

5-2018

# The Outlier Knowledge of Products Based On Grey Privacy Information

Huosong Xia

*Wuhan Textile University, bxxhs@sina.com*

Xiaolin Zhang

*Wuhan Textile University, 923567099@qq.com*

Yuting Meng

*Wuhan Textile University, 940976357@qq.com*

Follow this and additional works at: <http://aisel.aisnet.org/confirm2018>

---

## Recommended Citation

Xia, Huosong; Zhang, Xiaolin; and Meng, Yuting, "The Outlier Knowledge of Products Based On Grey Privacy Information" (2018).  
*CONF-IRM 2018 Proceedings*. 48.

<http://aisel.aisnet.org/confirm2018/48>

This material is brought to you by the International Conference on Information Resources Management (CONF-IRM) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CONF-IRM 2018 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# THE OUTLIER KNOWLEDGE OF PRODUCTS BASED ON GREY

## PRIVACY INFORMATION

Huosong Xia  
Wuhan Textile University  
bxxhs@sina.com

Yuting Meng  
Wuhan Textile University  
940976357@qq.com

Xiaolin Zhang  
Wuhan Textile University  
923567099@qq.com

### ***Abstract:***

Product attributes are closely related to consumers' willingness to buy on e-commerce. Some products are considered that few people would buy them online, but in fact they have large sales volume, such as underwear products. This study analyzed from the perspective of the product grey private information of why consumers buy and who is buying these products, what kind of relationship between the purchasing comments and product attributes of these products. Firstly, we select the companies with certain influence to the exploratory research to study the company's comments, and do the data mining of the top 10 products of the best sales and the last 10 products of the poor sales. Secondly, we analyzed the product attributes of the 14 enterprises again by the iterative method and compared with the first exploration result. Thirdly, we studied on the value of common knowledge and outlier knowledge for these products. The study found that common knowledge has the sensitive interval for the price of grey private products, the product price, the discount of the product and the sales volume of the product have significant relationship. The discovery of outlier knowledge is open and the comfort is very different. There is a significant positive correlation between the time of marketing and the sales volume: simple, natural, health products and sales are related; the extreme situations are not universal, but have good room for growth.

### ***Keywords:***

Outlier knowledge; grey privacy information; measurement of privacy information; grey private product information measurement; degree of openness

## **1. Introduction**

According to a forecast from iReserch ConsultingWorks, a market research firm, online shopping spending in China will grow to RMB12,198 in 2017, which is up 7% from 2016. Varieties of online purchases also gradually expanded, many are considered suitable for offline purchases of goods, but also gradually favored by young consumers in online shopping.

The biggest feature of e-commerce is the achievement of personalized custom products. As people's awareness of consumption increases, consumers are not only pursuing the functional attributes of products but also increasingly expecting personalized non-functional

requirements. Personalized customization is also a very popular mode of production and service in nowadays. In order to realize the individualized needs of users, such as the Han dynasty clothing store, the concept of "style, quick update and high cost performance" set off a wave of kin in China. Personalized customization will inevitably lead to increased costs, reduced production scale and other issues. Grey private products such as underwear, mostly small and medium-sized micro-production scale and comfort occupy the market advantage, personalized custom relatively small.

Some interesting phenomena are: Privacy products, such as underwear, young consumers not only purchased them in large quantities, but also comment a lot. The issue of "Helpfulness of Online Reviews" has drawn considerable attention and exploration by business and academia since Chatterjee (2001) first proposed the concept. In addition to grey private product reviews, explosions and slow-moving product features are very characteristic. What are the reasons that make young consumers willing to buy these products? Is there a certain sensitive interval to meet the common needs of consumers? What differentiated, personalized needs need to continue to meet? To answer these questions, the study analyzes consumers' purchasing decisions based on the online shopping habits. As a reference for their own products, first, consumers usually collect and browse the prices of the products, sales volume of the products and other consumers' online reviews. Second, compared to the quality of the price, it is possible to see the extent to which the merchant's products meet the demand. Finally decide whether to buy. Do consumer buy grey private products also follow these steps?

From the perspective of grey private information, this study analyzes how consumers can transform their common knowledge based on product online reviews, sales, product prices and other prior information, and promote sales of single products to a large sales volume. At the same time, analysis of why products that are also considered to be able to meet the needs of the consumer are unsalable? From the analysis of data sources, it is mainly based on online reviews of consumer products, sales volume, product information and so on, then refined these into regular knowledge and refined into outlier knowledge by analyzing some outlier phenomenon. The remaining main structure of the paper is as follows: The second part is the grey private product information measurement and outlier knowledge measurement model, the third part is the research design and research process, the fourth part is the conclusions and discussion.

## **2. Grey private product information measurement and outlier knowledge measurement model**

### **2.1. Grey private product information measurement**

#### ***2.1.1. Measurement of privacy information***

From the sociological point of view, information privacy refers to the time, way and degree of its private information being transmitted by individuals or organizations (Westin, 1968). There is no definition from a product perspective. In fact, privacy information has been exposed when consumers buy some products. To some extent, these private information is being used by

other consumers or businesses. The degree of information disclosure (DoIP: degree of Information Privacy) can be measured as Formula 1.

$$\text{DoIP} \in [0,1] \tag{1}$$

If DoIP = 1, the openness of information is the most transparent and it can be used publicly and legally;

If DoIP = 0, the openness of information is the least transparent and cannot be used publicly and legally;

If  $0 < \text{DoIP} < 1$ , the openness of information is close to 0, the degree of grey privacy is high, the degree of disclosure of information is close to 1, and the degree of grey privacy is low.

**2.1.2. Grey private product information measurement**

Closely related to product information is product attributes. By analyzing product attribute information, a measurement variable can be described by 18 attributes (x1, x2 ... x18), which can be measured according to the DoIP(degree of Information Privacy).Some of the information is shown to be a little bit more private, such as the degree of openness x9 is shown in Table 1.

The measurement variable x9	describes	for example a description of a underwear product	remarks
O1	Conservative	4/4	
O2	common	3/4	
O3	More open	2/4	
O4	open	1/4	

Table 1 Degree of openness x9 Measurement

**2.2. Grey private products outlier knowledge measurement model**

The work of grey private information product measurement framework mainly focuses on three aspects: firstly, through the product feature description and the product attribute domain knowledge, and selects the representative brand to make the comment crawler to establish the comment feature of interest. Secondly, the data set of 18 characteristic attributes is studied by using the product attribute characteristics. Explore the relationship between variables through statistical software and select the scatter plot between attributes to find the common knowledge and outlier knowledge. Thirdly, expand the scope of validation data from 14 enterprises, each for 20 products (burst sales of 10 products, slow sales of 10 products), a total of 280 data experiments to test the model effect. Details are presented in Figure 1. By climbing online transaction data and interaction data, we transformed product feature show measurement data into variables, then we got outlier knowledge and common knowledge through data analysis and processing.

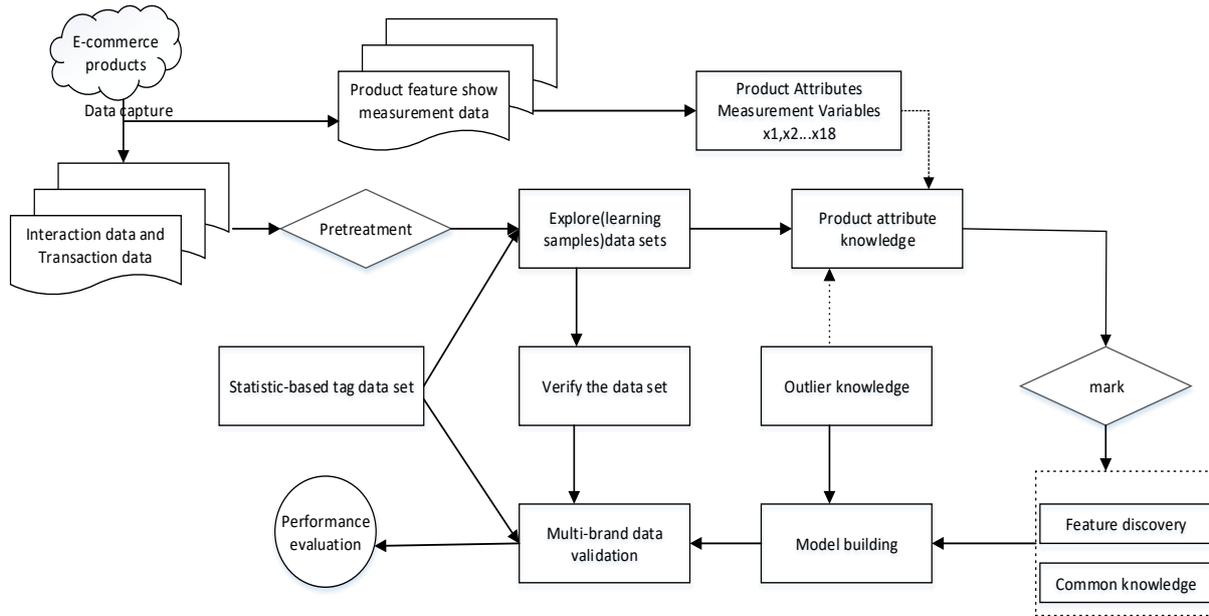


Fig 1: Outlier knowledge measurement model of grey private products

In the process of outlier knowledge discovery, outliers and outlier data are found by outlier detection method. Hawkins first proposed that outlier is different from other data in the data set (Hawkins, 1980). The improvement of algorithm in outlier detection method has been well developed from the technical level, such as statistical detection method. The method mainly identifies Outlier through the method of hypothesis testing. By detecting the isolated point in the data set as normal or Outlier, Outlier is the least probable data, if it is true, the other data are Outlier; otherwise, it is not, such as EkFk statistical algorithm (Marasinghe, 1985); such as distance-based statistical research methods: Index-Based algorithm , Mao R, Liao H, et al, 2016), Cell-based algorithm (Shaikh SA, Kitagawa H, 2014) and so on. DBSCAN, K-modes and other detection methods, the minimum cluster size needs to be set in advance, the distance threshold between the minimum cluster and other clusters needs to be determined, and the outlier cannot be clustered in the data set. However, the disparate causes and mechanisms have been neglected. This is precisely the focus of the discovery of outlier knowledge. Figure 1 is a general model for exploring the discovery of outlier knowledge of grey private products.

### 3. Research design

#### 3.1. Data Collection

In order to illustrate the value and effectiveness of the general model for the discovery of outlier knowledge of grey private products, this research chose 15 brands as the research object of grey private products in an e-commerce platform and labeled the presentation information of each product attribute, including the product time to market, openness information, style, buckle rows, inserts, comfort level, style, pattern, size, color and other product properties of the 18 kinds of related information description to carry out research. Research data was collected in mid-December 2017, which was a special time. Having just

experienced double 12, the data was objectively presented. Fifteen brands of grey private goods were selected and their top 10 and sales information of 10 products were captured. Which analyzed the scatter plot of the first 10 product attributes X4 of a certain brand of implicit private burst .Details are presented in Figure 2, we can see some common knowledge of X4: the outlier of the X4-sensitive and scatter-sensitive intervals.

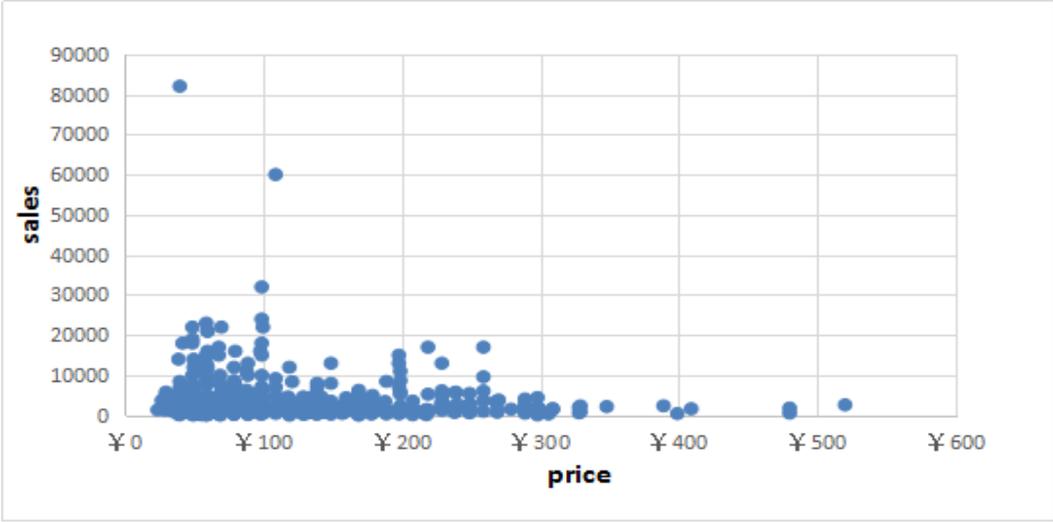


Fig 2: A brand grey private section of the first burst of product properties of a scatter plot 1

**3.2. Data Analysis and outlier Knowledge Discovery**

Analytical data shows that online product information has a positive impact on consumer purchasing decisions and further affecting product sales. Offline store research found that: Consumers will be more inclined to buy grey private products in the store. But after online data analysis and offline survey is inconsistent: online sales of private products have a large amount of sales. To analyze the reasons for who bought the grey private product, we use monthly data to replace the sales data to study the relationship between grey private attributes and product sales. Details are presented in Table 2.

variables	description	minima	maxima	mean	Standard deviation
sales	monthly sales	0	15709	753.63	1619.567
price	price	55	2690	349.441	277.8573
discount	discount	55	2690	308.174	286.1188
time	old medium new	1	3	1.5	0.756
size	quantity	0.5	1	0.8033	0.31489
Shoulder style	quantity	1	5	1.47	0.666
buckle	quantity	1	6	2.98	1.206
insert	quantity	1	4	1.28	0.580
rims	have or not	0	1	0.71	0.453
style	quantity	1	5	1.71	0.895
pattern	quantity	1	6	2.39	1.557

Table 2 Main variables and their description

In order to detect the relationship between variables, the study intended to use SPSS19.0 statistical software for regression analysis of variables to determine. In order to analyze the impact of grey private product information on product sales, we established a model as in Equation 2 for estimation.

$$\text{Sales}_i = a_0 + a_1 \text{price}_i + a_2 \text{discount}_i + a_3 \text{time}_i + a_4 \text{size}_i + a_5 \text{shoulderstyle}_i + a_6 \text{buckle}_i + a_7 \text{insert}_i + a_8 \text{rim}_i + a_9 \text{style}_i + a_{10} \text{pattern}_i \quad (2)$$

We use subscripts ( $i = 1, \dots, N$ ) on behalf of different products, the model of the dependent variable for the monthly sales, price on behalf of product prices, discount on behalf of product discounts, time on behalf of product time to market, size on behalf of product openness, shoulder style On behalf of the shoulder strap style, buckle on behalf of the number, insert on behalf of product inserts, style on behalf of product design, pattern on behalf of product design. Equation 3 is the formula after regression, Details are presented in Table 3.

$$\text{Sales}_i = 1643.674 + 3.081 \text{price}_i - 3.719 \text{discount}_i + 315.996 \text{time}_i + 475.201 \text{shoulderstyle}_i - 829.83 \text{rim}_i - 340.38 \text{style}_i - 119.078 \text{pattern}_i \quad (3)$$

model	Non-standardized coefficient		Standard factor	t	Sig.
	B	Standard error	trial version		
constant	1643.674	554.171		2.966	.003
x3	3.081	1.152	.529	2.675	.008
x4	-3.719	1.115	-.657	-3.334	.001
x6	315.996	122.686	.148	2.576	.011
x9	-340.059	273.636	-.066	-1.243	.215
x10	475.059	132.377	.195	3.590	.000
x11	14.288	76.198	.011	.188	.851
x12	-236.560	149.614	-.085	-1.581	.115
x13	-829.838	202.407	-.232	-4.100	.000
x14	-340.380	104.054	-.188	-3.271	.001
x15	-119.078	58.328	-.115	-2.042	.042

Table 3 Description of Major Regression Analysis

#### 4. Discussions and Conclusions

From the above analysis, it is found that there is a significant relationship between the product price, the product promotion price and the product sales volume. In general, the actual price of the consumer purchasing product is not the product price but the price after the product's promotion. From regression analysis results, ( $a_2 = -3.719$ ,  $p = 0.001$ ), indicating that the sales volume of the product is significantly affected by the price and the product price has a sensitive range. When the price exceeds a certain range, the sales volume will decrease. One of the findings is that the common knowledge is that there is a sensitive interval in the price of grey private products, and there is a significant relationship between product price, product promotion price and product sales volume.

There was a significant positive correlation between the time to market and the sales volume

( $a_3 = 315.996$ ,  $p = 0.011$ ). The earlier the time to market was, the more mature the information consumers knew. The stronger the desire to buy, the higher the sales volume. There was a significant positive correlation between shoulder strap style and product sales ( $a_5 = 475.201$ ,  $p = 0.000$ ), and consumers were more inclined to buy simple products for implicit private products. There is a significant negative correlation between the product presence and the product sales ( $a_8 = 475.201$ ,  $p = 0.000$ ). The product with steel rings shows low comfort and low sales, which is consistent with the related research. Become a burst of online private sales of personal secrets, indicating that consumers are increasingly concerned about the health of private property bias. The second finding is that there is a significant positive correlation between the time-to-market and the sales volume of the products, and the products are simple, natural, health and sales are positively correlated.

Product style and product sales there is a significant negative correlation ( $a_9 = -829.838$ ,  $p = 0.000$ ), the more novel style, the lower sales. There is a significant negative correlation between product design and product sales ( $a_{10} = -119.078$ ,  $p = 0.042$ ). The more complex the product design, the lower the product sales. There are also some product attributes have no significant impact on sales of privately-held products. The third finding is that the discovery of outlier knowledge is openness, comfort varies widely. The extreme does not have universal, but has good room for growth.

In a word, from the perspectives of common knowledge and outlier knowledge, we found that grey private products in the general public are in a conservative state. However, we also found some relatively open sprouts which are related to the differences between Chinese culture and western culture. This is also a future consideration for further research. From the sensitive interval for the price, the sales volume with high discounts is very high, which is consistent with the common knowledge. But some outlier data also see the new hope that the discovery of outlier knowledge of high prices products and its mechanism have not yet been completely revealed.

## **ACKNOWLEDGEMENT**

This research has been supported by the National Natural Science Foundation of China (71571139); We deeply appreciate the suggestions from fellow members of Xia's project team and Research center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences in Universities of Hu Bei Province (DSS20150215 & DSS20150108).

## References

- Moretti, I.C., Junior, A.B., Colmenero, J.C. (2017). Using the analytic hierarchy process for selecting prototypes in the development process of fashion garment products[J]. *Acta Scientiarum Technology*, 39(3):367-374.
- Jerger, J., Wirtz, J. (2017). Service Employee Responses to Angry Customer Complaints: The Roles of Customer Status and Service Climate[J]. *Journal of Service Research*, 20(4):362-378.
- Rahman, O., Fung, B.C.M., Chen, Z., et al. (2017). A Cross-National Study of Apparel Consumer Preferences and the Role of Product Evaluative Cues[J]. *Asia Pacific Journal of Marketing & Logistics*, 29(4):796-812.
- Hall, A., Towers, N. (2017). Understanding how Millennial shoppers decide what to buy: digitally connected unseen journeys[J]. *International Journal of Retail & Distribution Management*, 45(5):498-517.
- Iran, S., Schrader, U. (2017). Collaborative fashion consumption and its environmental effects[J]. *Journal of Fashion Marketing and Management*, 21(4):468-482.
- Kennedy, A.M., Kapitan, S., Bajaj, N., et al. (2017). Uncovering wicked problem's system structure: seeing the forest for the trees[J]. *Journal of Social Marketing*, 7(1):51-73.
- Hawkins, D.M. (1980). Identification of Outliers[J]. *Biometrics*, 37(4):860.
- Marasinghe, M. (1985). A Multistage Procedure for Detecting Several Outliers in Linear Regression[J]. *Technometrics*, 27(4):395-399.
- Xu, H., Mao, R., Liao, H., et al. (2016). Index Based Hidden Outlier Detection in Metric Space[J]. *Scientific Programming*, 2016(2):1-14.
- Shaikh, S.A., Kitagawa, H. (2014). Efficient distance-based outlier detection on uncertain datasets of Gaussian distribution[M]. Kluwer Academic Publishers, 17(4):511-538.
- Patrali Chatterjee. (2001). Online Reviews: Do Consumers Use Them?[J]. *Advances in Consumer Research*, 28(01):129-133.
- Westin, A.F. (1968). Privacy and Freedom[J]. *Washington and Lee Law Review*, 25(1):166-170.