Finding Comparative Facts and Aspects for Judging the Credibility of Uncertain Facts

Yusuke Yamamoto* and Katsumi Tanaka

Graduate School of Informatics, Kyoto University,
Yoshida-Honmachi, Sakyo, Kyoto, Japan
{yamamoto, tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract. Users often encounter unreliable information on the Web, but there is no system to check the credibility easily and efficiently. In this paper, we propose a system to search useful information for checking the credibility of uncertain facts. The objective of our system is to help users to efficiently judge the credibility by comparing other facts related to the input uncertain fact without checking a lot of Web pages for comparison. For this purpose, the system collects comparative facts for the input fact and important aspect for comparing them from the Web and estimates the validity of each fact.

Keywords: Support of credibility judgment, Mining of comparative facts and aspects for comparison, Web mining, Credibility.

1 Introduction

Nowadays, there is a great deal of information on the Web and people can easily obtain it. As the Web increases in popularity, however, problems with the credibility of Web information have emerged. Most of Web information is anonymous and not authorized unlike conventional mass media. This is why people often encounter uncertain facts. For example, many Japanese Web pages report the uncertain fact that "soybeans are effective for weight loss" without any evidence. As another example, the question "Which country makes the most famous beer?" has a variety of answers on the Web, such as "Mexico is famous for beer" and "Germany is famous for beer", on Yahoo! Answers[1], a well-known question answering (QA) site. In a case like this, users can have difficulty judging which answers are correct or incorrect because some answers have little evidence and other comparative answers may be given on other Web sites. If users are unaware of such information credibility, they can easily be misled. Therefore, a system is needed for analyzing or helping users judge the credibility of information on the Web.

For this purpose, we have developed a system called *Honto Search* to help users judge the credibility of uncertain facts ("Honto?" means "Is it true?" in Japanese) [1, 2]. The aim of Honto Search is to enable a user to judge the credibility of an uncertain

* He also works as a Research Fellow (DC2) of the Japan Society for the Promotion of Science.

[1] The question "Which Country Makes The Best Beer?" in Yahoo! Answers: http://answers.yahoo.com/question/index;_ylt=AgY58h8jhIJVGArKDTlYu5966xR.;_ylv=3?qid=20080329195454AAGTWWC&show=3#yan-answers

G. Vossen, D.D.E. Long, and J.X. Yu (Eds.): WISE 2009, LNCS 5802, pp. 291–305, 2009.

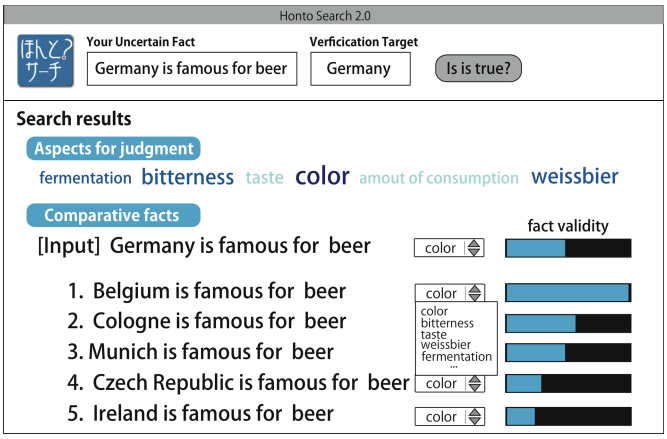

Fig. 1. Example of a Honto Search 2.0 result

fact by comparing related facts and checking the information in each fact. After the user inputs an uncertain fact and a verification target within the fact into Honto Search, the system collects alternative or countervailing facts from the Web and shows the temporal changes of each fact's popularity and sentiment. The system is imperfect, however, for credibility judgment. First, it often fails to collect enough comparative facts for comparing the input uncertain fact. Secondly even if the system collects enough comparative facts, it unexpectedly also extracts unrelated facts as well. Thirdly, temporal analysis and sentiment analysis are not always useful. The really desired function is to notice aspects necessary for checking the credibility of uncertain facts and to enable the user to check the information about a fact in terms of the aspects.

In this paper, we propose an improved, real-time system for more substantial credibility judgment of uncertain facts. This system extends Honto Search, and we call it *Honto Search 2.0*. A target fact in Honto Search 2.0 is an uncertain fact whose credibility can be judged by comparison with other related facts. To judge the credibility of such facts, the system has two functions:

- **Comparative Fact Finding:** The system collects only comparative facts for judging the credibility of the uncertain fact and estimates the validity of each fact by using a Web search engine.
- **Aspect Extraction for Checking the Fact Credibility:** The system also extracts aspects necessary to check the credibility of facts in detail from the Web.

Fig.1 illustrates one example of *Honto Search 2.0* where the user has checked whether Germany is more famous for beer than any other country. Given the two types of inputs, the system returns several comparative facts, such as "Belgium is famous for beer", "Cologne is famous for beer", and "Munich is famous for beer". At the same time, the system estimates the validity each fact on the Web. In this example, the system estimates the fact "Belgium is famous for beer" as the most valid comparative fact in the comparative facts including the fact "Germany is famous for beer". In addition, the system shows several aspects for checking credibility, such as "color",

"bitterness", and "weissbier". By doing a Web search with these aspects and each fact, the user can check Web pages describing them. In this way, the system enables users to judge the credibility of uncertain facts.

The remainder of the paper is organized as follows. In the next section, we discuss related work. Section 3 provides a system overview of Honto Search 2.0, while Section 4 discusses the details of our method to extract valid comparative facts from the Web. In Section 5, we explain the proposed method to collect aspects for judging the credibility of facts. In Section 6, we report experimental results obtained using our system. The last section concludes the paper and outlines our future research directions.

2 Related Work

There are previous studies focused on the credibility of information. Fogg et al. studied various metrics to evaluate Web site credibility from the viewpoint of users [3] and analyzed the effectiveness of all these metrics by doing a large-scale experiment [4]. Nakamura et al. surveyed around 1000 participants on their attitudes toward the credibility of Web search engines [5]. Few such projects, however, have proposed concrete methods of evaluation or applications. We proposed a system, Honto Search, to support assessment of the credibility of uncertain facts [1, 2]. Kobayashi et al. proposed the method to judge whether a product name is a brand name or a glorified term [6].

There are various works on comparative search and browsing in the Web research field. Sun at el. proposed a system named *CWS* for searching Web page sets that are appropriate for comparing two specified topics [7]. Nadamoto at el. developed a system, called a comparative web browser (CWB), for comparing one Web page with another on the same topic [8, 9]. The CWB has two browser windows and concurrently displays two Web pages in a way that enables the page contents to be automatically synchronized. Nakamura et al. proposed the system named *SyncRerank* to enable users to simultaneously rerank two kinds of search results when users compare two search results [10]. These works are based on the assumption that users need to specify comparative topics or aspects for comparison. On the other hand, our system can automatically find the comparative topics and the aspects for comparison.

Many works have examined Web mining of comparative objects or information for such comparison. Kurashima et al. proposed a method to rank the entities in a specific category by using comparative sentences like "The quality of X is better than Y" [11]. Liu et al. proposed a system called *Opinion Observer* [12]. Once several product names are input to the system, it extracts aspects for comparing products and aggregates the review information about each product from the viewpoint of each aspect. Zhai at el. proposed a generative probabilistic mixture model to extract aspects for comparing topics [13]. Ohshima et al. proposed a method to find coordinate terms of a given term by using a Web search engine in real time [14]. Most of these approaches, however, have focused only on entities like products, people, and so on. Moreover, they assume off-line data processing. On the other hand, our proposed method extracts comparative facts in the form of phrases from the Web and does not depend on any specific domains. Moreover, our method can extract comparative facts and aspects for comparison in real time.

3 System Overview

In this section, we describe an overview of our system, *Honto Search 2.0*. The purpose of Honto Search 2.0 is to enable users to compare an uncertain fact with related facts for judging the credibility of the uncertain fact in real time. Fig. 2 shows a flowchart of the operation of Honto Search 2.0.

The system requires two kinds of inputs. One is an uncertain fact whose credibility a user wants to check. This fact is input to the system as a phrase. The other input is a specific part of the uncertain fact, which we call a *verification target*. For example, if a user wants to know whether Germany is famous for beer as compared with other regions, the user inputs the phrase "Germany is famous for beer" as the uncertain fact and "Germany" as the verification target. After that, it collects comparative fact candidates from the Web in order to check the credibility of the input fact. The comparative fact candidates are extracted by applying syntactic pattern mining to Web search indexes. In the next step, from among the collected fact candidates, the system selects only meaningful facts for comparison with the input fact. Examples of such comparative facts are "Belgium is famous for beer", "The Czech Republic is famous for beer", and so on. At the same time, the validity of each fact is estimated by using statistical information on the Web. The collected comparative facts and the input fact are thus ranked according to validity scores.

To enable the user to check the credibility of facts in more detail, the system also gives the user viewpoints for comparison between facts. We call these viewpoints *aspects*. Possible aspects corresponding to the example fact above, i.e., "Germany is famous for beer", are *taste*, *amount of consumption*, *brand*, *manufacturing method*, and so on. The system also provides a function to link to Web pages that describe each fact and specific aspect. As a result, the user can easily check the details of each fact in terms of each aspect.

Our system collects all necessary data for analysis by using a conventional Web search engine and analyzing the search engine's indexes in real time. That is, the system contains no data of its own. Therefore, it can flexibly deal with a wide variety of input facts.

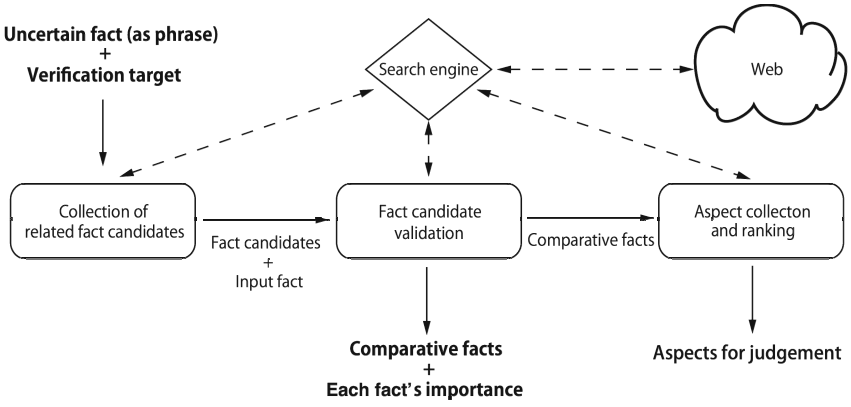

Fig. 2. Flowchart of the operation of Honto Search 2.0

4 Finding Comparative Facts

The objective of our system is to enables users to judge the credibility of an uncertain fact by comparing related facts. In this section, we propose a method to achieve this goal by finding comparative facts and estimating their validity.

4.1 Model and Task Definition

When a user enters an uncertain fact, as a phrase, and a verification target, within the fact, into our system, it collects several comparative facts from the Web. In the case of the uncertain fact given above, "Germany is famous for beer", with the term "Germany" as the verification target, the previous version of our system extracted sentences matching the pattern "(*) is famous for beer" from Web pages. Here, we regard the fact "Germany is famous for beer" as a *proposition* generated from a certain *predicate*. Moreover, we regard the verification target "Germany" as an *object* for the *variable* of the predicate. If we use these ideas, we can consider that the operation of inputting an uncertain fact and specifying its verification essentially consists of identifying the predicate. For example, if we define the predicate *FamousBeerPlace(x)* to mean that *x* is famous for beer, then our system's first task is to find a set of candidate objects $X = \{x \mid FamousBeerPlace(x) \text{ could be true}\}$.

The second task of our system is to estimate the validity of each of the collected fact candidates. In the case of "Germany is famous for beer", we need to estimate the importance or validity of comparative facts such as "Belgium is famous for beer", "The Czech Republic is famous for beer", and so on. We suggest that this operation means evaluating the strength of the relation between the predicate *FamousBeerPlace(x)* and various objects for the variable *x*, like *Germany, Belgium, Czech Republic*.

Hence, in this section, we formally define our system's tasks as the following:

- **Task1: Comparative fact Candidate Extraction:** First, given an uncertain fact *u* and a verification target *x*, we find a predicate *p* such that *p*(x) is equivalent to *u*. Next, to collect as many comparative facts as possible, we extract other possible predicates $P' = \{ p'_1, p'_2, \ldots, p'_n \}$ from the Web, where each *p'*(x) is semantically similar to *p(x)*. After that, we extract a set of *comparative objects* $X = \{ x \mid$ *the predicate p(x) or p'(x) in P' can be true* $\}$ from the Web. We define each p(x) for x in X as a *comparative fact candidate*.
- **Task2: Precise Estimation of Fact Validity:** To estimate the validity of each fact, we evaluate the strength of the relation between the predicate *p* and each comparative object *x*.

4.2 Extraction of Syntactic Patterns to Find Comparative Facts

In the previous version of our system, users often fail to get enough comparative facts to check the credibility of an uncertain fact, because the system used just one predicate (syntactic pattern) for collecting comparative facts. If the predicate does not appear on the Web, we cannot extract any comparative facts from the Web. To increase the variety of comparative facts, we need to collect several predicates representing the semantics of the original predicate.

Our basic strategy for achieving this goal is to extract several important keywords from the uncertain fact and then collect patterns, which include all of the keywords, as new predicates. The actual procedure is defined as follows:

1. The system does morphological analysis of the input fact. It then extracts only nouns, adjectives, and verbs (other than the verb "to be"). We denote the extracted terms as $T = \{t_1, t_2, \ldots, t_n\}$. In the case of "Germany is famous for beer", the system extracts *{Germany, famous, beer}* as *T*.
2. The system issues the query "t_1 *AND* t_2 *AND* … *AND* t_n" to a Web search engine and gets the top-N Web search results.
3. The system first collects sentences containing all of the keywords in *T* from each result's snippet, and then extracts substrings with arbitrary keywords at both ends. Examples of such substrings are *"famous beer in Germany", "Germany developed world famous beer", "Germany famous for beer"*.
4. The system replaces the verification target in each substring with <*variable*>. For the above examples, this gives *"famous beer in <variable>", "<variable> developed world famous beer",* and *"<variable> famous for beer"*.
5. The system ranks the resulting patterns by considering their lengths and their frequencies in the search results collected in step 2. We assume that the more frequent and shorter a pattern is, the appropriately it works as an alternative predicate. The score for pattern ranking is defined as the following:

$$score(pattern) = freq(pattern) * \log \frac{C}{length(pattern)} \quad (1)$$

   Here, *freq(pattern)* is the frequency of *pattern* appearing in the search results, and *length(pattern)* is the number of terms in *pattern*, and *C* is a constant number.
6. Finally, the system selects the top N patterns from the ranked patterns as the set of other possible predicates, *P'*. It also replaces the verification target in the input fact with <*variable*> and denotes the resulting predicate as *p*.

4.3 Finding Comparative Fact Candidates

Next, we collect comparative fact candidates for checking the credibility of the input fact. In this step, we collect Web pages about each predicate by using a Web search engine and then extracting comparative fact candidates from the search results. We then select the final comparative fact candidates by using the assumption that comparative facts can concurrently appear in documents including the input fact. The procedure is as follows:

1. The system converts each predicate to a query for extracting comparative fact candidates. If the predicate has the form *"<pattern1> <variable> <pattern2>"*, like *"beer cup features many of <variable>'s famous"*, then the system generates the query "*<pattern1> AND <pattern2>*". Otherwise, the system eliminates <*variable*> from the predicate and uses the rest of the predicate as a query. For example, the predicate *"<variable> is famous for beer"* is converted to the query *"is famous for beer"*.

2. The system issues each query to a Web search engines and gets the to-N search results. The system extracts the parts of result snippets that matches <*variable*> of the predicate *p* or one of the predicates in *P'* from the snippets of the results.
3. The system checks whether the part of speech (POS) of each extracted term is the same as the POS of the verification target. If they are the same, the term can be an object candidate *c* for the variable of the predicate *p*.
4. Finally, the system collects the top N Web pages including the input fact by using a search engine. If an object candidate *c* is included in some of the collected pages, the system regards *p(c)* as a final comparative fact candidate.

4.4 Verification of Candidates as Comparative Fact

To estimate the validity of a fact *p(c)*, our previous system issued each fact as a phrase to a Web search engine and used the page counts as a query to the search engine. Phrase search is so strict, however, that we could not often get a page count for a fact. Moreover, comparing the page count of an uncertain fact with the page counts of comparative facts was too rough for estimating the validity of the uncertain fact. We instead need to estimate fact validity more precisely.

We can assume that the validity of a fact *p(c)* is given by the strength of the relation between the object candidate *c* and the predicate *p*. Applying this assumption, we use the WebPMI [15] to estimate the fact validity. The WebPMI is one kind of point-wise mutual information and is estimated by calculating the co-occurrence degree between two terms on the Web. We define the fact validity *FactValidity(c, p)* of the fact *p(c)* by using the WebPMI as follows:

$$FactValidity(c,p) = \log(\frac{H(c \wedge p)}{H(c)H(p)}) \tag{2}$$

Here, *H(q)* means the page count for query *q* in a Web search engine.

By using this definition, the system estimates the fact validity for all comparative fact candidates collected by the procedure given in Section 4.3. Only if the fact validity of a comparative fact candidate is greater than zero, we regard the candidate as a comparative fact. Finally, the system shows the comparative facts and their scores to the user.

5 Extraction of Aspects for Judging Credibility of Facts

In checking the credibility of an uncertain fact, it is insufficient to compare related facts from the viewpoint of fact validity. The credibility of a fact often depends on specific aspects in comparing facts. In addition, the most important issue in judging credibility is to understand why a fact is credible or not. When a user checks the credibility of an uncertain fact, if the user knows specific aspects necessary to check the credibility, there is no problem. At that time, the user can search for clues to check the credibility on his or her own. Users do not always know, however, which aspects are important in checking the credibility of uncertain facts. In this section, we describe an approach to extract aspects for judging the credibility of an uncertain fact from the Web.

5.1 Collection of Aspect Candidates

When an uncertain fact is described in terms of an aspect in a document, certain assessment words are often used around the aspect. For example, in the case of the fact "Germany is famous for beer", phrases such as "amount of consumption is high" and "good taste" can be found in documents about this fact, and we can thus select "amount of consumption" and "taste" as aspects. For the purpose of finding aspects, one possible method is to collect pages including the uncertain fact and extract frequently neighboring assessment terms such as adjectives. Although this method cannot collect enough aspects and not all of them are useful, the collected aspects can be used as candidates. Therefore, we apply this method to first collect aspect candidates.

In this paper, we have focused on uncertain facts whose credibility is checked by comparison with other, related facts. In this case, the necessary aspects depend on the facts with which users want to compare the uncertain fact. For example, one user might want to check the credibility of the fact "Germany is famous for beer" by comparison with the fact "Belgium is famous for beer", while another user might want to compare with the fact "Germany is famous for sausage". That is, the extracted aspects depend on the predicate for the uncertain fact. Therefore, we use the predicate to collect aspect candidates. The procedure for aspect extraction is described as follows:

1. The system converts the predicate for the uncertain fact to a phrase query, as described in Section 4.3. The system then issues the phrase query to a Web search engine and collects the top-N search results.
2. The system extracts phrases with the following syntactic patterns from each search result snippet: *"<Adjective> <Noun phrase>" or "<Noun phrase> <the verb "to be"> <Adjective>"*. In this step, we use a morphological analyzer to identify each term's POS.
3. The system extracts <Noun phrases> as aspect candidates from the collected phrases. At this time, aspect candidates that appear in a stopword list are eliminated from the list of aspect candidates.

5.2 Aspect Ranking by Using Comparative Facts

In the case of using only adjectives as clues for extracting aspect candidates, inappropriate terms can appear as aspects. Therefore, we need to estimate the appropriateness of aspect candidates and distinguish worthless candidates.

The simplest approach is to calculate the probability for each aspect candidate, $Pr(a|p)$, that the aspect candidate a appears given a predicate p. Unexpectedly, however, inappropriate terms can have a high probability of appearing as appropriate aspects, because candidates are collected by using a simple syntactic rule. To solve this problem, we introduce another hypothesis. Because expected aspects are useful for comparing an uncertain fact with related facts, we suppose that the more useful an aspect is, the more facts appear with it in Web pages, as illustrated in Fig. 3. By using this assumption, we determine the probability $Pr(a|p)$. Let C be a set of comparative terms that are objects for the variable of the predicate p and are collected by the method proposed in Section 4. The probability $Pr(a|p)$ can be defined as follows:

$$\Pr(a \mid p) = \frac{\Pr(a, p)}{\Pr(p)}$$

$$= \frac{\sum_{c \in C} \Pr(a, c, p)}{\Pr(p)}$$

$$= \frac{\sum_{c \in C} \Pr(a \mid c, p) \Pr(c, p)}{\Pr(p)}$$

$$= \sum_{c \in C} \Pr(a \mid c, p) \Pr(c \mid p) \quad (3)$$

Here, *Pr(a|c,p)* means the probability that aspect *a* appears in Web pages containing the comparative fact *p(c)*, while *Pr(c|p)* means the probability of occurrence of the comparative term c if predicate *p* is specified. Hence, *Pr(c|p)* is the co-occurrence degree between *c* and *p*, which we estimate as the following:

$$\Pr(c \mid p) = \frac{FactValidity(c, p)}{\sum_{c' \in C} FactValidity(c', p)}, \quad (4)$$

where the function *FactValidity* was defined in Section 4. To enhance the validity in calculating the probability *Pr(a|c,p)*, we introduce a semantic factor into the calculation, as with the syntactic patterns shown in the previous subsection. We redefine *Pr(a|c,p)* as the probability of occurrence of aspect term *a* in Web pages containing fact *p(c)*, where *a* appears in the form "<Adjective> <*a*>" or "<*a*> <the verb "to be"> <Adjective>". In fact, it is difficult to collect all Web pages containing the fact *p(c)* and then calculate *Pr(a|c,p)*. Therefore, we calculate *Pr(a|c,p)* approximately by using the top-N Web search results collected with the query "*p AND c*".

By using these definitions, we can finally calculate the probability *Pr(a|p)* for all aspect candidates collected as described in Section 5.1. The system orders all aspect candidates by score and reports the top N terms to the user as important aspects for judging the credibility of an uncertain fact.

6 Evaluation

We conducted two kinds of evaluations for our algorithm. The first was for evaluating the performance of the algorithm in collecting comparative facts. The second was for evaluating how many meaningful aspects were collected by the algorithm for comparison. For these experiments, we prepared 20 pairs of uncertain facts and verification targets as a test set for evaluation. These pairs are listed in Table 1. The pairs in the test set were categorized into 6 groups. As described above, the credibility of the uncertain facts is supposed to be checked by comparison with other related facts, such as the examples of comparative facts given in the table. These experiments were conducted in Japanese, and then the results were translated into English. Constant number C was set to 30.

6.1 Validity and Variety of Collected Facts

To evaluate the performance of comparative fact collection, we used our algorithm with all pairs listed in Table 1. We set the system to use Google as the Web search engine and to collect the top 100 search results. In this experiment, we ordered the collected comparative facts by fact validity and evaluated the number of collected facts that is appropriate for comparison with the top 3, top 5, top 10, and all results. The adequacy of the facts was judged by checking whether they were reported in Wikipedia. In addition, we evaluated the processing time of our algorithm. The algorithm proposed in our previous work was used as a baseline, and the previous results were compared with the results of our new algorithm.

Table 1. Test set for the experiments

Fact type	Input fact	Verification target	Example of comparative facts
Local product of nation	Germany is famous for beer	Germany	Belgium, Denmark
	Germany is famous for beer	beer	sausage, wine, car
	Poland is famous for amber	Poland	Lithuania, Russia
	Poland is famous for amber	amber	escargot, rye
Compound's effectiveness	Anthocyanin helps eyestrain	anthocyanin	polyphenol
	Anthocyanin helps eyestrain	eyestrain	atherosclerosis, cyanocobalamin
Geographic characteristic of nation	England is a rainy country	England	Thailand, Malaysia
	Italy is an earthquake-prone country	Italy	Japan, Armenia, Turkey
	Japan has many volcanoes	Japan	Indonesia, Guatemala
	Drought often happens in India	India	Australia
Social characteristic of nation	Agriculture is important in Poland	Poland	China, Australia, France
	Moscow is an expensive city	Moscow	Tokyo, London
	London has a large population	London	Mumbai, Tokyo, New York
Leading country for sport	Japan is strong in judo	Japan	France, Netherlands
	Japan is strong in baseball	Japan	United States, Korea, Cuba
	Argentina is strong in football	Argentina	Brazil, England, Spain
	Japan is strong in volleyball	Japan	Cuba, Brazil
Common religion of nation	Buddhism is common in Japan	Japan	Thailand, Korea
	Islamism is common in Indonesia	Indonesia	Egypt, Turkey
	Hinduism is common in India	India	Nepal, Bangladesh

Table 2 summarizes the results of the two algorithms. The proportion of appropriate comparative facts obtained by the proposed method was higher than the proportion obtained by the previous method in all cases except the top-3 results. In the case of all results, the proportion obtained by the proposed method was about 5% higher than obtained by the previous method. As shown in Table 2, there was little difference between the proposed method and the previous one from the viewpoint of the validity of collected comparative facts. From these results, both methods can be expected to perform well in collecting valid comparative facts.

When focusing on the variety of collected comparative facts, however, we find a significant difference between the algorithms' performances. The number of valid comparative facts collected by the proposed method was 26.1% greater than that collected by the previous method, on average. The proposed method was slower than the previous method, because the system must access a Web search engine three times for one calculation of the WebPMI, but we think this extra time is necessary for better judgment of the credibility of uncertain facts.

For the detail study, we pick up some examples of comparative fact extraction. Table 3 illustrates the results. When we focused on the predicates used in fact extraction, we found a problem. In case an input uncertain fact does not often appear on the Web, the system fails to find alternative predicates for collecting comparative

facts and to collect comparative facts. One of the examples is the fact "Poland is famous for <amber>". In such a case, we need to transform the input fact to another expression before applying our proposed method. Actually, when we uses the query "<Amber> is a specialty product of Poland" by substitutes the term "famous" with "specialty product", we got vodka, mushroom, hot pepper, chess, etc. as comparative terms. On the other hand, in case the input uncertain fact often appears to some extent on the Web, the system succeeded in increasing the amount of valid comparative facts. The most typical case is the input "<Germany> is famous for beer". In the case of this input, the previous algorithm collected 5 valid comparative facts while the proposed method collected 9 valid comparative ones. "Ireland", "United States", "Thailand", "Korea" were collected by our proposed method as additional valid comparative facts.

Table 2. Results of using Honto Search 1.0 (previous method) vs. Honto Search 2.0 (proposed method) for comparative fact collection

Algorithm	Proportion of valid comparative facts in top-N results				Average of processing time(s)	Increase in number of valid comparative facts
	@1	@3	@5	All		
Proposed method	85.0%	70.4%	69.1%	70.9%	11.53	26.1%
Previous method	85.0%	80.0%	63.3%	65.6%	2.57	-

Table 3. Examples of comparative fact extraction. The term between brackets is the verification target in the uncertain fact. The number in parenthesis is the fact validity score. Underlined terms are valid terms as variable of the predicate for the input fact.

Input fact (verification target)	Result
<Germany> is famous for beer	**Belgium(2.97)**, Cologne(2.78), Munich(2.77), **Czech Republic(2.53)**, **Denmark(2.43)**, **Ireland(2.42)**, **United States(2.03)**, **Thailand(2.02)**, **Germany(1.97)**, **Korea(1.93)**, **England(1.91)**
Agriculture is important in <Poland>	**Macedonia(4.00)**, **Poland(3.39)**, **Rumania(3.30)**, Spain(2.54), **Thailand(2.15)**, Japan(1.15), **China(0.58)**
Drought often happens in <India>	**Australia(3.52)**, **India(3.18)**
<Japan> is strong in baseball event	**Cuba(3.20)**, **Korea(2.16)**, Asia(2.04), Osaka(1.41), Hokkaido(1.28), **Japan(0.96)**, Ehime(0.95), Wakayama(0.90), Taiwan(0.52)
<Japan> is strong in judo event	**Nederland(2.70)**, **France(1.83)**, **Japan(1.01)**
Poland is famous for <amber>	**amber(0.31)**

We also need to improve the validity calculation of extracted comparative facts. For example, when we input the fact "<Japan> is strong in baseball event" to the system, we got 9 comparative facts, but only 3 facts of them were valid. The invalid comparative facts are superclass or subclass objects of Japan such as Asia, Osaka prefecture, and Hokkaido prefecture. For solving this problem, it is effective to use the ontological information of the verification target as well as the POS information.

6.2 Aspect Extraction

Next, we evaluated the performance of aspect extraction for judging the credibility of uncertain facts. We applied our aspect extraction algorithm with all pairs listed in Table 1. As described in Section 6.1, we collected the top 100 search results in analyzing Web pages. In this experiment, we ordered the extracted aspects for each

uncertain fact by the probability *Pr(a|p)* in formula (3) and extracted the top-10 aspects as the final ones. Then, we evaluated how many aspects were meaningful for comparing the collected comparative facts with each input fact and judging its credibility. We manually judged whether the extracted aspects were meaningful.

Table 4. Three best cases of aspect extraction. The term between brackets is the verification target in the uncertain fact. The underlined terms are appropriate aspects.

Uncertain fact (verification target)					
<Germany> is famous for beer		Agriculture is important in <Poland>		Drought often happens in <India>	
amount of consumption	0.03907	**vegetable**	0.19545	snow	0.47420
color	0.03654	**fruit**	0.05869	**climate**	0.47420
weissbier	0.02200	**future**	0.05639	**wildfire**	0.17527
bitterness	0.01579	industry	0.04210	**weather**	0.08000
japan	0.01279	**production**	0.02950	**rain**	0.07052
ferment	0.01253	nature	0.02666	amount	0.05897
Pilsner Urquell	0.01183	**region**	0.01837	**influence**	0.04000
culture method	0.01142	**amount** of crop	0.01809	**region**	0.03897
restaurant	0.00960	**amount**	0.01809	green	0.03052
street	0.00897	land	0.01548	**problem**	0.03052

Table 5. Three worst cases of aspect extraction. The term between brackets is the verification target in the uncertain fact. The underlined terms are appropriate aspects.

Input fact (verification target)					
Poland is famous for <amber>		<Japan> is strong in judo		<Japan> is strong in baseball	
soup	0.02000	**player**	0.06868	**high school**	0.03087
-		news	0.03756	**team**	0.03049
-		feelings	0.03290	game	0.02993
-		gossip	0.02567	school	0.02541
-		rugby	0.02315	**impression**	0.02278
-		factor	0.02000	awareness	0.02179
-		think	0.01951	feelings	0.02132
-		fighting sport	0.01685	**reason**	0.01827
-		Tokyo	0.01190	United States	0.01379
-		opponent	0.01025	foundation	0.01336

In the case of applying our algorithm to all pairs in the test set, the average number of meaningful aspects was 4.75. We assume that our system would show 10 aspects for an uncertain fact and the user could then select specific ones to check detailed information about those aspects of the uncertain fact on the Web. Therefore, we consider the performance of our system in this experiment to be adequate.

To more intuitively understand the results, we studied some specific cases of aspect extraction. Table 4 summarizes the three best cases, in terms of obtaining many meaningful aspects, among the pairs of an uncertain fact and a verification target. When we input the fact “Germany is famous for beer” and specified “Germany” as the verification target, we obtained 7 appropriate aspects. In this case, we assume that a user would want to check the credibility of the fact by comparing Germany with other countries famous for beer, such as Belgium. Therefore, aspects like “amount of consumption”, “color”, and “bitterness” would be useful for judging the credibility of the fact. Most of the collected aspects for the other pairs were also regarded as useful. The common features in these uncertain facts are that we could obtain many comparative facts for any pair and that they were appropriate facts for comparison.

For example, of the 11 comparative facts collected for the uncertain fact "<Germany> is famous for beer", 9 were appropriate.

Table 5 summarizes the three worst cases of aspect extraction. We could not collect any meaningful aspects for the uncertain fact "Poland is famous for <amber>, with only one aspect collected at all. This was because there were few pages including the fact, and therefore, the system could not extract syntactic patterns for collecting comparative facts. As for the facts "<Japan> is strong in judo" and "<Japan> is strong in baseball", the system could obtain only a few meaningful aspects, although it did extract a several number of aspects for each fact. In the case of the former fact, the number of extracted comparative facts was very small. Therefore, we think that the quality of the extracted aspects was low. In the case of the latter fact, we actually obtained many comparative facts, but the system could not collect many meaningful aspects. This was because most of the comparative facts were not valid. We expected that, ideally, the comparative facts of the latter fact should represent which countries are other leading baseball countries. However, most of the collected comparative facts represented that specific Japanese prefectures are strong in baseball event. This is why we need a certain number of valid comparative facts to collect meaningful aspects.

7 Conclusion

In this paper, we developed the system to help users to judge the credibility of the uncertain facts by showing comparative facts and important aspects for judgment of the credibility. Evaluations showed that our proposed method could find comparative facts and important aspects for comparison from the Web if input facts are relatively often seen on the Web.

However the technique of our proposed method is immature. First, if descriptions about uncertain facts do not appear to an extent on the Web, the system fails to collect comparative facts and aspects for comparison. In such a case, we need to modify input facts to collect as many Web pages for analysis as possible. Moreover, we need to consider how to extract comparative facts with negative polarity such as "Soy bean is totally not effective for weight loss" against "Soy bean is effective for weight loss". As a more substantial problem, ideally, users expect the system to directly check whether the input uncertain fact is true or not and to estimate which comparative fact is the most credible. Our system estimates each comparative fact using WebPMI, but this estimation is still based on the statistical relation between the variable and the predicate. We have the assumption that the credibility of a fact involves various factors and should be estimated by considering all of them. We proposed the method to find aspects for checking the credibility of uncertain facts. We have a plan to estimate the credibility of uncertain facts by using these aspects as parameters.

There is a lot of unreliable information on the Web. Especially, Web 2.0 contents are critical. Lately, Web 2.0 contents such as Yahoo! Answer and Wikipedia are getting more attention as collective knowledge. However, most of them are not authorized. Our proposed method can be useful from the viewpoint of estimating the credibility of Web 2.0 contents by aggregating Web 1.0 contents. In the future, we plan to estimate the credibility of Web 2.0 contents and support users to judge the credibility by using our proposed method.

Acknowledgment

This work was supported in part by a MEXT Global COE Program "Informatics Education and Research Center for Knowledge-Circulating Society" (Project Leader: Katsumi Tanaka), a MEXT Grant-in-Aid for Scientific Research on Priority Areas "Cyber Infrastructure for the Information-explosion Era", Planning Research "Contents Fusion and Seamless Search for Information Explosion" (Project Leader: Katsumi Tanaka, A01-00-02, Grant#: 18049041), the National Institute of Information and Communications Technology, and a MEXT Grant-in-Aid for JSPS Fellows (Project Leader: Yusuke Yamamoto, Grant#: 211243).

References

1. Yamamoto, Y., Tezuka, T., Jatowt, A., Tanaka, K.: Honto? Search: Estimating Trustworthiness of Web Information by Search Results Aggregation and Temporal Analysis. In: Dong, G., Lin, X., Wang, W., Yang, Y., Yu, J.X. (eds.) APWeb/WAIM 2007. LNCS, vol. 4505, pp. 253–264. Springer, Heidelberg (2007)
2. Yamamoto, Y., Tezuka, T., Jatowt, A., Tanaka, K.: Supporting Judgment of Fact Trustworthiness Considering Temporal and Sentimental Aspects. In: Bailey, J., Maier, D., Schewe, K.-D., Thalheim, B., Wang, X.S. (eds.) WISE 2008. LNCS, vol. 5175, pp. 206–220. Springer, Heidelberg (2008)
3. Fogg, B.J., Tseng, H.: The elements of computer credibility. In: Proc. of CHI 1999, pp. 80–87 (1999)
4. Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M.: What makes web sites credible? A report on a large quantitative study. In: Proc. of CHI 2001, pp. 61–68 (2001)
5. Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S., Tanaka, K.: Trustworthiness analysis of web search results. In: Kovács, L., Fuhr, N., Meghini, C. (eds.) ECDL 2007. LNCS, vol. 4675, pp. 38–49. Springer, Heidelberg (2007)
6. Kobayashi, T., Ohshima, H., Oyama, S., Tanaka, K.: Evaluating Brand Value on the Web. In: Proc. of WICOW 2009, pp. 67–74 (2009)
7. Sun, J.T., Wang, X., Shen, D., Zeng, H.J., Chen, Z.: CWS: A Comparative Web Search System. In: Proc. of WWW 2006, pp. 467–476 (2006)
8. Nadamoto, A., Tanaka, K.: A Comparative Web Browser (CWB) for Browsing and Comparing Web Pages. In: Proc. of WWW 2002, pp. 727–735 (2002)
9. Nadamoto, A., Qiang, M., Tanaka, K.: Concurrent Browsing of Bilingual Web Sites by Content-Synchronization and Difference-Detection: In Proc. of WISE 2003, pp. 189—199 (2003)
10. Nakamura, S., Yamamoto, T., Tanaka, K.: SyncRerank: Reranking Multi Search Results Based on Vertical and Horizontal Propagation of User Intention. In: Bailey, J., Maier, D., Schewe, K.-D., Thalheim, B., Wang, X.S. (eds.) WISE 2008. LNCS, vol. 5175, pp. 120–135. Springer, Heidelberg (2008)
11. Kurashima, T., Bessho, K., Toda, H., Uchiyama, T., Kataoka, R.: Ranking entities using comparative relations. In: Bhowmick, S.S., Küng, J., Wagner, R. (eds.) DEXA 2008. LNCS, vol. 5181, pp. 124–133. Springer, Heidelberg (2008)
12. Liu, B., Hu, M., Cheng, J.: Opinion Observer: Analyzing and Comparing Opinions on the Web. In: Proc. of WWW 2005, pp. 342–351 (2005)

13. Zhai, C., Veliveli, A., Yu, B.: A Cross-Collection Mixture Model for Comparative Text Mining. In: Proc. of SIGKDD 2004, pp. 743–748 (2004)
14. Ohshima, H., Oyama, S., Tanaka, K.: Searching Coordinate Terms with Their Context from the Web. In: Aberer, K., Peng, Z., Rundensteiner, E.A., Zhang, Y., Li, X. (eds.) WISE 2006. LNCS, vol. 4255, pp. 40–47. Springer, Heidelberg (2006)
15. Bollegala, D., Matsuo, Y., Ishizuka, M.: Measuring Semantic Similarity between Words Using Web Search Engines. In: Proc. of WWW 2006, pp. 757–766 (2006)