International Journal of Innovative
Computing, Information and Control
Volume **14**, Number **6**, December **2018**

ICIC International ©2018 ISSN 1349-4198
pp. 2353–2360

RESEARCH ON EXPERT OPINION CREDIBILITY RATING IN VERTICAL FIELD

Yao Liu[1], Qingxu Wang[2] and Yi Huang[1]

[1]Institute of Scientific and Technical Information of China
No. 15, Fuxing Road, Haidian Dist., Beijing 100038, P. R. China
liuy@istic.ac.cn

[2]China Meteorological Administration
No. 46, Zhongguancun South Street, Beijing 100081, P. R. China

Received March 2018; revised July 2018

Abstract. *Nowadays, online celebrities became highly influential in shaping public opinion and disseminating information without supervision. This paper proposes a method of calculating credibility scoring of expert opinion so as to provide some useful information to correctly guide public opinion. This paper studies the credible evaluation of expert opinion from two aspects: the extraction of rating information and the credibility rating of expert opinion. We extract score information from expert opinions by using sequence annotation, construct the credible rating standard considering a number of factors, compare it with the public official statistics in the vertical field, and generate the credibility score of expert opinion based on the comparison results.*
Keywords: Expert opinion, Credibility rating, Vertical field

1. **Introduction.** With the rapid development of network media, online celebrities became highly influential in shaping public opinion and disseminating information without supervision, and have deeply penetrated into the social politics, economy, culture, life and so on. How to correctly guide public opinion is an important challenge for the Internet media. Information credibility is a qualitative assessment or quantitative measure of the credibility of information. At present, research on the credibility of online information mainly focuses on the credibility of website information with user reviews, the credibility of online news, and credibility of search engine results.

Research abroad mainly focuses on evaluation methods, evaluation algorithms, and evaluation systems. Zheng and Liu [1] proposed a quantitative information credibility evaluation model based on fuzzy mathematics, but its fuzzy measurement lacked strict monotony. Liu et al. [2] used a dynamic programming method with minimum cost flow formula and maximum weight matching algorithm to measure the feasibility of constructing online information credibility model from a numerical perspective. Caslillo et al. [3] studied the credibility assessment method for newsworthy information based on the characteristics of the social media environment and users' active participation, and further proposed a model for predicting the credibility of information.

The domestic research on information credibility focuses mostly on qualitative analysis, and less on the quantitative evaluation methods of information credibility. Zhang [4] conducted a survey through questionnaires and analyzed the evaluation factors of credibility of online information. Sun and Liu [5] studied the credibility of online purchase reviews based on the characteristics of review information structure. Qu and Xie [6] mainly used qualitative analysis methods to combine four aspects: site level, layout level, topic level,

DOI: 10.24507/ijicic.14.06.2353

and content level to construct an online academic forum information credibility evaluation index system and a fuzzy comprehensive credit evaluation model. He et al. [7] used the SVM model to predict whether Weibo information was a rumor with features such as keyword distribution, and time difference. Cheng et al. [8] proposed a model based on BP neural network and improved its excitation function to construct a Weibo rumors detection model. Zhang et al. [9] built an information credibility index system, and used the analytic hierarchy process (AHP) structural model to determine the weight coefficient of the index. Yang et al. [10] proposed a rumor detection method based on momentum model for sudden topic detection and domain expert discovery, to identify the authenticity of topic information by discovering domain-related microblog users in the candidate pool, based on the domain relevance of topic and user's personal information.

Compared with foreign research, from the content point of view, the current research on credibility analysis in China is mostly based on the research on credibility of credential information and the credibility of microblog information. From the perspective of research methods, most of them are qualitative methods, and few of them are automated evaluation methods.

The rest of this paper is organized as follows. Section 2 describes the method and framework. Section 3 describes the expert opinion scoring information extraction. Section 4 describes expert opinion credibility rating. Finally, conclusion and future work are given in Section 5.

2. **Method and Framework.** The overall framework is shown in Figure 1. Score information extraction includes the extraction of elements, and conditional random field model is adopted for extraction. The expert opinion credibility rating includes classification and calculation of rating information and formulation of the credibility rating standard. We perform multivariate analysis of the scoring information in the opinion sentence, mainly constructing a four-tuple {subject, time, trend, numerical value} as the extraction target, training and extracting with the conditional random field method. In view of the accuracy rate and the number of opinions involved in the comparison results, we use the IMDB algorithm to propose the calculation standard of the credibility score of the expert opinion. The overall idea and framework of this section are shown in Figure 1.

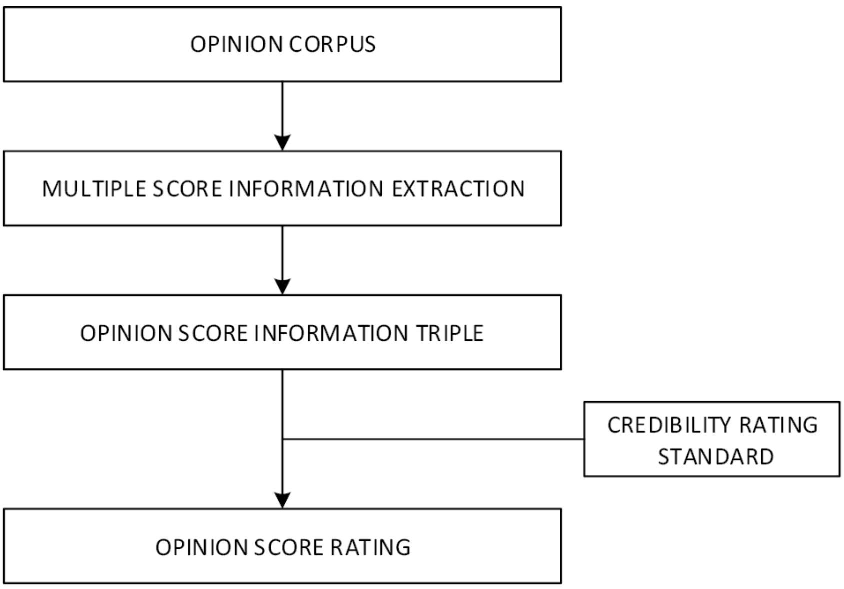

FIGURE 1. The overall framework

3. Expert Opinion Scoring Information Extraction.

3.1. Expert opinion scoring information analysis. Expert opinion scoring information refers to the information contained from the expert opinion that can be compared with public data. We propose to construct a four-tuple of score information {subject, time, trend, value}.

Considering that we need to extract four different types of information, we choose to use a statistical-based sequence labeling method rather than rule-based one to extract the score information, by setting the feature template trained with a manually labeled corpus to obtain the corresponding probability model for extraction.

3.2. Score information extraction. We choose the conditional random field (CRF) model to extract the scoring information in the vertical domain.

Using an economic corpus as an example, we extracted 5,765 sentences that contain expert opinions out of 8,000 articles from "Sina Finance Blog" and "NetEase Finance", and chose 4,233 sentences that contain score information as basic corpus by manual screening. The labeling rule of scoring information for the four-tuple {subject, time, trend, value} is shown in Table 1.

The four symbols (B, M, E, F) indicate the position of a term, and the six symbols (H, I, J, K, L, O) indicate the type of scoring element. Taking the sentence "China's economy can achieve a 7% increase in the next five years" for an example, the result is shown in Table 2.

The first column is the segmentation result, the third column is the result of the part-of-speech tagging, and the fourth column is the manual annotation.

Then, we design the feature template. According to the corpus, the maximum length of a chunk containing a complete scoring element is generally 5, and we set the window

TABLE 1. Rating information labeling rules

Symbol	Description	Symbol	Description
B	Beginning of a scoring element	H	Subject
M	Middle of a scoring element	I	Time
E	End of a scoring element	J	Time interval
F	A full scoring element	K	Value
		L	Trend
		O	None

TABLE 2. Example of rating score information

Segmentation (Chinese)	(English)	POS	Annotation
未来	next	nt	B-J
五	five	m	M-J
年	years	q	E-J
中国	China	ns	O
经济	economy	n	F-H
可	can	v	O
实现	achieve	v	O
7%	7%	m	F-K
的	–	u	O
增长	increase	v	F-L

of feature template to 5, that is, considering the relationship of the current element with the forward two items and the backward two items in vertical direction, and relationship between word segmentation results and part-of-speech tagging results in horizontal direction.

"U:%x[row,col]" indicates the position of the feature relative to the current tag, "row" indicates the relative line position of the feature relative to the current marker, and "col" indicates the absolute column position of the feature.

3.3. **Experimental results and analysis.** The basic corpus is divided into a training set and a test set, and we conducted a 10-fold cross-validation. The evaluation results are shown in Table 3.

TABLE 3. CRF opinion score information quad tuple selection evaluation

Element	Accuracy	Recall	F1
Subject	0.8525	0.8033	0.8272
Time	0.7954	0.7678	0.7814
Trend	0.7843	0.7496	0.7666
Value	0.8324	0.8065	0.8192

Judging from the experimental results, the extraction of subject and value elements is better than time and trend element, and the accuracy rate, recall rate, and F1 score all reach 0.8. And the reason is that the designed CRF feature template mainly considers word segmentation and part of speech tagging results, and the wrong ones have a certain impact on the extraction results. The POS information of words representing subject and value element is relatively single, while that representing time and trend elements include a variety of POS information, including noun, verb, numeral, etc. So the word segmentation and POS results have multiple combinations and have different probabilities in labeling samples.

4. **Expert Opinion Credibility Score.**

4.1. **Classification and calculation of opinion scoring information.** The scoring information can be classified into categories and compared with the corresponding public data in the vertical field. We consider factors such as the number of expert opinions, the accuracy rate, and the minim number of opinions. The "time", "trend", and "value" information extracted by the CRF model contains multiple types, and it is necessary to further extract the score information according to the rules of extraction and classification. The design rules are as follows.

(1) Time element classification and extraction rules

The elements of time information in expert opinions fall into two main categories: time points and time intervals, as shown in Table 4.

For Type 1, a time-trigger vocabulary is constructed based on the time information obtained by extracting the CRF in the basic corpus, and the trigger words such as "year,

TABLE 4. Example of time element types and descriptions

Time Type	Example (Chinese)	(English)
Time	"2015年、今年"	in 2015, this year
Time Interval	"未来15年、未来五年到十年"	in the next 15 years, in the next five years to ten years

month, and day" are extracted. For information that does not have a definite time (for example, this year, next year), it is judged by the time when the corpora are published.

For Type 2, we need to extract the start time and end time. We sum up the time information template, such as: "in xx years", "up to xx years", and "from xx year to xx year". For those with a definite start and end time, the start and end time is extracted according to time trigger words such as "year, month, and day"; for those without a clear start and end time, the publication time of the article can be regarded as the start time, and the end time is calculated based on the time interval extracted.

(2) Value element classification and extraction rules

The elements of value information in expert opinions can be divided into three major categories: integers, decimals, and percentages, as shown in Table 5.

Table 5. Numerical element types and description examples

Value Type	Example (Chinese)	(English)
Integer	"达到10245元、10亿"	reached 10245 yuan, 1 billion
Decimal	"2.5倍、5.04美元"	2.5 times, \$5.04
Percentage	"7.5%、6.5%至8%"	7.5%, 6.5% to 8%

Percentages and non-percentages can be distinguished by trigger words such as "%, ¥, RMB, and dollar"; integers and decimals are classified by regular expressions; numerical modifiers such as "reach, over, under" are also classified. The specific categories are shown in Table 6.

Table 6. Classification of numerical value modifiers

Example (Chinese)	(English)
"超过、超、大于、高于"等	over, above
"低于、跌至、更低、降至"等	below, under
"达到、达、到"等	reach
"至、~、之间、左右"等	to, between, about

(3) Trend elements classification and extraction rules

The elements of trend information in expert opinions can be divided into four categories: rising, falling, stable, and change. The specifics are shown in Table 7.

Table 7. Example of trend element types and descriptions

Trend Type	Example (Chinese)	(English)
Rising	"增加、升值、高涨"	increase, rise
Falling	"贬值、降息、下跌"	decrease, fall, drop
Stable	"稳定、均衡"	stable, balanced
Change	"波动、逐渐企稳"	fluctuated, gradually stabilized

Considering elements in the four-tuple might be absent, those without any "subject", "time", "value", "trend" information are filtered as noise data, and for those with incomplete information, the missing elements are recorded as 0. The opinion scoring information extraction and classification results are shown in Figure 2.

As can be seen from above, the results are separated by "#" lines. The final score information result of each opinion includes "Value Type", "Value", "Value Modifier", "Subject", "Trend", "Time Type" and "Time Value".

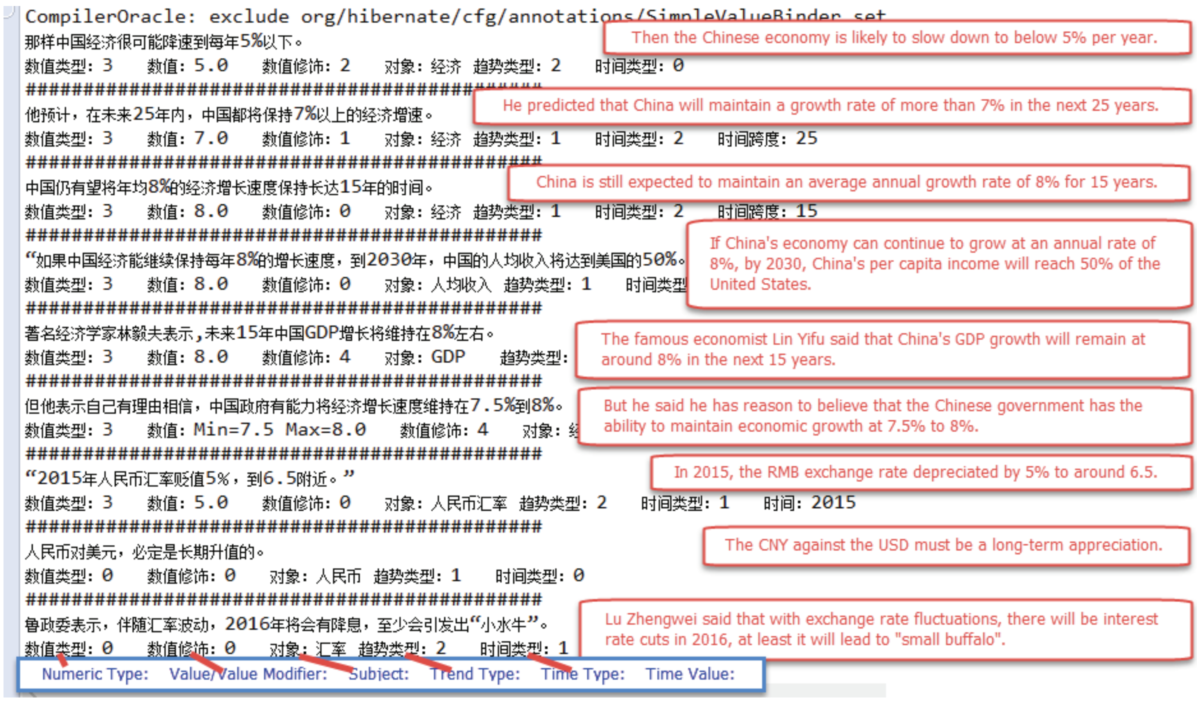

FIGURE 2. Rating score information parameter results

By constructing a "subject-indicator map", the mapping relationship between the "subject" in the expert opinion and the "indicator" in the national public database is established, and the score parameters <indicator name, time type, trend type, value type, and value modifier> are obtained. The calculation formula of these indicators is configured according to the index formula commonly disclosed in the vertical domain, and the calculation method calling method is exemplified by the corpora in the economic field. The configuration file format is shown in Figure 3.

```
"HousePrice1130","房价1130": ["lambda x:(x[1]-x[0])*100/x[0]"]
"HousePrice1230","房价1230": ["lambda x:(x[0]-x[1])*100/x[1]"]
```

FIGURE 3. Index calculation formula configuration file

Each of the calculation formulas has a key-value pair: key is the type of the calculation formula, which is composed of the sentence rating information configuration parameters; value is the corresponding calculation formula.

4.2. **Criteria for credibility of expert opinion.** Expert opinion credibility scores should consider science and fairness in the formulation of scoring standards. We refer to the Top 250 scoring standards of IMDB, and propose the Credibility Scoring (i) formula for expert i on topic T:

$$\text{Credibility Score } (i, T) = \frac{N_i}{N_i + m} * R_i + \frac{m}{N_i + m} * C \tag{1}$$

In the formula, N_i is the number of opinions published by expert i on topic T, and m is the minimum number of opinions published by an expert on subject T that can be participated in the score calculation. R_i represents the average accuracy of opinions expressed by expert i on topic T. R_i is calculated as follows:

$$R_i = \frac{\sum_{j=1}^{N_{i(j)}} \frac{a_{i,j}}{s_{i,j}}}{N_{i(j)}} \tag{2}$$

$N_{i(j)} > 0$ represents the number of articles related to expert i on topic T; $\sum_{j=1}^{N_{i(j)}} \frac{a_{i,j}}{s_{i,j}}$ represents the sum of accuracy of expert i on topic T. The accuracy rate of article j is calculated by $a_{i,j}/s_{i,j}$ where $a_{i,j}$ is the number of credible sentences in article j, $s_{i,j}$ is the number of scoring opinions in article j. ($a_{i,j} > 0$, $a_{i,j}/s_{i,j}$ is in range $[0, 1]$).

C represents the average accuracy of opinions expressed by all experts on topic T. The formula is as follows:

$$C = \frac{\sum_{i=1}^{N_{(i)}} \sum_{j=1}^{N_{(j)}} \frac{a_{i,j}}{s_{i,j}}}{\sum_{i=1}^{N_{(i)}} N_{i(j)}} \tag{3}$$

$\sum_{i=1}^{N_{(i)}} N_{i(j)}$ is the number of linguistic data for all experts on the topic T. Since $N_{i(j)} > 0$, the denominator > 0; $\sum_{i=1}^{N_{(i)}} \sum_{j=1}^{N_{(j)}} \frac{a_{i,j}}{s_{i,j}}$ is the sum of accuracy of all experts on topic T.

According to Formulas (1), (2), (3), the Credibility Score $CS(i)$ of expert i on topic T can be calculated, and the average credibility of expert i on all topics can also be calculated. The formula is shown as follows:

$$\overline{CS(i)} = \frac{\sum_{t=1}^{N_{i(t)}} RS(i)_t}{N_{i(t)}} \tag{4}$$

In the above formula, $N_{i(t)}$ is the number of topics in the vertical field. Since the range of $a_{i,j}/s_{i,j}$ is $[0, 1]$, the ranges of the credibility index $CS(i)$ of expert i on topic T and the credibility index $\overline{CS(i)}$ of expert i under all topics are within range $[0, 1]$.

4.3. **Experiment and result analysis.** 45,761 articles from the "NetEase Finance" and "Sina Finance Blog" websites are used to obtain opinions and information. The scoring information is compared with the data in the National Bureau of Statistics. When comparing, the matching error e is set to 0.1, i.e., the error range within 10% is a trustworthy opinion. We use the Credibility Score (i, T) formula to calculate the expert's credibility score, by setting the minimum number of published opinions m to 30 to make the expert credibility score value within range $[0, 10]$. The results are shown in Figure 4.

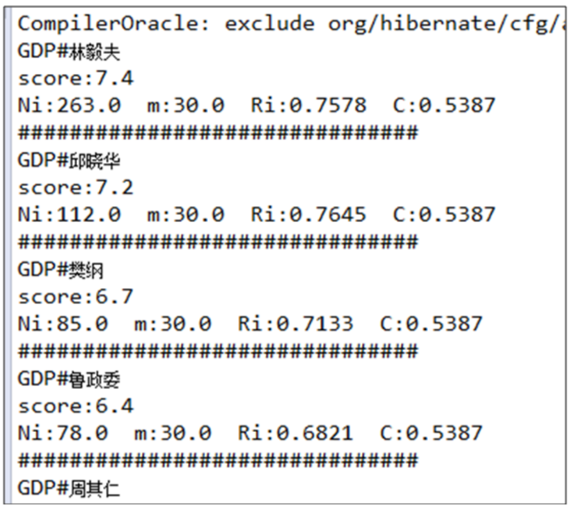
```
CompilerOracle: exclude org/hibernate/cfg/
GDP#林毅夫
score:7.4
Ni:263.0  m:30.0  Ri:0.7578  C:0.5387
###############################
GDP#邱晓华
score:7.2
Ni:112.0  m:30.0  Ri:0.7645  C:0.5387
###############################
GDP#樊纲
score:6.7
Ni:85.0  m:30.0  Ri:0.7133  C:0.5387
###############################
GDP#鲁政委
score:6.4
Ni:78.0  m:30.0  Ri:0.6821  C:0.5387
###############################
GDP#周其仁
```

FIGURE 4. Expert credibility score results

As can be seen from above, the parameters and results of expert opinion credibility score with respect to the "GDP" topic are displayed. The first line of each result unit is "subject # expert", such as "GDP # Lin Yifu(林毅夫)"; the second line is the opinion credibility score; the third line is the value of each parameter and the scoring result: N_i is the number of articles published by the expert on the subject, m is the number of minimum opinions that an expert is able to participate in the rating on topic T, R_i is the average accuracy of expert's opinion on topic T, and C is the average accuracy rate of all experts on topic T.

According to the results, the Credibility Score (i, T) formula can give experts who have published a certain number of opinions on a topic a reasonable score within range $[0, 10]$, according to the number of opinions, the accuracy rate of expert opinion. It can be concluded that Credibility Score (i, T) formula is feasible in actual use in vertical field, and the parameter (e, m) should be fine-tuned according to features in other vertical fields to make the result more reasonable and reliable.

5. Conclusions.

This paper studies the credibility evaluation of expert opinion from two aspects: the extraction of rating information and the credible rating of expert opinion. We extract score information from expert opinions by using sequence annotation, construct the credibility rating standard considering a number of factors, compare it with the public official statistics in the vertical field, and generate the credibility score of expert opinion based on the comparison results. In the future, we plan to explore expanding the training corpus and optimizing the CRF extraction results of scoring element, and make some application in other fields rather than economic field.

Acknowledgment. This work is supported by National Key Project of Scientific and Technical Supporting Program No. 2013BAH21B02, which have improved the presentation.

REFERENCES

[1] Y. Zheng and B. Liu, Fuzzy vehicle routing model with credibility measure and its hybrid intelligent algorithm, *Applied Mathematics and Computation*, vol.176, no.2, pp.673-683, 2006.

[2] B. Liu et al., Optimizing information credibility in social swarming applications, *IEEE Trans. Parallel and Distributed Systems*, vol.23, no.6, pp.1147-1158, 2012.

[3] C. Castillo, M. Mendoza and B. Poblete, Predicting information credibility in time-sensitive social media, *Internet Research*, vol.23, no.5, pp.560-588, 2013.

[4] M. Zhang, Research on credibility of network information: Perspectives of netizens, *Journalism & Communication Research*, vol.12, no.2, pp.17-27, 2005.

[5] C. H. Sun and Y. Z. Liu, Experimental study on credibility of Internet word of mouth, *Finance and Economics Review (Journal of Zhejiang University of Finance and Economics)*, vol.145, no.4, pp.96-102, 2009.

[6] W. Qu and D. Xie, A gray analysis of information reliability of network academic forum, *Documentation, Information & Knowledge*, vol.2, pp.112-118, 2013.

[7] G. He et al., Research on Weibo proverbs recognition, *Library and Information Service*, vol.57, no.23, pp.114-120, 2013.

[8] L. Cheng, Y. F. Qiu and L. Sun, Research on Microblog booster detection methods, *Computer Applications and Software*, vol.30, no.2, pp.226-228, 2013.

[9] T. Zhang, Y. S. Zhang and Q. S. Liu, Microblog user information credibility model based on analytic hierarchy process, *Journal of Beijing Information Science and Technology University*, vol.4, pp.61-65, 2015.

[10] W. T. Yang et al., Weibo rumor detection method based on sudden topic and domain expert, *Journal of Computer Applications*, vol.37, no.10, pp.2799-2805, 2017.