

# The Comparative Analysis between Sensory and Cognitive in Online Customer Reviews of Baby Formula

Yu Bai, Guiping Zhang, Peiyan Wang, Jianjun Chen,

Tianhao Zhang, Fang Cai, Peng Lian, Liang Liu

**Abstract**—Based on a big data service platform for industry informatization, we made a comparative analysis between sensory and cognitive in online customer reviews of baby formula. Three aspects of the customer reviews are analyzed. In the first part, the descriptive analysis was conducted to evaluate the customer reviews propensity between sensory and cognitive; in second part of analysis, we made the comparison of the reviews propensity among buyers who have different frequency of consumption; Finally, the comparison of the propensity between buyers and sellers was made according to the base data from reviews and selling point index. According to this study, product brochures emphasize on the cognitive side of the infant formula milk powder, which is rather less intuitive and perceivable for average buyers than the sensory side. This analysis can guide the changing of promotion strategy and the design of the brochure.

**Index Terms**—Big Data Service, Comparative Analysis, Sensory, Cognitive, Online Customer Reviews

## I. INTRODUCTION

THE role of psychology in consumer buying behavior is essential because it affects the attitude of customers through feeling, emotion desire and response [1]. A brand is successful when it matches human characteristics, lifestyles and preferences [2]. As understood, sensory involved five senses of human namely feel, touch, taste, sight and hear. For cognitive, it is commonly understood as thinking ability.

Manuscript received July 10, 2014; revised July 29, 2014. This work was supported in part by National Key Technology R&D Program of China (2012BAH14F00).

Yu Bai is with the Knowledge Engineering Research Center, SAU, Shenyang, 110136, China, and with the College of Computer Science and Technology, NUAA, Nanjing, 210016, China (Phone: 86-024-89726878; e-mail: baiyu@sau.edu.cn).

Guiping Zhang (e-mail: zgp@ge-soft.com) is with the Knowledge Engineering Research Center, SAU, and the Shenyang Bit power Co., Ltd, Shenyang, China

Peiyan Wang (e-mail: wangpy\_kerc@163.com) is with the Knowledge Engineering Research Center, SAU, Shenyang 110136 China.

Jianjun Chen (e-mail: chenjj4003@163.com) is with the Shenyang Bit power Co., Ltd, Shenyang, China.

Tianhao Zhang (e-mail: tianhao.zhang.2016@gmail.com) and Fang Cai (e-mail: fangcai@berkeley.edu) are with the EECS, UC Berkeley, Berkeley, USA.

Peng Lian (e-mail: lianpeng7909@163.com) and Liang Liu (e-mail: liuliang8123@163.com) are with the Shenyang Bit power Co., Ltd, Shenyang, China.

Previous researches [3,4] proved that there is a significant relationship of human senses towards shopping behavior. However, little attempts have been made to compare the sensory with cognitive in behavior of online shopping.

Nowadays, retail success is no longer all about physical stores. Online shopping or e-shopping as a form of electronic commerce, allows consumers to directly buy goods or services from a seller over the Internet using a web browser, and it appeals to a wide set of consumers. The largest of these online retailing corporations are Alibaba, Amazon.com and eBay.[5] Taobao [6] is a Chinese website for online shopping similar to eBay and Amazon that is operated in China by Alibaba Group. As well as Tmall.com [7] offers a shopping experience for increasingly affluent Chinese consumers. Tmall.com and Taobao Marketplace set a record for highest single-day transaction volume during a special promotion on November 11, 2012, facilitating the sales of goods totaling RMB19.1 billion on the day. For the year ended March 31, 2013, the company claimed that combined gross merchandise volume (GMV) of the two platforms exceeded RMB1 trillion.[8] At the same time, those two platforms also produce massive amounts of data, include trading records, and especially the customer data like reviews of commodities.

With its large population, China is the world's biggest consumer of baby formula, as revealed by one study of Euromonitor, a consumer research agency. The baby formula, with its broad representation of everyday consumer products, has a large number of active online shoppers. In this study, a large number of customer reviews about baby formula product from Taobao and Tmall.com were used for the comparative analysis between sensory and cognitive in online shopping.

This paper is organized as follows: In the second section, we give the methodology on data analyzing, and illustrate how we use online customer reviews to perform the comparative analysis between sensory and cognitive on online shopping behavior of baby formula buyers. The observed data will be discussed in details in the third section. And the related works will be described in the fourth section. Finally, the conclusions of this study will be made.

## II. METHODOLOGY

In this section, we will illustrate how we analyzed the data of customer reviews which reflect online buying behavior.

Following steps give a sketch of the process of data access and analysis.

- Step1. Data access
- Step2. Clustering on Customers reviews data set
- Step3. Reviews factors extraction (extract topic words from each cluster)
- Step4. Mapping Reviews factors to psychology dimensions
- Step5. Knowledge discovery from data

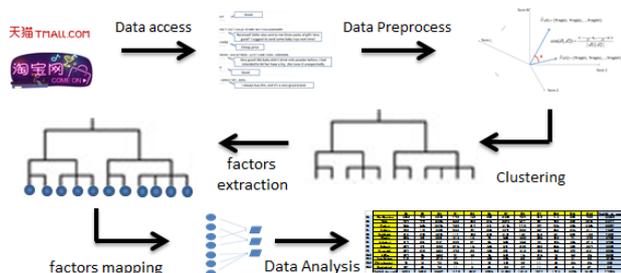


Fig. 1. A sketch of the data access and analysis process.

### A. Data Access

In order to obtain authentic data, this paper chose a direct approach which utilizes a data pipeline to access public customer reviews data from the online retailing corporations. Furthermore, the product brochures of 12 brands of Infant formula milk powder were collected to understand the present positioning of these brands. Each brand was represented by a corresponding element in the set of  $M = \{M1, M2, \dots, Mx\}$ ,

with  $x$  equal to the number of brands of interest.

The customer reviews of these 12 brands of infant formula milk powder were collected from Taobao.com and Tmall.com. All reviews were made during January 2014 to June 2014, with a total of **79516** reviews.

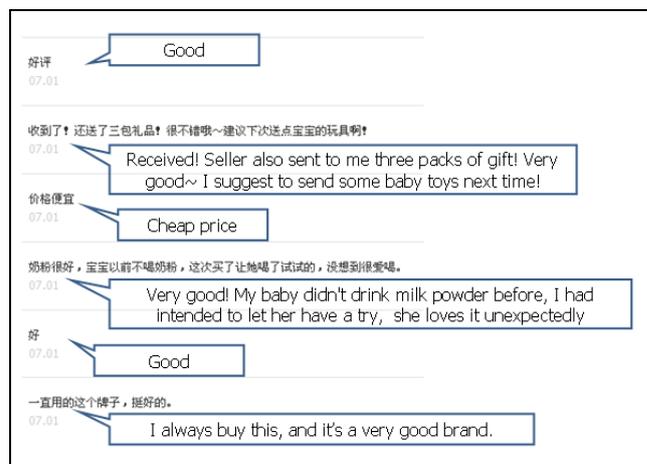


Fig. 2. Examples of online customer reviews data.

### B. Data Preprocess

Since customer reviews and product brochures are in natural language, the data in the data layer is not yet ready for analysis. NLP techniques, Such as Chinese word segmentation, feature selection and expansion are essential in feature representing for those unstructured texts.

TABLE I  
 TOP REVIEWS AND SELLING POINT INDEX

	Categories	Reviews factors	Brochures
R1	包装(Packaging)	灌(tin), 桶(barrel), 袋(bag), 破损(broken), 完好(intact)	×
R2	物流(delivery speed)	快递(Express), 配送(distribute)	×
R3	价格(Price)	便宜(cheap), 实惠(affordable), 划算(cost-effective), 不贵(inexpensive)	×
R4	赠品(Gifts for sales promotion)	优惠券(coupons), 促销活动(Sales Promotion)	×
R5	品质(Quality of Presentation)	保质期(warranty), 新鲜(fresh)	√
R6	味道(Taste)	口感(Taste), 口味(Taste)	√
R7	推荐(Advice)	习惯(custom), 广告(ad), 口碑(word of mouth)	×
R8	品牌(Brand)	牌子(Brand), 名牌(Brand)	×
R9	安全(Scurity)	放心(secure), 信赖(reliable), 安全(safe), 安心(secure), 可靠(reliable)	√
R10	反应(ADRs)	腹泻(diarrhea), 便秘(constipation), 过敏(allergies), 呕吐(vomiting), 湿疹(eczema)	√
R11	产地(Location)	国产(Domestic), 进口(imported), 原装(original)	√
R12	发育(Physiogeny)	健康(Healthy), 抵抗力(resistance), 体质(physical) 智力(intelligence), 体重(weight)	√
R13	营养(Nutrient)	配方(Recipes), 成分(ingredients), 配料(ingredients), DHA, ARA	√

TABLE II  
THE QUANTITY OF CUSTOMER REVIEWS OF 12 BABY FORMULA BRANDS FROM THE ONLINE RETAILING CORPORATIONS TAOBAO.COM & TMALL.COM.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	Total in row	
R1	Packaging	1502	973	3812	795	418	502	2180	929	840	441	288	958	13638
R2	Vol.	789	999	2366	632	335	246	2271	677	694	925	532	1011	11477
R3	Price	666	636	2396	925	249	314	1496	770	597	736	386	1126	10297
R4	Gifts	595	448	1369	833	194	251	1397	852	617	513	246	344	7659
R5	Quality	573	392	1891	447	105	210	1437	467	645	424	230	438	7259
R6	Taste	686	187	1054	141	43	91	651	135	526	135	94	190	3933
R7	Advice	549	264	817	238	87	66	681	162	404	196	186	271	3921
R8	Brand	289	194	880	276	46	121	694	210	235	266	102	239	3552
R9	Scurity	289	240	603	195	28	101	504	142	155	165	170	214	2806
R10	ADRs	309	66	602	84	12	16	613	46	346	43	60	56	2253
R11	Location	740	39	273	37	4	101	366	22	30	28	14	82	1736
R12	Physiogenv	93	43	93	86	4	15	112	18	44	23	26	26	583
R13	Nutrient	127	23	92	44	3	23	85	19	44	25	2	27	514
	Total in col	6751	6672	16677	4470	2117	2036	14213	4245	7377	4971	3566	6421	79516

Cluster analysis

In data mining, hierarchical clustering, a method of cluster analysis, seeks to build a hierarchy of clusters. In this paper, an optimal agglomerative method [9] was used to cluster the reviews set. Here, the metrics for hierarchical clustering is cosine similarity of the word vectors. Finally, all reviews were clustered into 13 categories.

Reviews factors extraction

A Mutual Information Based Feature Selection method, called minimal-redundancy-maximal-relevance (mRMR) [10] was used to extract topic words from each cluster as the cluster’s reviews factors. And the selling point words were also extracted from product brochures. Combined with the reviews factors and selling points, an index was made, as listed in table1. Table 2 shows the quantity of customer reviews in details.

TABLE III  
DIMENSIONS FOR CUSTOMER REVIEWS

Dimension	Measurement	Reviews factors
Sensory	Vision	包装 (Packaging)
		品质(Quality of Presentation)
		价格(Price)
		赠品(Gifts for sales promotion)
		物流(delivery speed)
Cognitive	Knowledge	味道(Taste)
		推荐(Advice)
		品牌(Brand) 安全(Scurity) 反应(ADRs) 产地 (Location) 发育 (Physiogeny) 营养 (Nutrient)

Dimensions

Customer review is a form of customer feedback on electronic commerce and online shopping sites. Since customer reviews include sensory and cognitive description of the brands from buyers, by analyzing customer reviews, we can reveal the influence of sensory and cognitive on online shopping behavior. Table3 shows the map of reviews factors from psychology dimensions.

III. OBSERVATION AND DISCUSSION

For the first part of the analysis, the descriptive analysis was conducted to evaluate the customer reviews propensity between sensory and cognitive. Figure 3 summarizes the descriptive statistic for distribution listed in the study.

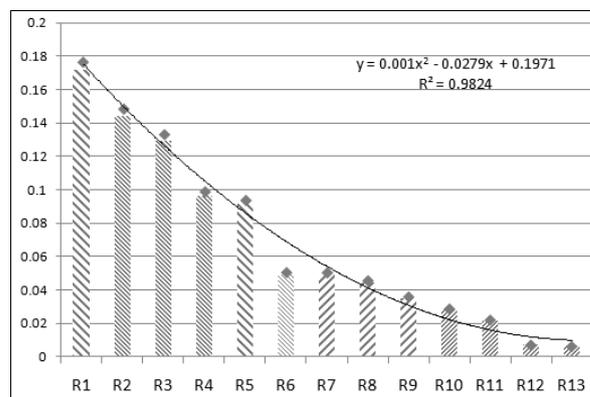


Fig. 3. Sensory and cognitive factors distribution in all reviews

As the data shows, the most frequent review type is Packaging (R1), which is under vision category, along with Quality of Presentation (R5). It is indeed generally accepted that people like to view beautiful things and hate to see bad things. Vision is the most powerful sense that enables human to observe the good and the bad of any situation. In buying behavior, vision is the major factor that triggers other senses to react.

As the second most frequent review type, the delivery speed (R2), Price (R3) and Gifts for sales promotion (R4) all come from the Feelings category. The fundamental difference between human and the machine is emotion. In human life, emotion plays an influential role especially in expressing their satisfaction and dissatisfaction. Fast delivery, reasonable prices and gifts from sellers can bring pleasant shopping experience for each of customers, and vice versa.

For the sense of taste and smell (R6), although closely linked to and communicated with the brain, the sense of taste and smell always take place after the vision and shopping

feelings for an online customer. This explains why the number of reviews about taste and smell is less than those of sensory factors.

Different from sensory, cognitive is defined as thinking. It refers to the mental structures and development involved in understanding memories. As information, meanings, and beliefs that human keep in one mind, cognitive will be processed to understand the stimuli or event, remembering the past, forming evaluations and making decisions purchase [11]. Cognitive relies on interpreting, thinking, decision making, analyzing, sensing and belief that will in turn transform into the experience and knowledge. It is a long and complicated process, and it's hard to customers who usually spend a short amount of time on reviews after receiving products. For this reason, all cognitive factors, regardless belief type, such as *advice (R7)*, *brand (R8)*, *security (R9)*, or knowledge type, such as *ADRs (R10)*, *location (R11)*, *Physiogeny (R12)*, *Nutrient (R13)*, got lower attention rates of customer reviews. And the attention rates of reviews in the knowledge type is even lower than the those of belief type.

In order to establish the validity of the analysis, we investigated sensory and cognitive factors distribution of the reviews for the 12 brands separately. Figure 3 shows the distribution of the factors of the reviews for each brand.

In Figure 4, for each brand, from M1 to M12, the nonlinear polynomial function was used to form a trend of the attention rates of customer reviews. As we see that, all the trend curves are decreased as with the difficulty of the feelings interpretation increased. For all brands, the attention rates of custom reviews in sensory aspect is significantly higher than that in the cognitive aspect.

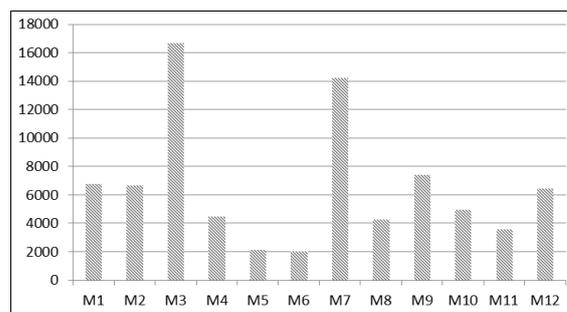


Fig. 5. The reviews number of all 12 brands

Moreover, figure 5 is a histogram of the number of reviews for the 12 brands. As shown in this graph, the M3 and M7 have more customs reviews than the others brands, so that the reviews for brand M3 and M7 are more representative of online shopping behavior of baby formula buyers. The review trend curves of M3 and M7 are consistent in shape, and also these two curves resemble the overall trend of sensory and cognitive factors are identical. This illustrates the consistency of the above analysis results.

For the next part of analysis, the comparison of the factor focus between different level buyers was made according to the base data from reviews point index. According to the explicit description in reviews, buyers are divided into three groups: (a) those buying thus brand for the first time, (b) those buying thus brand twice and (c) those buying thus brand at least three times.

As shown in figure 6, There are reviews trends of all 13 factors in 6 month (2014.01 – 2014.06) described by different

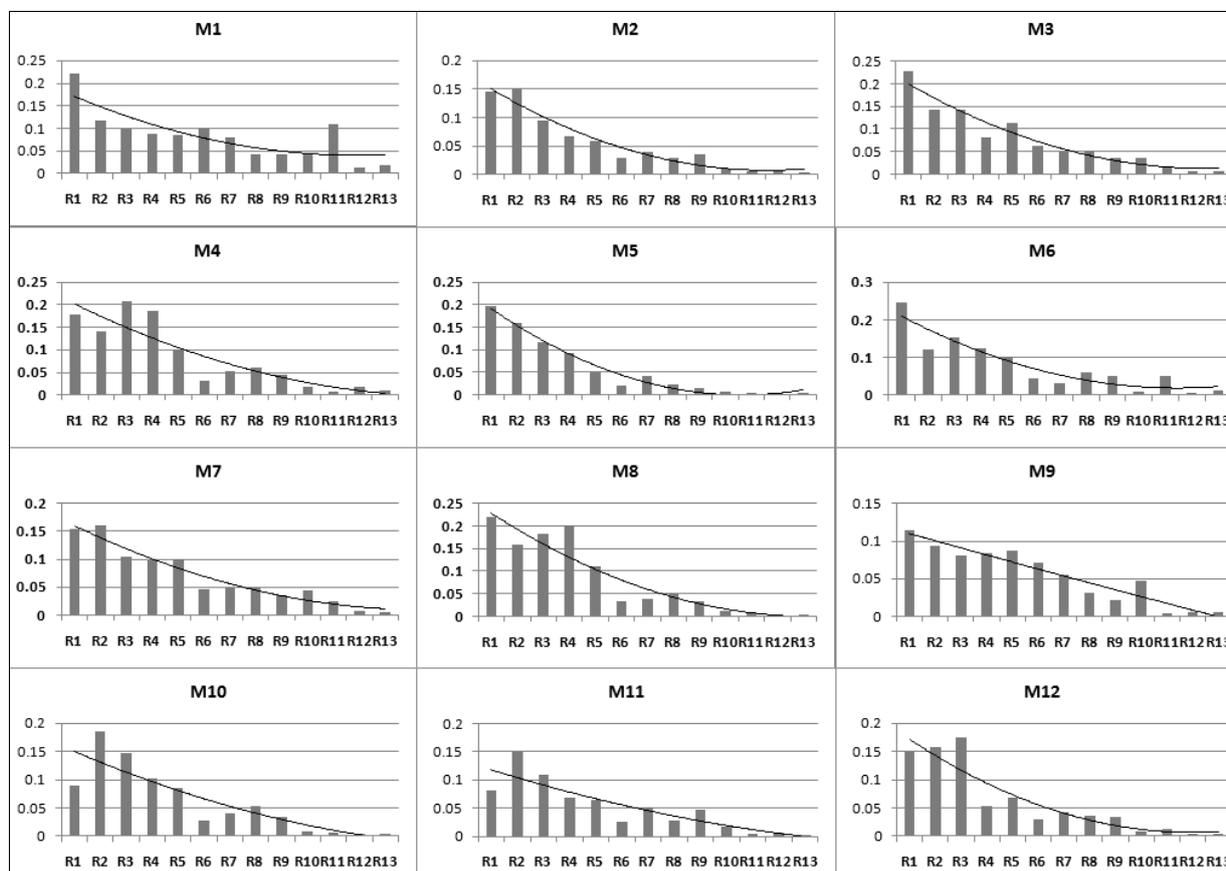


Fig. 4. Sensory and cognitive factors distribution in reviews of each brand.

level buyers. For each level of buyers, the trends in six month are basically the same. The customs reviews in sensory factors were mainly interpreted.

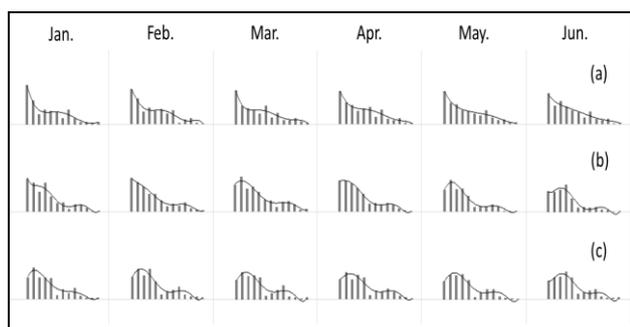


Fig. 6. The reviews trends of all 13 factors in 6 month (2014.01 – 2014.06) described by different level buyers. Customs who were buying thus brand (a) for the first time, (b) twice (c) at least three times.

For first-time buyers, as shown in figure 7, the number of reviews about sensory factors was more the cognitive ones. *Packaging (R1)* is the most frequent review type. In cognitive factors, reviews on *brand (R8)* receive more attention rates than others. One possible reason is that, for an ordinary person, without buying experience, brand dominates all other factors in purchase decision.

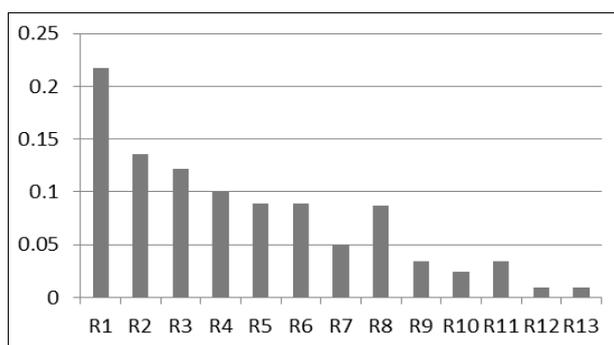


Fig. 7. The overall types distribution of reviews described by first level (buying thus brand for the first time) buyers.

For second-time buyers of this brand, as shown in figure 8, although the sensory factors still dominate customs reviews, the *Packaging (R1)* is not the most frequent review type (its attention rate decreased from 21% to 16%). While the attention rate of the *delivery speed (R2)*, *Price (R3)* and *Gifts for sales promotion (R4)* and *Quality of Presentation (R5)* increased by 4.7%, 2.8%, 6.1% and 1.5% respectively. As a belief factor, the *security (R9)* in cognitive category begins to be more involved (from 3.4% to 4.6%). The attention rate of *brand (R8)* decreased from 8.7% to 3.2%, and became lower than that of *advice (R7)* and *security (R9)*.

For the buyers buying this brand for at least three times, as shown in figure 9, the attention rates of the *Packaging (R1)* and *taste (R6)* declined to 12.5% and 2.6% respectively. While the attention rates of *Quality of Presentation (R5)* raised to 13.4%. As a belief factor, the *security (R9)* in cognitive category raised to 7.1%. Meanwhile, others factors remained at a stabilized level.

The comparison result of the reviews propensity among different types of buyers was shown in figure 10. It shows that

the emphasis on the proportion of sensory and cognitive in online shopping behavior remains in a stabilized high rate, and it was almost independent of purchase frequency.

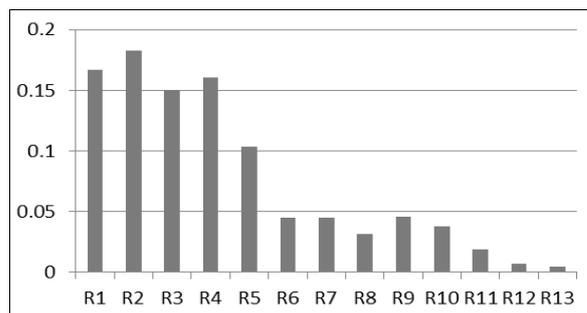


Fig. 8. The overall types distribution of reviews described by second level (buying thus brand twice) buyers.

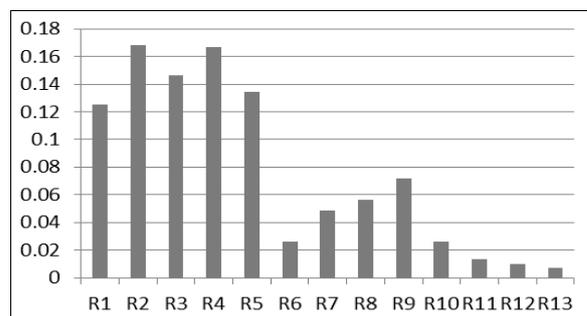


Fig. 9. The overall types distribution of reviews described by second level (buying thus brand at least three times) buyers.

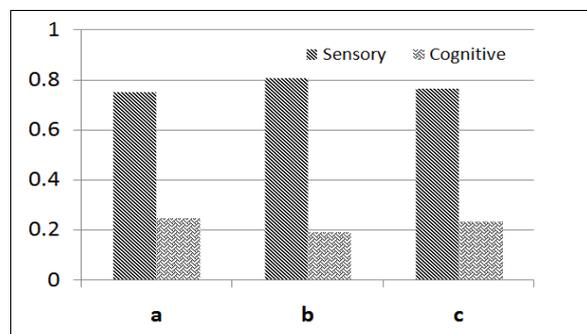


Fig. 10. The comparative between sensory and cognitive in behavior of buying baby formula online based on the reviews described by different level buyers.

For the third part of analysis, the comparison of the factor focus between buyers (reviews) and sellers (brochures) was made according to the base data from reviews and selling point index. As listed in table 2, the factor *R5*, *R6*, *R9*, *R10*, *R11*, *R12* and *R13* were related to both buyers and sellers.

Interestingly, as shown in figure 11, the reviews trend curve is similar to the reverse of the brochures trend curve. As explicit competitive factors, the *security (R9)*, *ADRs (R10)*, *location (R11)*, *Physiogeny (R12)* and *Nutrient (R13)* were contained in brochures of each brand generally. On the other hand, because of the perceptibility, the *Quality of Presentation (R5)* and *Taste (R6)* as the sensory factors should be the potential competitive factors. Which means that the easily expressed sensory factors are not emphasized by sellers, while the cognitive factors which sellers focused on, prove difficult to interpret for buyers.

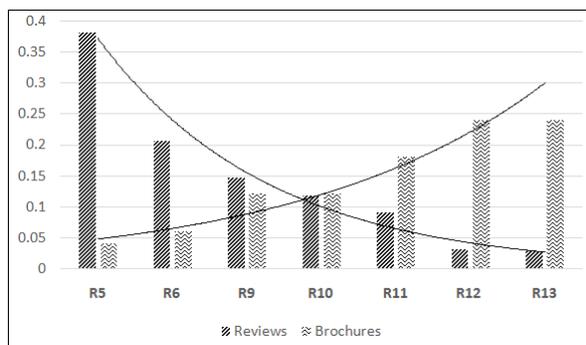


Fig. 11. The reviews number of all 12 brands

#### IV. RELATED WORKS

According to the research by MGI and McKinsey's Business Technology Office [12], the amount of data in our world has been exploding, big data is emerging as the world's newest resource for competition, underpinning new waves of productivity growth, innovation, and consumer surplus. Therefore, more and more analytics-driven organizations have recently emerged to help companies reduce costs, increase revenues and improve competitiveness.

In 2013, our group firstly proposed the concept of "bit power" [13]. The bit power is an integrated source of data, information and knowledge. Just as nuclei release large amount of energy through fusion reaction, the bit power generates huge energy urging industry development through connected data. It will be a primary energy of the next generation of business intelligence, and that will ensure the enterprise have the right information, at the right time, to make the right business decisions. And the data collection and analytical process in this paper were based on the bit power plant—a big data service platform for industry informatization [14, 15].

#### V. CONCLUSIONS

It is important to consider the consumer's motivation for buying products. The study on consumer behavior helps firms and organizations improve their marketing strategies by understanding issues such as how consumers think, feel, reason, and select between different alternatives. For this purpose, we compared the online buying behavior of baby formula on sensory and cognitive aspects through customers review analysis. Findings are as follows:

First, for all brands, the attention rate trend curve goes down as moving in direction from sensory to cognitive factors. The attention rate of custom reviews in sensory aspect is significantly higher than that in the cognitive aspect. This analysis result shows that people tend to make reviews on the aspects that can be easily perceived.

Furthermore, the comparison result among different types of buyer shows that the proportion of sensory reviews is always far more than cognitive reviews. For baby formula, this phenomenon is independent of buyer's purchase frequency.

Finally, the comparison of the factor focus between buyers and sellers shows that the reviews trend curve is the reverse of the brochures trend curve. This indicated that, the descriptive sensory factors are not emphasized by sellers; However, the cognitive factors, which sellers focus on, prove difficult to interpret for buyers. With respect to brochures, the cognitive

factors constitute explicit competitive elements among brands. On the other hand, because of the perceptibility, the sensory factors should be the potential competitive elements for sellers, and these elements will have significant influence on baby formula brand promotions.

In conclusion, the future of online marketing begins with predictive analytics on big data. According to the findings of this study, the emphases in current product brochures of infant formula milk powder are more cognitive, but it is not immediately understandable for average buyers. Therefore, we suggest that the baby formula sellers make changes on promotion strategy and use more sensory product brochures.

#### REFERENCES

- [1] N.M. Puccinelli, R.C. Goodstein, and D. Grewal, et al, "Customer Experience Management in Retailing: Understanding the Buying Process," *Journal of Retailing*, Vol 85, Mar, 2009. pp. 15-30
- [2] M.R. Jamaluddin, M.H. Hanafiah, and M.I. Zulkifly, "Customer-Based Psychology Branding," *In Proceedings of Asia Pacific International Conference on Environment-Behaviour Studies*, London, UK, Sep. 2013. pp. 772-780
- [3] T. Wagner, T Rudolph, "Towards a hierarchical theory of shopping motivation," *Journal of Retailing and Consumer Services*, Vol. 17, Sep, 2010. pp. 415-429
- [4] H. Kim, K.S. Suh, and U.K. Lee, "Effects of collaborative online shopping on shopping experience through social and relational perspectives," *Information & Management*, Vol. 50, Jun, 2013. pp. 169-180
- [5] The Economics. (March 23, 2013). The Alibaba phenomenon [Online]. Available: <http://www.economist.com/news/leaders/21573981-chinas-e-commerce-giant-could-generate-enormous-wealthprovided-country-s-rulers-leave-it/#sthash.cOkLXOJV.dpbs>
- [6] Taobao.com [Web site]. Available: <http://www.taobao.com/>
- [7] Tmall.com [Web site]. Available: <http://www.tmall.com/>
- [8] Alibaba Group. Alibaba Group Company Overview (21 October 2013.) [Online]. Available: <http://www.alibabagroup.com/en/about>
- [9] William H.E. Day, H Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of Classification*, Vol. 1, Dec, 1984. pp. 7-24
- [10] William H.E. Day, H Edelsbrunner, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, Aug, 2005. pp. 1226-1238
- [11] J.P. Peter, J.C Olson (2010). *Consumer Behavior and Marketing Strategy* (9th ed.). McGraw-Hill Companies. (Section 2).
- [12] J. Manyika, M Chui, and B. Brown, et al, Big data: The next frontier for innovation, competition, and productivity (May. 2011.) [Online]. Available: [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)
- [13] Guiping Zhang, Transport Bit Power to Enterprise Accurately (*in SCIENCE AND TECHNOLOGY DAILY*, in Chinese) (Nov. 2013.) [Online]. Available: [http://digitalpaper.stdaily.com/http\\_www.kjrb.com/kjrb/html/2013-11/18](http://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2013-11/18)
- [14] GYTaobao.cn (Jun. 2014) [Web Site] Available: <http://www.gytaobao.cn/>
- [15] D.F. Cai, Y. Bai, and G.P. Zhang, et al, "The Knowledge Service Project in the Era of Big Data," *In Proceedings of IEEE International Congress on Big Data*, CA, USA, Jun. 2013. pp. 429-430