Analysis of Competition in Chinese Automobile Industry based on an Opinion and Sentiment Mining System

Xie Xinzhou[*], Wang Qiang[**], Chen Anqi[**]

* Competitive Intelligence and Competitiveness Research Center of Peking University, Beijing, China.
xzxie@pku.edu.cn

** Key Laboratory of Competitive Intelligence and Innovation Evaluation, Beijing Academy of Science and Technology, Beijing, China.
wq.malmsteen@gmail.com

ABSTRACT: In this paper a methodology for a mining system is introduced. The architecture of the system is based upon what is called opinion and sentiment mining. The mining system is used to analyze competition in the auto industry. The results show the advantages with each of the two cars used for this study. Instead of offering theory this is a hands-on approach to help solve specific problems by describing a complex process.

KEYWORDS: Competitive Intelligence, Opinion Mining, Chinese Automobile Industry

1. Introduction

Internet has become the main source for Competitive Intelligence (CI). The reason is that internet users express their opinion and attitude towards products and images of enterprises online. This paper presents a concept for how to analyze the competition in the automobile industry. The main focus is based on what is called opinion and sentiment mining. A comparative analysis between two auto brands in China is shown as an example.

First the role of opinion and sentiment mining in CI will be introduced. Further on we present the methodology for this study as well as key issues of opinion and sentiment mining. Finally the architecture of the opinion and sentiment mining system and how to use this system to analyze the competition in the auto industry is discussed.

1.1 The Role of Opinion and Sentiment Mining in CI

As shown in Table 1, internet users increased dramatically with the development of internet over the past years. The number of internet users has

reached close to 2 billion, and is about 30% of the world's population. The number is higher in developed countries and developed areas. (Table 1)

World Regions	Population (2010 Est.)	Internet Users Dec. 31, 2000	Internet Users Latest Data	Penetration (% Population)	Growth 2000–2010	Users % of Table
Africa	1,013,779,050	4,514,400	110,931,700	10.9 %	2,357.3 %	5.6 %
Asia	3,834,792,852	114,304,000	825,094,396	21.5 %	621.8 %	42.0 %
Europe	813,319,511	105,096,093	475,069,448	58.4 %	352.0 %	24.2 %
Middle East	212,336,924	3,284,800	63,240,946	29.8 %	1,825.3 %	3.2 %
North America	344,124,450	108,096,800	266,224,500	77.4 %	146.3 %	13.5 %
Latin America/Caribbean	592,556,972	18,068,919	204,689,836	34.5 %	1,032.8 %	10.4 %
Oceania / Australia	34,700,201	7,620,480	21,263,990	61.3 %	179.0 %	1.1 %
WORLD TOTAL	6,845,609,960	360,985,492	1,966,514,816	28.7%	444.8 %	100.0 %

Table 1. World internet users and population statistics

Users express their thoughts online, making internet the main information distribution and access channel. This provides new opportunities and challenges for the development of CI as a discipline. It opens up user preferences and topics such as:

- How do users evaluate the products?
- Do users like the products?
- Which properties of the products make users like or dislike them?
- How do internet users perceive the image of the enterprise?
- Which practices of the enterprises do users like or dislike?
- How do users choose between different products?
- What properties make users buy the products?

Opinion and sentiment mining provide views and preferences of internet users for different companies. The users' comments are important for companies and for product development.

Take Windows Vista as an example. Vista has been selected by Time magazine as one of the 10 biggest tech failures. Mr. Nash, Windows vice president of product, confirmed the hesitation to launch the product, based on early users opinions. It said that the service was not being user-friendly, which again influenced other users in a negative way.

Users of products are an important information source for CI, and their opinions can provide companies with rich contents, making them an important reference for enterprises.

2. The Methodology

Opinion and sentiment mining goes through five major steps as shown in figure 1:
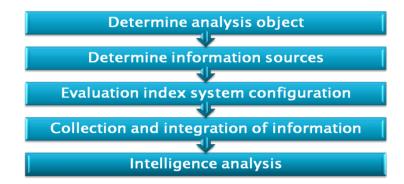

Figure 1. Framework of opinion and sentiment mining

1. Determine object analysis

   In the object analysis stage we answer the following questions:

   - Which competitors should be analyzed?
   - What are the products, brands and services of the competitors?

   According to our needs, these aspects are defined as our objects.

2. Determine information sources

   In the stage of determining information sources, an alternative information source list can be created, containing authoritative forums, web stations, and blogs. It can be filtered according to the influence and quality of the information. It can also be filtered and complemented with help from industry experts.

3. Evaluation index system configuration

   The third step is to build an evaluation index system to describe the properties of our objects. For example, the index system may contain engine, computer screen, wheel, seat and so on in an auto industry analysis. The index system creates an alternate property list. A sentiment vocabulary need to be built, which describes the "sentiment" of the properties like good, excellent, terrible and so on. In this step, the participation of industry experts who will help us filter and complement the property list and sentiment vocabulary is necessary. The relationship and weight of properties should be determined, after which a complete index system is constructed.

4. Collection and integration of information

   The properties of index systems are used as the query words to retrieve from the information sources. At the same time, the opinion and sentiment words are extracted. This information will be integrated into the opinion and sentiment database.

5. Intelligence analysis

   The final step is to analyze the data. Before the analysis, some provisions need to be done, including error correction and elimination of duplicates. Then we need to identify the emotion tendency, which can be positive, neutral or negative. Some intelligence analysis methods like association, comparative and trend analysis are used to research the competitive situation further.

3. Key Issues

The introduction above is the framework of the methodology, and in almost every step there are some key issues including:

- How to select the more authoritative information sources?
- How to obtain and integrate the information which is heterogeneous?
- How to build index systems which can describe our objects comprehensively?
- Choose an opinion and sentiment mining algorithm.

(1) Selection of authoritative information source

In the source selection, methods such as web metrics can be used to evaluate the information source, and inputs from industry experts are essential.

(2) Acquisition and integration of multiple heterogeneous information sources

During the acquisition and integration of multiple heterogeneous information, spam and filter noise should be removed through metadata standards, using segmentation algorithms to process unstructured and semi-structured information.

(3) Evaluation index system

For different CI tasks, the index system is different. This step is a semi-automated process and some work must be done manually. In order to improve efficiency, software to help industry experts build or modify the index system was developed.

(4) Opinion and sentiment mining algorithms

The core part of the opinion and sentiment mining system is the algorithms, which include the corpus-based approach, dictionary-based approach, supervised machine learning methods, image segmentation algorithm and other opinion extraction algorithms. During the development of this system, a dictionary-based algorithm is more suitable for Chinese information processing, and the accuracy is about 82%. That is acceptable for a commercial operation.

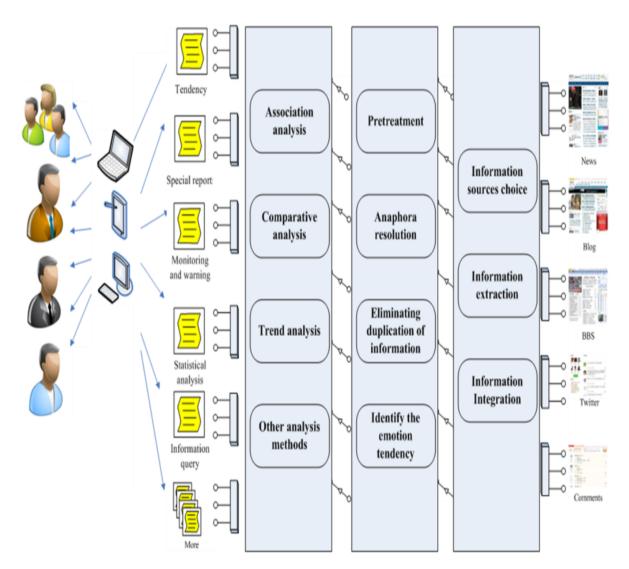**4. Architecture of Opinion and Sentiment**

Figure 2. Architecture of opinion and sentiment mining system

The Opinion and Sentiment Mining System is developed to gather data about opinions and sentiments related to products and services. The system consists of four parts: data acquisition, data pretreatment, data analysis and user interface, as shown in figure 2.

- The function of the data acquisition part is information selection, information extraction and information integration;
- The function of the data pretreatment is to eliminate duplication of information, do error correction, emotion tendency judgment and so on;
- The main task of the data analysis part is to do association research, comparative research and trend research;
- The analysis of the result will be shown through different types of terminals.

5. Analysis of Competition in the Chinese Auto Industry

How to use this system to analyze the competition in China's auto industry will be illustrated through a case study. In this case, Peugeot 307 and Ford Focus (shown as figure 3), are used as examples. Both cars have a high selling rate and the competition between them is fierce. We performed an analysis of the competition of the two cars through analyzing the comments of internet users.

Figure 3. Auto products used in the case study

(1) Information Source

The information was mainly collected from auto forums using systems and saved information in Databases which provided information about the targeted cars. The information sources are shown in table 2.

No	Url	Logo
1	http://www.autohome.com.cn/	
2	http://www.xcar.com.cn/	
3	http://www.chetx.com/	
4	http://auto.sina.com.cn/	
5	http://auto.qianlong.com/	
6	http://www.ieche.com/	
7	http://auto.sohu.com/	
8	http://auto.huanqiu.com/	
9	http://www.feelcars.com/	

Table 2. Information Source

(2) Evaluation Index System

The index system was established containing properties, such as sunroof, abs, air-condition and engine. Indicators used in the index system are shown in table 3.

Sunroof	Seat belt	Airbags	Sunshade
Chassis	Alloy wheels	Single-disc CD	Rear brake
Power Window	Head airbags	Multi-disc CD	Auto parking
EBD	Car phone	Cylinders	Front passenger airbags
Side airbags	Bluetooth	Max.hp	Engine
CD Support	Rear outlet	Temperature zone	Maximum torque
Center armrest	CND	control	Windshield wiper
Rearview mirror	Electric trunk	Trim	Stroke
Valve structure	Seat	ASCD	Sport kit
External audio interfaces	Vehicle door	Steering wheel	View Camera
Center console	ABS	Cylinder bore	Air conditioning
Air-condition	BA	EAS	Maximum speed
Spare wheel	Central locking	Rear side airbags	Multi-disc DVD
GPS	Keyless Go	Car TV	Power Assisted Steering
Body side molding	Rear LCD screen	Drive mode	Computer screen
Sun visor mirror	Single-disc DVD	DRL	Front Suspension
Speaker	Cylinder cover	internal hard disk	Head lamp
Brake pedal	QA Quality Assurance	Displacement	Tumbler holder
Front brake	Appearance	Maximum power	Fuel way
Fuel consumption	RKE	Compression ratio	Man-machine interactive
Transmission	Rear suspension	Cylinder stator	system
Headlight	Tire	Rear head airbags	Others
		HUD	

Table 3. Indicators used in the index system

(3) Sentences Extracted by the System

A data set can be obtained through opinion extraction. Take Peugeot 307 for example (shown in figure 4), the first line is the sentence about appearance, the second is about other properties that is not described in the remaining part, the third is about air-condition and the fourth is about doors.

Figure 4. Sentences extracted by the system (in Chinese)

(4) Attention Comparison

Figure 5 is the comparison of the attention between our targeted cars. Attention is measured by the number of posts about the given car. The red line is the attention of Ford Focus and the green line is for Peugeot 307. In this figure it is shown that users pay more attention to Ford Focus than to the Peugeot 307.
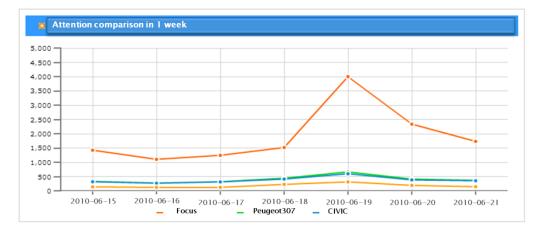

Figure 5. Attention comparison between Peugeot 307 and Ford Focus

(5) Positive Comments

After identifying the emotional tendency, we summed up the positive comments through which a trend of the users' positive comments are shown. The number of positive comments for Ford Focus is higher than for Peugeot 307, which indicates that users prefer the Ford Focus over Peugeot 307. This results may help people who want to buy a family car make their decision. It can also attract the attention of staff from Peugeot 307 who should like to change the image of the car.
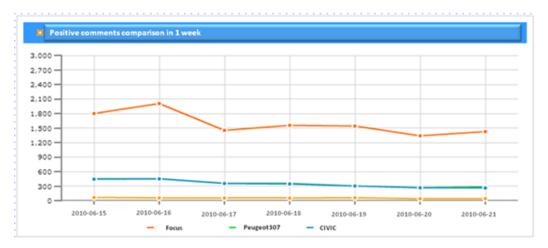

Figure 6. Positive comments of target cars

(6) Negative Comments

Figure 7 shows the comparison of the negative comments. In this figure we see that the negative comments about these two cars are similar. After combining the positive and the negative analysis, the conclusion is that the negative comments occupy much larger proportions of the users' comments of Peugeot 307 than for Ford Focus.

Figure 7. Negative comments of target cars

(7) Skylight Comparison

A comparison of the selected properties of the two cars is valuable because it tells us why the users like or dislike the products. The comparison in figure 8 shows that users prefer the skylight of Peugeot 307 over the skylight of Ford Focus.
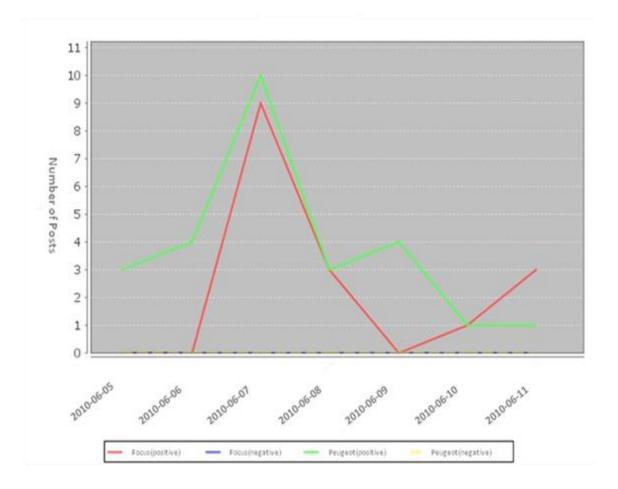

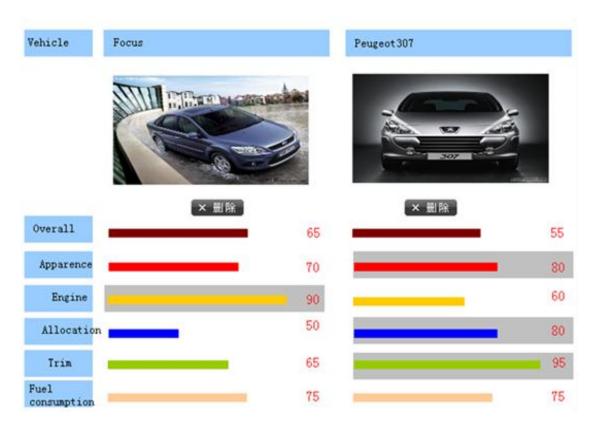*Figure 8*. Skylight comparison between target cars

Figure 9. Overall comparison between target cars

(8) Overall Comparison

Other properties are compared in a similar way achieving this overall result. We see that compared to Peugeot 307, users prefer Ford Focus, but the appearance and trim of the Peugeot 307 is preferred to its rival.

Peugeot 307 is better on	Ford Focus is better on
Skylight, Fuel consumption, Seat, Appearance, Trim, Headlight, Door, RKE, Cruise Control System, ABS, Electronic anti-theft, Speaker	Engine, Air-condition, Rear suspension, Tire

Table 4. Comparison result of Peugeot 307 and Ford Focus

(9) Comparison Result

We came to the conclusion that the advantages of Ford Focus is the car's power and performance, which is embodied in the engine, air-condition, rear suspension and tire. Peugeot 307 on the other hand has an advantage in appearance and design which is embodied in the skylight, fuel consumption, seat and so on.

Peugeot 307	Ford Focus
Increase the PR about appearance and design.	Let consumers understand the importance of vehicle performance.
Fix engine deficiencies.	Strengthen the design of appearance and trim.

Table 5. Recommendation according to opinion and sentiment mining

6. Outlook

Further research in this field could include:

- Use Opinion and Sentiment Mining System to perform other industry analysis, such as for cosmetic industry and health industry and see what are best applied areas.
- Improve the accuracy of the opinion extraction and sentiment judgment;
- Embed natural language processing algorithms of other languages, which can make this system analyze the information of several languages at the same time.

References

A HowNet Word List for Sentiment Analysis (beta version). Retrieved 2010-04-30. Available online at URL: http://www.keenage.com/html/c_index.html.

Agarwal, A. & Bhattacharyya, P. (2005). Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified. *Proceedings of the International Conference on Natural Language Processing (ICON)*.

Chao, L., Jian, S., Yi, G., Xingjun, X., Lei, H. & Sheng, L. (2009). etc. Chinese Chunking With Maximum Entropy Models. *Proceedings of CIPS-ParsEval-2009*.

Fuld & Company. *Intelligence Software Report 2008-2009*. London, United Kingdom. Fuld & Company, Inc. 2009.

Gang, L. & Qiangbin, D. (2008). An Approach Based on Words Numbers for Extracting Text from Web Pages. *Information Science*, 26(3).

Hatzivassiloglou, V. & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 35th annual meeting of ACL*.

Internet World Stats. Available online at URL:http://www.internetworldstats.com/stats.htm.

Pang, B. & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2): 1-135.

Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. *Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management, Bremen, Germany*. pp. 625–631.

Zhao, J., Xu, H., Huang, X., Tan, S., Liu, K. & Zhang, Q. (2008). Overview of Chinese Opinion Analysis Evaluation 2008. *Proceedings of the First Chinese Opinion Analysis Evaluation (COAE 2008)*. pp. 1-20.