

Athens Institute for Education and Research

ATINER



ATINER's Conference Paper Series

LNG2013-0833

**Automatic Survey of Customer
Sentiments from Chinese Social Media**

Wei Li

Chief Assistant

Netbase

USA

Lei Li

Chief Assistant

Netbase

USA

Tian Tang

Chief Assistant

Netbase

USA

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209
Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

Printed in Athens, Greece by the Athens Institute for Education and Research.
All rights reserved. Reproduction is allowed for non-commercial purposes if the
source is fully acknowledged.

ISSN 2241-2891

23/1/2014

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Hanna, S.M. (2013) "Automatic Survey of Customer Sentiments from Chinese Social Media" Athens: ATINER'S Conference Paper Series, No: LNG2013-0833.

Automatic Survey of Customer Sentiments from Chinese Social Media

Wei Li

Chief Assistant

Netbase

USA

Lei Li

Chief Assistant

Netbase

USA

Tian Tang

Chief Assistant

Netbase

USA

Abstract

Social media in Chinese has seen an explosive growth in the last few years. Due to the quantity of social media big data, it contains consumers' spontaneous comments on almost any brands, hence a valuable source for surveying customers' sentiments for business insights. Automatic survey from Chinese social media enabled by Natural Language Processing is a way of doing just that, in scale. This paper presents the general architecture and implementation of a Chinese customer insight system mining from social media, to serve businesses for cases like risk management as well as brand comparison. Benchmarks show that the sentiments extracted by this system reach high precision (about 80%) and fair recall that is comparable with the English system.

Keywords: NLP; Chinese parsing; social media; sentiment extraction; public opinions; text *mining*

Acnolwdgement: Thanks go to Zixin Jiang, Martin Min, Ray Hsu, Cheng-Ying Wu, Xin Lu and Karin Golde for their involvement and contributions in the Chinese development, deployment and testing work. HowNet is used as part of lexical resources for this study.

Corresponding Author:

Introduction

The Chinese social media has seen explosive growth in recent years, especially for the Sina Weibo (Microblog) and Tencent Weixin (WeChat). The popularity of the social media makes it a gold mine of customers' intelligence that no serious businesses can ignore. Recent estimates indicate that on average one in every three blog posts and one in five tweets involve comments on products, services or brands (Hogenboom et al. 2013). Random users freely talk about whether they love or hate a brand, and lots of times they compare it with other brands in the same category. Apparently, such information would be really important for businesses to keep track of consumers' attitudes toward their brands. The management can make faster decisions based on the social intelligence when it is extracted from the huge social data pool and analyzed properly. The big data nature of such information calls for automatic mining instead of manual collection. This paper presents the work on how we support automatic survey of any brands by mining social media deep sentiments based on a robust Chinese parser.

Section II is literature review for related work. Section III discusses the concept of deep sentiments and the operational definition as a target for supporting automatic survey. Section IV describes the architecture and implementation of a customer insight mining system supported by deep sentiments based on parsing. Section V shows some experiments of automatic survey, with discussions on the test results.

Related Work

Sentiment analysis has been a focus of research in the last decade, especially when associated with social media. Natural language mainly involves two types of expressions, one called subjective language, and the other objective language (Bruder & Janyce 1990; Yu and Vasileios 2003). The former is used for stating facts or evidence while the latter is to express sentiments or judgments. Traditional information extraction systems target facts such as relationships and events (Chinchor and Marsh 1998). The more recent research on extraction of sentiments from subjective language has drawn special interests from industry as the underlying technology opens the door to the untapped unstructured text world in social media, to help gain business insights and listen to the customer's voice. Whether the sentiments represent customers' happiness with a product or complaints, they are invaluable customer insights which the businesses so far can only manually collect from surveys.

In the research literature, the main stream of sentiment analysis has been sentiment classification based on BOW (Bag of Word) model using machine learning algorithms (Pang and Lee 2008). Such keyword-based learning mainly supports course-grained sentiment classification, basically tagging the incoming document or post as positive or negative, also called thumbs up and

down classification (Pang, Lee and Vaithyanathan 2002; Turney 2002; Taboada et al 2011). This approach to coarse-grained sentiment analysis has the benefits of quick implementation and good performance in a narrow domain such as movie reviews or product reviews when labeled corpus is available for training (e.g. Titov and McDonald 2008). In a narrow domain, the vocabulary for subjective language is fairly limited and unambiguous, the sentiments are black and white, hence, the sentiment classification based on keyword density using even the simple learning algorithms can still achieve high precision (80% or above), as long as the input is not a short message. In websites like Amazon and Yelp, there is no lack of online reviews data which are already labeled by the users using either thumbs up and down icons or 5-star ratings. The increasing availability of such labeled data creates an ideal condition for using simple machine learning approach to sentiment classification.

However, all BOW sentiment classification faces a number of challenges.

The first is domain portability challenge. Traditional text classification approaches are domain dependent (Turney 2002). A sentiment classifier trained on movie reviews performs poorly on electronics reviews. In a commercial sentiment analysis system, social data are harvested from almost everywhere on the Web, including blogs, micro-blogs, instant messages, and various review sites and different domains, such as hotels, airlines, retailers, banking, automobile, foods, TV shows, and every other possible sector. With so many domains and verticals where a market researcher wants to apply sentiment analysis, training and maintaining numerous classifiers seem to be a daunting task, even if we assume that the data from all domains are available.

The second challenge comes from the length of a message. As mentioned before, the learning supported classification is good at classifying a document or a long review post, but the quality will compromise significantly once the incoming message is short (e.g. micro-blog posts from a mobile user). This is understandable as short messages do not provide sufficient number of data points (lexical evidence for sentiments) for an effective BOW system to work on. There is research on applying classification to short messages in an attempt to address this challenge (Yu and Hatzivassiloglou 2003; McDonald et al. 2007; Titov and McDonald 2008; Wilson et al. 2009), but most such study is either highly research-oriented or too domain-dependent, and hence can hardly be applied to a real world sentiment analysis scenario. As mobile platform for social media is getting popular, the social media world will soon be dominated by short messages. In fact, in terms of quantity and importance, micro-blogs are already dominating the social media, taking more than half of the entire social space while all the other thousands of social media sources together cannot match.

The third challenge comes from the unsatisfactory classification quality associated with the practical scenarios when big data becomes small. For example, almost all market researchers need the support for slicing and dicing the data for in-depth sentiment analysis from different angles, e.g. based on demographics, geo-locations, time, etc. In particular, representing sentiment

insights along the time dimension shows the trends of a brand and its history, which is particularly useful for tracking a brand's ups and downs in real time. However, when sliced and diced based on the users' needs, big data becomes small quickly, and a quality-challenged classifier is bound to have trouble with the users.

The fourth is the association challenge. Sentiments are not meaningful unless they are associated with an object, such as a brand or product. By nature, all BOW sentiment classifiers have trouble in associating with the target object (Davidov, Tsur and Rappoport 2010). Most such systems rely on co-occurrence and proximity heuristics for association. Since there is no structure or any relational understanding of the message in BOW-based systems, they are powerless with simple comparative expressions like 谷歌比雅虎强老鼻子了 (Google is a lot better than Yahoo). It happens that social media is abundant in comments involving multiple brands for contrast.

Finally, insights from sentiment classification only provide an overview of the sentiments, they are not actionable insights. Deep insights for sentiment analysis need to uncover the reasons behind the sentiments to answer business questions such as *why* customers like or dislike a product. Decoding the underlying reasons is a critical need for actionable insights in customer sentiment system as businesses can take actions with such insights.

The implementation and deployment of our sentiment analysis system in the customer insight platform is designed to address the above challenges, taking the approach of deep sentiment extraction based on natural language parsing. Our system is multi-lingual (currently, English, Chinese, Japanese, Spanish, German, Portuguese and Italian), but for this paper, we limit our study to Chinese.

Shallow sentiment analysis on Chinese has seen growing interest in the research community (Tsou 2005, 朱嫣岚 et al. 2006, 王素格 et al. 2009, 党蕾 et al. 2010, 赵妍妍 et al. 2010). Yang and Hou (2012) reports an impressive rule-based sentiment classification engine for Chinese. The unique output of this system is classification, coupled with a measure of confidence. However, unlike the confidence measure generated by the probability from a statistics-based learning system, this artificial confidence measure is calculated by using the manual encoding in the sentiment lexicon. This design makes it difficult to manage the consistency for both lexicon coding and the annotation of a gold standard for sentiment classification. Like most other Chinese sentiment systems, this system remains a lab prototype, and there is a long way before it can be deployed to the real world.

In summary, the major limitation of the sentiment work reported in the literature lies in three aspects: (i) coarse-grained, dominated by classification; (ii) shallow, only using lexicons and/or immediate context, with little or no support from parsing; (iii) mostly confined to labs. Our systems has addressed all these three issues.

Concept of Deep Sentiments

The concept of deep, fine-grained sentiments is our proposal in contrast to the dominant practice of shallow, course-grained sentiment analysis, i.e. thumbs-up and thumbs-down classification, coupled with sentiment association based on proximity. This concept is inspired by the needs from the field that call for more actionable insights and the answers to the why questions with regards to customer sentiments. Over time, the deep sentiments evolve in the process of engaging the analysts of customer insights and become mature as a standard to drive the research and development supported by parsing. To shed some light in the process, a deep sentiment system should be able to extract insights that can answer these questions in addition to sentiment classification:

- Which brand is this sentiment about? (association insight)
- Who made the sentiment comment? (customer background insight)
- How intense is the sentiment? (passion intensity insight)
- Can the system associate sentiments not only with a brand such as iPhone, but also with a feature of the brand, say, screen? (granularity insight)
- In addition to sentiments related to emotions of a customer (love/hate/happy/annoyed etc.), can the system identify the agent's positive or negative evaluations of the brand (e.g. cost-effective/poorly-designed)? (customer evaluation insight)
- How about the agents' needs or wish-list for brands? (market needs insight)
- How about agents' positive or negative action towards a brand (including consumers' purchase intent such as “will buy”; negative actions such as “abandon”, “stop using”)? (customer action insights)
- What are pros of a brand, including specs, features, functionality (designed to do what)? (pros insights)
- What are the cons of a brand, including weakness, loopholes and issues? (cons insight)
- Can the system identify comparisons between brands (iPhone is better than Blackberry)? (competition insight)
- Most important of all, what is the reason of the sentiment? (why insight)

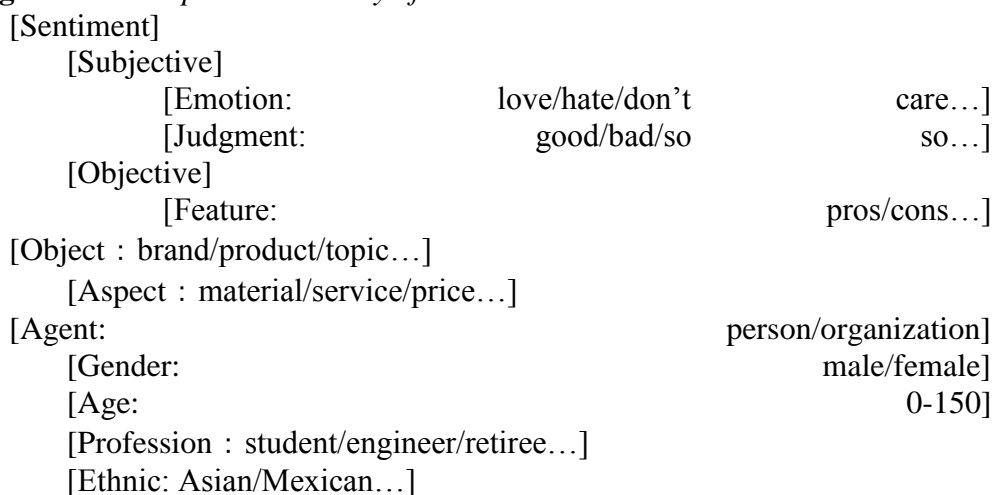
Systems that can answer these questions provide invaluable actionable insights to businesses. For example, it is much more insightful to know that consumers love the web browsing speed of iPhone 5 but are very annoyed by the lack of support to flash. This is an actionable insight, one that a company could use to redirect resources to drive the product's further development.

Extraction of such insights is enabled in our parsing-supported customer insight system.

A system that can answer most or all of the above questions is what we call a deep sentiment system. Apparently, some of the insights listed above, especially the actionable insight *why*, are beyond what can be attempted using the existing BOW models because it requires considerable natural language understanding to decode such elaborate and fine-grained insights. A full parser is called for to facilitate the deep sentiment extraction in order to meet the business users' real world needs.

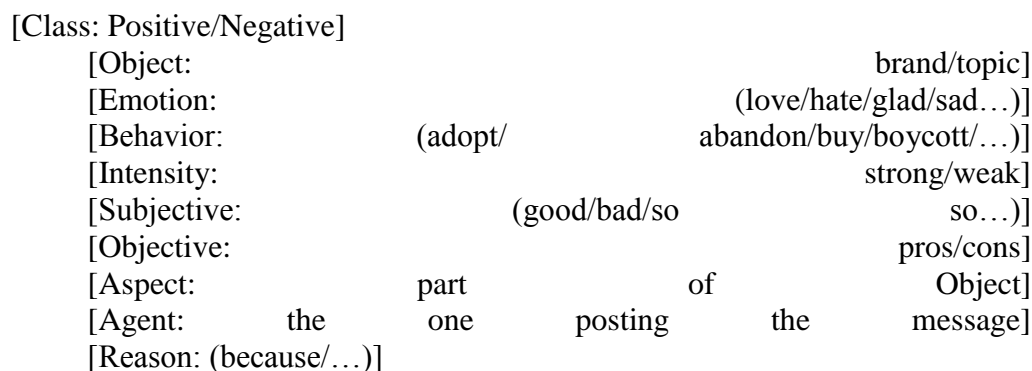
The hierarchy behind the deep sentiments is conceptualized in Fig.1.

Figure 1. Conceptual Hierarchy of Sentiments



Based on the concept hierarchy in Fig. 1, the sentiment frame in Fig. 2 has been defined as the target for our extraction engine in this stage. The agent background insight is left for future development.

Figure 2. Sentiment Frame Definition



Customer Insight Mining

The entire system involves the conceptual design of two-subsystems and four levels of processing. The two subsystems are often referred to as the front end search app and the backend indexing engine.

The four levels are:

- linguistic level
- extraction level
- mining level
- app level

These four levels of two (sub-)systems basically represent a bottom-up support relationship: 1 ==> 2 ==> 3 ==> 4. Clearly, the core engine of NLP (namely, a parser) sits in the first layer as an enabling technology, and at the 4th layer lies our customer insight application.

The back-end system is composed of three components: data acquisition involving Extraction, Transformation and Loading (ETL), NoSQL databases, and distributed indexing. Pooled from different sources in different formats in real time or non-real-time, the data is uploaded by the ETL into the distributed Cassandra NoSQL database. The uploading process involves language detection, spam filtering, web page extraction and de-duplication. The data integration goes through distributed processing after the data is loaded in the NoSQL database, using the open source text indexing engine Apache Lucene (lucene.apache.org) based on the Map-Reduce Framework. The NoSQL databases and the distributed indexing engines are configured in the computing cloud in order to ensure the elasticity of the entire backend system. A large social media data archive accumulated since one year ago (about 30 billion documents across 40 languages) can be indexed within seven days by using about 30 Cassandra database servers and 150 indexing servers. The frontend system is a SaaS-based application very similar to a search engine. Users interact with the apps through the browser and the application server. The application servers and the query node cluster communicate via multicast based on JGroups. The users' queries are broadcasted to the query node cluster for distributed search. The search results go through a process of integration, ranking and necessary transformation in the application server before they are presented to users using our configurable dashboard with a number of visualization widgets in the app. One year archive of social media generates about 25 terabytes of index files stored evenly in about 170 query servers using SSDs (solid state drives). In most cases, the user can see their search results within 2-3 seconds for all the required data reports in the app platform.

The NLP backend engine is a two-component system. It forms a highly modular processing pipeline, from shallow levels of linguistic processing to deep levels of sentiment extraction, using an expanded version of the seasoned NLP-oriented formalism named Finite State Automata (FSA, a high-level formal language that supports the encoding of finite state grammars for rule-

based NLP, Roche and Schabes 1997). Similar formalism and platform for NLP and information extraction include Silberztein (1999), Hobbs (1993) and Srihari et al. (2006).

The first component is a dependency parser based on basic phrase structures generated by chunking. The parser outputs a dependency-based hybrid tree structure involving basic phrases (Li 1989), representing the system's understanding of each incoming sentence. The hybrid tree is a system-internal, linguistic representation, similar to diagramming taught in grammar school. The system parses text in a number of passes (modules), starting from shallow levels and moving on to deeper levels. The tree output provides a logic-semantic basis for the subsequent extraction component for more generalized coverage of the sentiment phenomena.

The extractor component sits on top of the parser and outputs a table that directly meets the sentiment needs of products. This is where extraction rules, based on sub-tree matching, set to apply, for deep sentiment extraction. Considering the combination possibilities of surface structures, extraction rules built at logical level, on top of parsing tree structures, are often as powerful as hundreds of, or even thousands of, low-level linguistic rules in terms of covering the relevant surface patterns of language expressions for sentiments. The walk-through in Fig.3 and Fig.4 shows the process of how our customer insight system parses a social media Chinese post and generates the deep sentiments.

Figure 3. Parsing Illustration

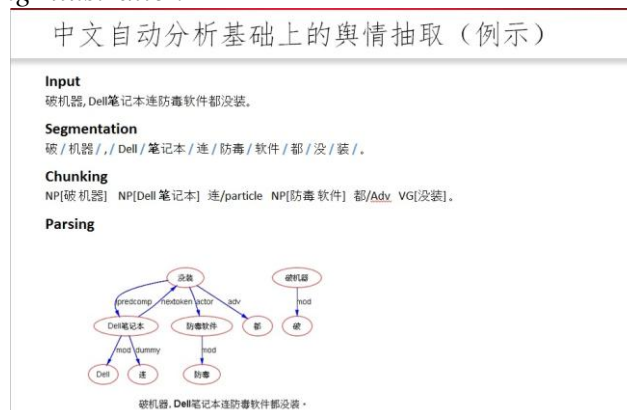
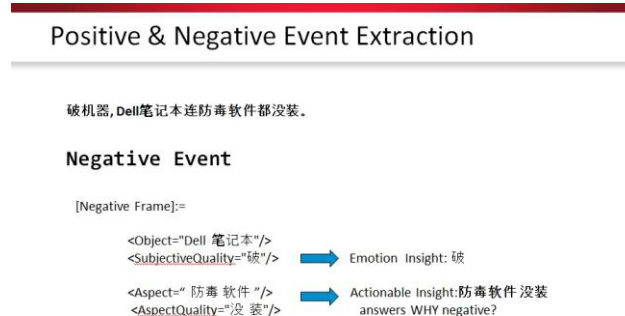


Figure 4. Sentiment Extraction Illustration



Automatic Survey Experiments

The customer insight mining system supports automatic survey of any brands from social media. This section illustrates some of the experimental results.

The content sources of one-year archive for the experiments consist of three hundred and fifty million posts or documents in simplified Chinese script, in which about one hundred million posts come from Baidu (Tieba, etc.), more than twenty million from Sohu, twenty-five million posts from the Tianya Forum. Unfortunately, there is only limited data from the most influential micro-blogging channel Weibo at this point (about one month data). But for hot topics, the current content sources seem to be sufficient to represent the general sentiments in social media.

In contrast to traditional manual survey, automatic survey demonstrates the following benefits: (i) speed; (ii) low-cost; (iii) objectivity, and (iv) multi-brand comparison. We will explain them one by one with illustrations.

First, manual survey takes weeks to complete while automatic survey gets results instantly within seconds.

Second, low-cost: considerable time and money are required before proceeding with a sizable meaningful manual survey for one brand while automatic survey only requires an annual license to access results for any number of brands any number of times in a year.

Third, objectivity: automatic survey is based on big data of the brand comments spontaneously posted by the customers while manual survey involves much smaller number of data points;

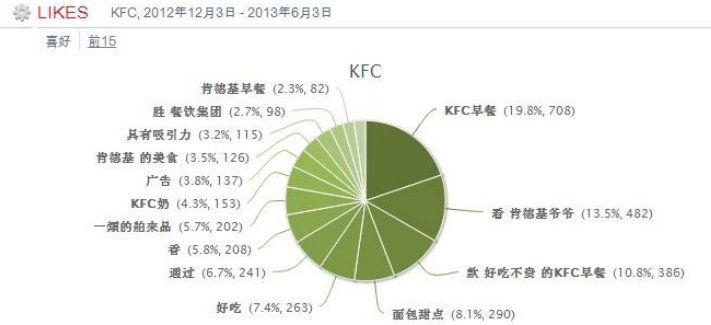
Finally, multi-brand comparison: it is an easily implemented, built-in functionality of our customer insight system to allow for multiple brands in one industry to be automatically surveyed and compared at the same time while it is almost impossible for the manual survey to do so due to the increasing manual costs. The Brand Passion Index illustration below surveys the relative standing of the four well-known international brands in the fast food industry: McDonald's, KFC, Pizza Hut and Yoshinoya, with Pizza Hut on the top and KFC most commented (biggest bubble).

Figure 5. Brand Passion Index for International Fast Food Brands



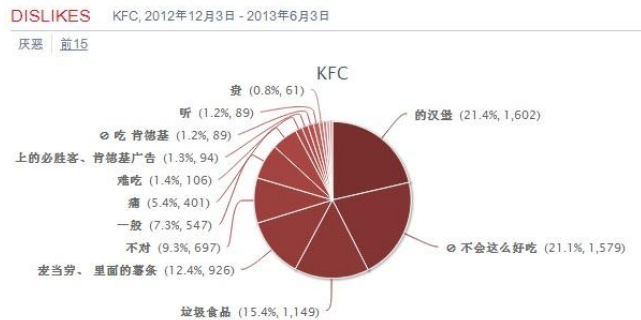
A follow-up experiment is performed in search of reasons behind the likes and dislikes of KFC in China, using one month of the Weibo data (April-May 2013) with 1,860,000 posts. The results are shown in Fig. 6 and Fig. 7.

Figure 6. Why like KFC



As shown in Fig.6, among the top 15 reasons why people like KFC, the delicious and inexpensive KFC breakfast is at the top (32.9%=19.8%+10.8%+2.3%), followed by “go watch KFC Grandpa” (13.5%).

Figure 7. Why dislike KFC



As for the dislikes in Fig. 7, the top three reasons are KFC hamburger (21.4%) (no specific criticism on its flagship chickens though), not delicious (22.5%=21.1%+1.4%), and junk food (15.4%).

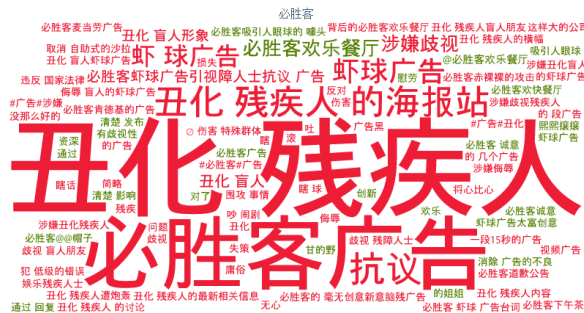
Fig. 8 shows the related sample Weibo data. They are typical short Microblog messages including several comparative expressions mentioning two brands, such as “麦当劳还是不如肯德基好吃” (*McDonals is after all not as tasty as KFC*). It is impressive to see the parsing supported sentiment system get right these cases which are known to be a challenge to the BOW models.

Figure 8. Sample Posts of Social Coverage of KFC



We give a recent survey of Pizza Hut in China for public sentiments in the use case of risk management.

Figure 9. Word Cloud for Negative Comments against Pizza Hut Ad



The word cloud in Fig. 9, which is derived at one point of June 2013 based on 1-week data, is dominated by anger and complaints in red fonts (our visualization uses red for negative and green for positive). It demonstrates strong negative sentiments accusing Pizza Hut's commercial of hurting feelings of the vision impaired. The online ad tried to introduce a shrimp ball, Pizza Hut's latest product, by making a pun. In Chinese, "shrimp" and "blindness" share the same sound. The advertisement portrayed a shrimp, holding a walking stick, rolling and forming a ball as the Chinese character for "blindness" appears on screen. This incident triggered huge criticism online with protesters' gathering in many locations. This is the biggest PR crisis for Pizza Hut China in history. As shown, the public survey system can help monitor and catch events like this to alert the businesses before the crisis runs wild. It is later reported that the Pizza Hut management quickly withdrew the ad and issued a public apology. Further mining shows that the effort for the apology has taken effect. Despite a sudden jump of the negative event in social media everywhere for a short time, the sentiments calmed down quickly. On

our radar are 86,000 Weibo posts in June when the event happened, there are only 3100 negative posts about this event, about 28% of the entire Weibo pool of the month, mostly concentrated in one week span.

The following automatic survey of the mobile phone market in China is a bit surprising (Fig. 10). We expected the divide between the survey brands iPhone, HTC, Samsung, Nokia and Xiaomi to be much larger than it is reflected in the following BPI, with iPhone assumably at the top and some local brands such as Xiaomi (“小米”) to be at the bottom. In fact, they all cluster together at the dislike corner of the graph.

Figure 10. Brand Passion Index for Mobile Industry



Despite bigger buzz of the mobile king Apple due to more comments on iPhone (with 1,600,000 data points), iPhone shows no significant superiority over the other brands (with 400,000+ data points) in terms of both net sentiment and passion intensity. It seems that consumers have more negative comments than positive for all these brands, and the sentiment intensity is not strong. It seems to indicate that the mobile market in China still has no clear leader and the competition is just started. This result is believed to be trustworthy as the number of data points (400,000 to 1,600,000) in one year supporting this automatic survey is more than 2 magnitude higher than a typical manual survey (usually with only several thousand data points).

To benchmark the sentiment quality objectively, we use anonymous human judges in the CrowdFlower service to perform the quality assurance (QA). Since our system is mainly developed for businesses to gather consumer sentiments about brands, QA usually selects 10-20 representative brands from various industries as queries. Every result is submitted to at least four native speaker-judges through CrowdFlower and at least 75% inter-judge (i.e. three out of four judges) agreement is used to count as gold standard. One recent QA uses the following 15 brands as queries:

*iPhone | 中国电信 | 丰田 | 伊利 | 南航 | 可口可乐 | 宝马 | 家乐福
| 必胜客 | 携程 | 淘宝 | 苏宁 | 茅台 | 蒙牛 | 麦当劳*

Fig. 11 shows the precision results of the last release versus the present release in comparison. Fig. 12 benchmarks the sentiment coverage with a measure called relative recall.

Figure 11. *Precision Change between Releases*

Sentiment	Release 1	Release 2
Positive	81.4%	73.2%
Negative	89.8%	89.1%
Overall	85.7%	79.6%

Figure 12. *Relative Recall Change*

	Release 1	Release 2	Change
Pos	2.8%	14.5%	417.1%
Neg	1.5%	5.8%	298.7%

In terms of balancing precision and recall, we set the bottom line for precision to be 80% as this is what most customers expect. Once it reaches 80% precision, the system is set to maximize recall. Fig. 11 shows that the previous release was more conservative, with precision as high as 85.7%, so we made some attempt for enhancing recall, by adding a few condition-loosened default rules. As a result, the precision is reduced to 79.6%, just reaching our target, but the recall has improved significantly, by 298.7% - 417.1%. The positive relative recall reaches 14.5%, and the negative relative recall turns out to be 5.8%, indicating that this adjustment is well justified.

Here the relative recall is not the recall as used in the community, but its indirect reflection. Since sentiments are fairly sparse in randomly collected running text, which also involves considerable amount of gray phenomena, it is not economically practical to annotate a large running text corpus, nor is it easy to maintain it. We turn to a relative coverage measure (relative recall) to track the progress in the coverage of the system. Relative recall is a reflection of the extracted sentiment concentration.

Figure 9 shows the relative recall which looks fairly low, but is it comparable to that of our English relative recall (6% -14%), so the recall is decent. English is the language we have developed and put to use for the longest time, having gone through numerous development iterations and debugging, with both precision and recall meeting the customer satisfaction and product requirements. On the other hand, we also need to keep in mind the difference between English and Chinese in terms of real life sentiment concentration. Statistics from lexical perspective shows that with the basic vocabulary of around 45,000-50,000 entries in English and Chinese, the number of the Chinese positive entries is three times the number of the English positive ones, and the number of the Chinese negative entries is more than

twice that of English. This seems to indicate that Chinese text may contain sentiments significantly more than that of English. Therefore, there is still considerable room for recall improvement in the on-going development.

References

- G. A. Bruder and W. M. Janyce 1990. A psychological test of an algorithm for recognizing subjectivity in narrative text. In Proceedings of 12th Annual Conference of the Cognitive Science Society, 1990: 947-952
- N. Chinchor and E. Marsh 1998. MUC-7 information extraction task definition (version 5.1). Proceedings of MUC-7
- D. Davidov, O. Tsur and A. Rappoport 2010. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. In Proceedings of COLING 2010: Poster Volume, pages 241–249, Beijing, August 2010
- J. R. Hobbs 1993. FASTUS: A system for extracting information from text. Proceedings of the DARPA workshop on Human Language Technology, pp. 133–137. Princeton, NJ.
- A. Hogenboom, D. Ball, F. Frasinca¹, M. Ball, F. Jong, U. Kaymak 2013. “Exploiting Emoticons in Sentiment Analysis”, Proceedings of the 28th Annual ACM Symposium on Applied Computing: 703-710
- W. Li 1989. A Dependency Syntax of Contemporary Chinese, BSO/DLT Research Report, the Netherlands.
- B. Pang, L. Lee and S. Vaithyanathan 2002. Thumbs up? Sentiment classification using machine learning techniques. In EMNLP 2002, 79-86.
- B. Pang and L. Lee 2008. Opinion mining and sentiment analysis. In Foundation and Trends in Information Retrieval. Volume 2, Issue 1-2, 1–135.
- E. Roche and Y. Schabes 1997. Finite-State Language Processing. MIT Press.
- M. Silberztein 1999. INTEX: a Finite State Transducer toolbox. Theoretical Computer Science 231(1). Elsevier Science.
- R. Srihari, W. Li, T.Cornell and C. Niu 2006. InfoXtract: A Customizable Intermediate Level Information Extraction Engine. Journal of Natural Language Engineering, 12(4), 1-37
- M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede 2011. “Lexicon-Based Methods for Sentiment Analysis,” Computational Linguistics, 2011,37(2): 267-307
- I. Titov and R. McDonald 2008. Modeling online reviews with multi-grain topic models. In WWW, pages 111–120, New York, NY, USA. ACM.
- P. D. Turney 2002, “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews,” Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. USA.2002:417-424.
- T. Wilson, J. Wiebe and P. Hoffmann 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics, 35(3):399–433
- J. Yang and M. Hou 2012. Sentence Sentiment Orientation Calculation Based on Rules. Proceedings of the 13th Chinese Lexical Semantics Workshop (CLSW2012), Wuhan, China. 2012.
- H. Yu and V. Hatzivassiloglou 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. EMNLP

- Y. Zhao, B. Qin, W. Che and T. Liu 2010. Sentiment Unit Identification Based on Parsing. Spoftware Journal. 21(8):1834-1848.
- 朱嫣岚,闵锦,周雅倩,黄萱菁,吴立德 2006,
“基于HowNet的词汇语义倾向计算,”中文信息学报, 2006,20(1):14-20.
- 王素格, 杨安娜, 李德玉 2009,
“基于汉语情感词表的句子情感倾向分类研究,”计算机工程与应用,
2009,45(24):153-155
- 党蕾,张蕾 2010, “一种基于知网的中文句子情感倾向判别方法,”计算机应用研究,
2010,27(4):1370-1372
- 赵妍妍,秦兵,车万翔,刘挺 2010, “基于句法路径的情感评价单元识别,”软件学报,
2010,21(8):1834-1848