistic data into numerical data represented as membership functions. Furthermore, by introducing experts' information and knowledge into the membership functions, and by using the fuzzy reasoning method with these membership functions, we perform a useful and flexible data analysis not describing from the only numerical data directly. Since we numerically deal with fuzzy reasoning methods to such various Web data, these Web data and consequences may be much available for the tourism.

Table 7
Sample data of occupied rates in bullet trains

SS	Departure time	7 day	6 day	5 day	4 day
SS4	10:00	Δ	○	Δ	Δ

Furthermore, from the simplified fuzzy reasoning method, the tourism planner may perform some favorable plans to promote more tourists. For instance, the following data of occupied rates in bullet train is received.

Using the simplified fuzzy reasoning method with this data and rule set shown in Table 1, we find that the appropriate label is set as "Δ" from the consequent value obtained by the simplified fuzzy reasoning method to input data in Table 7. However, if we enhance the SS index from SS4 to SS5 by performing some campaigns such as the fare discounting of trains or buses, the discount of admission fee and the special admission of important cultural properties, the appropriate seat availability label will be often "×" from the re-calculation result of consequent value, because the rule-6 in Table 1 exceedingly operates. Therefore, this approach will be useful to develop the other effective tourism plans.

4.2. *Case of specific data*

Second, we consider a specific data on 14 November in 2011 as shown in Table 8.

This date is Monday, and hence, it is not weekend. Furthermore, there is no big festival held in Kyoto. Therefore, the value of SS index becomes "small" (SS1) or "a little small" (SS2), i.e., this date is included in the off-season. Applying this data to the simplified fuzzy reasoning method, the consequent value is much less than 0.75, because almost all data of occupied rate in the off-season becomes "○ " from 7 to 4 days before.

Table 8
One data of occupied rates in bullet trains on 14 November in 2011

Train no	Departure time	7 day	6 day	5 day	4 day
17	8:30	○	○	○	○

However, the actual data of occupied rate one day before was "×", i.e., all seats in this train are occupied. Consequently, this data is very specific, and hence, we may discover some useful knowledge for the tourism by considering why all seats of this train in the off-season were fully occupied. As an observation of this analysis, the season for autumn leaves starts at the middle or end of November in Kyoto. Since Kyoto is one of famous and beautiful cities with respect to autumn leaves, many tourists may visit Kyoto according to information of autumn leaves in the last one or two days.

5. Conclusion and future works

In this paper, we analyzed tourism data, particularly bullet train in Japan, derived from the Web using the simplified fuzzy reasoning method which is one of the useful approaches in Data mining. Using the fuzzy reason method, we could deal with the ambiguous data such as gradual data, linguistic data, and expert's subjectivity, and hence, this approach will be used for the other data derived from the Web. For the remarkable development of ICTs, many people easily get and deal with various types of Web data. In many cases using Web data, the proposed approach in this paper will be the basis and important step for Web intelligence and Web data mining.

Of course, it obviously becomes important to deal with Web data in the tourism actively. Particularly, in this paper, we focused on the only bullet train data in the Web, since the access to Kyoto was mostly the bullet train. On the other hand, with respect to the other sightseeing cities, tourists visit them by various means of transportation such as not only the bullet train but also airplanes, buses, and local trains. For instance, Hiroshima is one of famous cities with some World Heritage cites in the world, and tourists in Japan visit Hiroshima by airplanes, buses, and trains including bullet trains. However, it is difficult to estimate the total number of tourist visiting Hiroshima exactly from the only bullet train data. Furthermore, it is also hard to get seat availability data of buses and local trains in the Web, and hence, it is also hard to divide tourist into some groups by means of transportation. In addition, in this paper, we assume that the value of season-scale index is initially given, but the SS index includes various parameters derived from historical data and Web data. Therefore, it may be difficult to analyze these data and to obtain the value of SS index easily. In the near future, using or extending the other approaches in Data mining as well as the fuzzy reasoning method, we will develop analysis techniques to various Web data for the tourism and find the specific features.

References

[1] G.I. Alptekin and G. Büyüközkan, An integrated case-based reasoning and MCDM system for Web based tourism destination planning, *Expert Systems with Applications* **38** (2011), 2125–2132.

[2] D. Buhalis, eTourism: Information technology for strategic tourism management, London: Pearson, 2003.

[3] D. Buhalis and R. Law, Progress in information technology and tourism next term management: 20 years on and 10 years after the Internet — The state of eTourism research, *Tourism Management* **29** (2008), 609–623.

[4] R.A. Carrasco and P. Villar, A new model for linguistic summarization of heterogeneous data: An application to tourism web data sources, *Soft Computing* **16** (2012), 135–151.

[5] C. Chang and L. Caneday, Web-based GIS in tourism information search: Perceptions, tasks, and trip attributes, *Tourism Management* **32** (2011), 1435–1437.

[6] R. Jakkilinki, M. Georgievski and N. Sharda, Connecting destinations with an ontology-based e-tourism planner, In: M. Sigala, L. Mich, & J. Murphy (Eds.), Information and communication technologies in tourism, (pp. 21–32), Springer, 2007.

[7] P.A. Johnson, R.E. Sieber, N. Magnien and J. Ariwi, Automated web harvesting to collect and analyse user-generated content for tourism, *Current Issues in Tourism* **15**(3) (2012), 293–299.

[8] JR CYBER STATION (2012). http://www.jr.cyberstation.ne.jp/ (in Japanese).

[9] J.R.P. Hanna and R.J. Millar, Promoting tourism on the Internet, *Tourism Management* **18**(7) (1997), 469–470.

[10] V.C.S. Heung, Internet usage by international travellers: Reasons and barriers, *International Journal of Contemporary Hospitality Management* **15**(7) (2003), 370–378.

[11] D.J. Kim, W.G. Kim and J.S. Han, A perceptual mapping of online travel agencies and preference attributes, *Tourism Management* **28** (2007), 591–603.

[12] D. Li and X. Dai, (2011). Research of web text fuzzy classification algorithm based on fuzzy comprehensive weighted evaluation reasoning. *Proceedings of SPIE - The International Society for Optical Engineering* 8285, Article number 82855Q.

[13] E.H. Mamdani, Application of fuzzy algorithms for control of simple dynamic plant, *Proceedings of IEE* **121**(12) (1974), 1585–1588.

[14] T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transaction on System, Man, & Cybernetics*, SMC-15(1), (1985), 116–132.

[15] W.K. Tan and T.H. Chen, The usage of online tourist information sources in tourist information search: An exploratory study, *Service Industries Journal* **32**(3) (2012), 451–476.

[16] Y. Tanaka, Review of the methods of quantification, *Environmental Health Perspectives* **32** (1979), 113–123.

[17] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning, *Part I: Information Sciences, 8, pp. 199-249. Part II: Information Sciences, 8, pp. 301-357, Part III: Information Sciences* **9** (1975), 43–80.

[18] A. Zimmermann, N. Lopes, A. Polleres and U. Straccia, A general framework for representing, reasoning and querying with annotated Semantic Web data, *Journal of Web Semantics* **11** (2012), 72–95.