Advance Journal of Food Science and Technology 9(1): 60-64, 2015
DOI: 10.19026/ajfst.9.1934
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: January 25, 2015 Accepted: February 27, 2015 Published: July 30, 2015

Research Article

Analysis on Information Sources of Consumers' Perception of Food Safety Risks by ID3 Algorithm

Guo Xiao-Yan and Gao Wei-Yi
School of Management, Wuhan University of Technology, Wuhan 430070, China

Abstract: The aim of this study is to research of the effect on different sources of information on consumers' perception of food safety risks to the level of consumers' perceived risk. This study through the questionnaire survey, using ID3 algorithm to do analysis. The results showed that consumers' food safety risk perception level was generally high, which adopted by the government and experts to get information was the highest. Finally, combining the research results were presented to suggestions of reducing consumers' perceived risk levels.

Keywords: Food safety, ID3 algorithm, information source, risk perception

INTRODUCTION

"People make food their heaven", food is the basic material conditions of humans' survival and development. At present china is in the period of social transition, due to a variety of complex reasons, a series of food safety accidents have happened successively, the food safety causes community more and more widespread attention., people have varying degrees of concern even have a panic on the food security situation. The 21th meeting of the 11th national people's congress held in June 29, 2011 suggested to include food security and financial security, grain security, energy security, ecological security in the "national security" system, suffice it to say that the food safety risk has become an extremely serious and major problem at the national level.

In recent years, consumers' risk perception level of food safety increased year by year, seemingly due to frequent accidents, but the key reason is consumers do not have enough information to provide opportunistic space for production operator in food market,. Some scholars at home and abroad have done the related research of cognition and demand for food safety risk information. Rutsaert *et al*. (2013) pointed out that the development of social media had brought a lot of information to the consumers, but also brought the related issues of information accuracy, the reliability of trust and sources of information. Vainio *et al*. (2013) said in the context of daily food consumption, individual had to through the food safety management mechanisms getting information to assess their health and environmental risks. At the same time, they have to trust to the agencies protecting them and providing them risk information. Wu and Xu (2009) inspected the Jiangsu residents about the causes of food safety cognition, behaviors of consumers demanding for food safety information. Results showed that consumers with a higher degree of concern on the label information, needing high quality food safety information and production information of whole process. Cha Hua (2009) proposed food production operators, the mass media consumers associations and government were the basic channels for consumers to get food safety information. Zhang and Liu (2010), through the survey found that the main source of consumers getting food quality and safety information was media of information.

Chinese and foreign scholars on the urgency of consumers demanding for food safety information and the main channels of obtaining information to do the corresponding analysis, but the studies of association between on consumers' perceived risk level and the sources of information were not much, based on this, the paper will through the questionnaire survey, using ID3 algorithm to mining the relationship between consumers' perceived risk level and source of information.

MATERIALS AND METHODS

Overview of information sources of consumers' perception of food safety risks: Consumers' perceived risk refers to the possibility when consumers make purchasing decision they feel the purchased products can't reach their expected. Once consumers perceive the risk of purchase, their behaviors will immediately make a change (Mahon and Cowan, 2004). There is an acceptable risk, as soon as the risk perception of consumers exceed this range, people will take a series of measures to reduce the risk level to an acceptable

Corresponding Author: Guo Xiao-Yan, School of Management, Wuhan University of Technology, Wuhan 430070, China

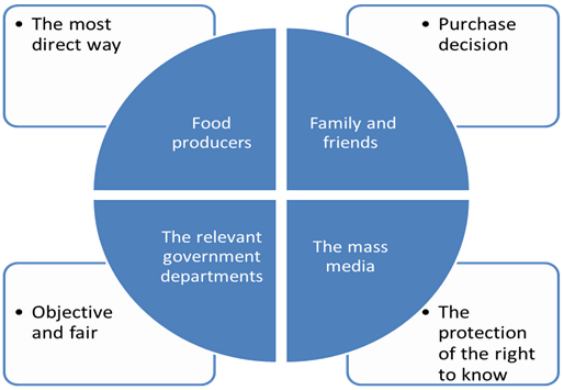

Fig. 1: Consumers access to food safety risk source of information

degree, the risk perception level is higher, consumers are more likely to change their behaviors to cut risk.

Through the summary of massive literatures, consumers' sources of access to food safety risk information can be divided into four categories: the food manufacturers, family and friends, the mass media, the relevant government departments, as in Fig. 1

The basic principles of ID3 algorithm: (Jody, 2012) ID3 algorithm is a kind of inductive learning method in the field of artificial intelligence, bases on the "attribute-value" information structure by setting fact instance to calculate the information gain and as a standard to select the best test attribute, so that tested in a non leaf node, can obtain the largest category information about the record tested to get the selection process of collection of instances and the decision tree. ID3 algorithm constructing decision tree is simple, introducing it in this study, the known data can be classified and analysed, which contribute to understand the correlation between different information sources and consumers' perceived risk level, provide the basis. To make strategies of reducing consumers' perceived risk level.

Suppose S is a collection of s data samples. Assume that the class label attribute have different value of m, the definition of m a similar C_i (i = 1, 2, …, m). Let S_i be the number of samples in the class C_i. The expected information of a given sample classification as follow:

$$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

where, P_i is the probability for arbitrary sample belongs to C_i and uses s_i/s estimation.

Set the A attribute has different values of v: $\{A_1, A_2, ..., A_V\}$. The S can be divided into v subsets $\{S_1, S_2, ..., S_V\}$ by the A attribute; wherein, S_i contains some samples like S that, they own the value a_i in the A. If A is chosen as the test attribute (that is the best splitting attribute), then these subsets correspond to branches which grown from node contains a collection of S.

Set s_{ii} is the number of samples of class C_i in a subset S_i. The entropy or desired information of subsets divided by A is given by the following formula:

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j}+s_{2j}+\cdots+s_{mj}}{s} I(s_{1j}, s_{2j}, \ldots, s_{mj})$$

Wherein, $\frac{s_{1j}+s_{2j}+\cdots+s_{mj}}{s}$ is the weight ofthe j subset and is equal to the number of samples of subset (that is A value of a_i) divided the total number of samples of S. Entropy is smaller, the higher the purity of subset partition. Note that, for a given subset s_j:

$$I(s_{1j}, s_{2j}, \ldots, s_{mj}) = -\sum_{i=1}^{m} p_{ij} \log_2(p_i)$$

Among them, $p_{ij} = \frac{s_{ij}}{s}$ is the probability of the samples of S_i belonging to the class C_i. Code information will be obtained by branched in A is:

$$Gain(A) = I(s_1, s_2, \ldots, s_m) - E(A)$$

Gain (A) is called the information gain, it is due to know the value of the attribute A that lead the desired of entropy to be compressed. The attribute with the highest information gain is chosen as the test attribute of a given set of S. Create a node and to tag the attribute, create branches on each value of attribute and then based on that to divide the sample.

The analysis step of information sources of consumers' perception of food safety risks by ID3 algorithm:

Step 1: The research phase. This phase uses a questionnaire survey to obtain the condition of consumers' food safety risk perception and the channels of accessing to information, gets the attributes' set List = {Gender, newspapers or related books channel, the channel of the friends and relatives, the government or expert advice channel, television broadcasting and network media channels, channel of brand advertising or manufacturer guide consumers} (due to so much influence of personal attributes of consumers' perceived risk level, in order to facilitate the modeling, this study only selects gender as consumers characteristic attribute; and the mass media categories for the press or the relevant books and television broadcasting network media).

Step 2: If all the nodes are the final nodes, then terminate the calculation, or turn to step three.

Step 3: If the Attribute Collection List = (), then terminate the calculation, or turn to step four.

Step 4: Selecting for each attribute from the attribute collection to divide non final nodes instance

set, by calculating the information entropy respectively, comparing the average information contents of attributes to get the smallest one, forming the classification results. Analysis of the classification results one by one, if the results are consistent and not empty, then the results of classification are the final nodes of decision tree, which are marked by the classification results, if empty, the classification results are the final nodes and to be deleted.

Step 5: Just delete attributes used to divide example set and turn to step two.

Step 6: Each attribute completes to be calculated, get the decision tree of instance set S.

EXPERIMENTS

Test materials: This study adopted the network questionnaire survey method to collect the public food safety risk perception level and sources of accessing to information, to confirm the effect of the domestic consumers' different sources of information to their food safety risk perception.

A total of 110 questionnaires were filled out, removed the questionnaires of draining answer key information and error information occurred, there are 99 valid ones which had 90% efficiency. The obtained questionnaires summary as shown in Table 1.

By investigation, the public mainly through "relatives and friends", "television broadcasting network media" to obtain related information of food safety crisis, there were 24.24 and 42.42% respondents choosing the source of information accordingly and 18.18% of the public by means of "newspapers or books related "understanding of the relevant information, only 9 and 6% respondents from the "government or expert advice" and " brand advertising or manufacturer guide consumers "getting information.

Build the ID3 decision tree: Used ID3 algorithm to analyze the summary results above, the process was as follows:

s = 99, the class label attribute "food safety risk perception level" had three different values (i.e., {high, medium, low}), so there were three different types (m = 3). Assumed that the class C_1 corresponded to the "high", C_2 corresponded to the "medium ", C_3 corresponded to the "low". Then $s_1 = 31$, $s_2 = 50$, $s_3 = 18$, $p_1 = 31/99$, $p_2 = 50/90$, $p_3 = 18/99$.

- The calculation of the desired information as required for the given samples classification:

$I(s_1,s_2,s_3)=-\sum_{i=1}^{3} p_i \log_2(p_i)=-\frac{31}{99}\log_2\frac{31}{99}-\frac{50}{99}\log_2\frac{50}{99}-\frac{18}{99}\log_2\frac{18}{99}=1.469$

Calculate the entropy of each attribute.

First calculate the entropy of attribute "gender":

For the gender = "male": $s_{11} = 9$, $s_{21} = 16$, $s_{31} = 7$, $p_{11} = 9/32$, $p_{21} = 16/32$, $p_{31} = 7/32$

$I(s_{11},s_{21},s_{31}) = -\frac{9}{32}\log_2\frac{9}{32}-\frac{16}{32}\log_2\frac{16}{32}-\frac{7}{32}\log_2\frac{7}{32} = 1.494$

For the gender = "female": $s_{12} = 22$, $s_{22} = 34$, $s_{32} = 11$, $p_{12} = 22/67$, $p_{22} = 34/67$, $p_{32} = 11/67$

$I(s_{12},s_{22},s_{32}) = -\frac{22}{67}\log_2\frac{22}{67}-\frac{34}{67}\log_2\frac{34}{67}-\frac{11}{67}\log_2\frac{11}{67} = 1.452$

If the samples divided according to the "gender", the desired information as required for a given sample classification:

$E(\text{gender}) = \frac{32}{99} I(s_{11},s_{21},s_{31}) + \frac{67}{99} I(s_{12},s_{22},s_{32}) = 1.466$

Therefore, the information gain of the division was:

Gain (gender) = $I(s_1,s_2,s_3)$-E(gender) = 1.469-1.467 = 0.003.

Similarly computed the attribute "information sources entropy of food security":

For attribute = "1": $s_{11} = 9$, $s_{21} = 6$, $s_{31} = 3$, $p_{11} = 1/2$, $p_{21} = 1/3$, $p_{31} = 1/6$

$I(s_{11},s_{21},s_{31}) = -\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{3}\log_2\frac{1}{3}-\frac{1}{6}\log_2\frac{1}{6} = 1.459$

Table 1: The results of the investigation summary table

Serial number	Gender	Sources of the food safety information	The food safety risk perception level
	1 = Newspapers or books related 2 = Relatives and friends 3 = The government or expert advices 4 = Television broadcasting network media 5 = Brand advertising or manufacturer consumer guide		
1	Female	2	Medium
2	Male	1	Medium
3	Male	3	High
……			
99	Female	4	High

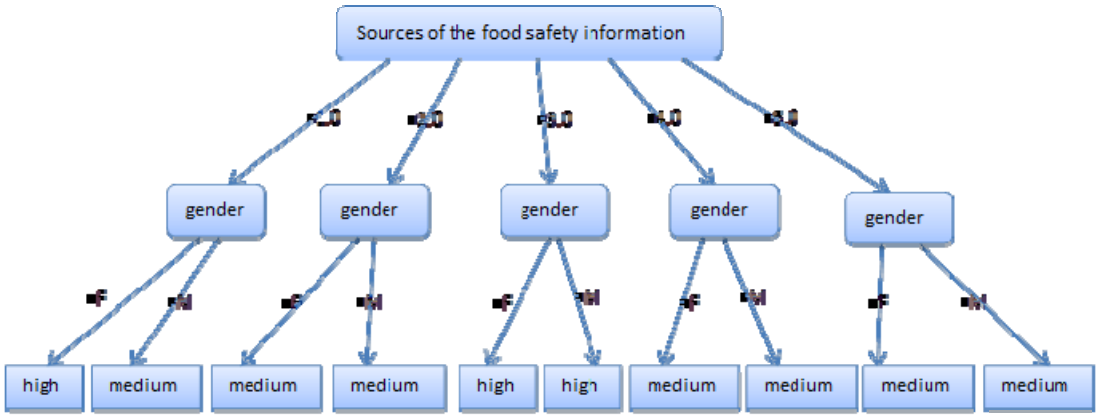

Fig. 2: The decision tree of using ID3 algorithm to analysis the food safety information sources on. (1.0 = newspapers or books related; 2.0 = family and friends; 3.0 = government or expert advices; 4.0 = television network media; 5.0 = brand advertising or manufacturer guide)

For attribute = "2": $s_{12} = 4$, $s_{22} = 14$, $s_{32} = 6$, $p_{12} = 1/6$, $p_{22} = 7/12$, $p_{32} = 1/4$

$I(s_{12},s_{22},s_{32}) = -\frac{1}{6}\log_2\frac{1}{6} - \frac{7}{12}\log_2\frac{7}{12} - \frac{1}{4}\log_2\frac{1}{4} = 1.384$

For attribute = "3": $s_{13} = 4$, $s_{23} = 3$, $s_{33} = 2$, $p_{13} = 4/9$, $p_{23} = 1/3$, $p_{33} = 2/9$

$I(s_{13},s_{23},s_{33}) = -\frac{4}{9}\log_2\frac{4}{9} - \frac{1}{3}\log_2\frac{1}{3} - \frac{2}{9}\log_2\frac{2}{9} = 1.530$

For attribute = "4": $s_{14} = 13$, $s_{24} = 22$, $s_{34} = 7$, $p_{14} = 13/42$, $p_{24} = 22/42$, $p_{34} = 1/6$

$I(s_{14},s_{24},s_{34}) = -\frac{13}{42}\log_2\frac{13}{42} - \frac{11}{21}\log_2\frac{11}{21} - \frac{1}{6}\log_2\frac{1}{6} = 1.443$

For attribute = "5": $s_{15} = 1$, $s_{25} = 5$, $s_{35} = 0$, $p_{15} = 1/6$, $p_{25} = 5/6$, $p_{35} = 0$

$I(s_{15},s_{25},s_{35}) = -\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6} = 0.650$

If the samples were divided according to the "food safety information sources", the desired information as required for a given sample classification:

E (food safety information sources) $= \frac{18}{99}I(s_{11},s_{21},s_{31}) + \frac{24}{99}I(s_{12},s_{22},s_{32}) + \frac{9}{99}I(s_{13},s_{23},s_{33}) + \frac{42}{99}I(s_{14},s_{24},s_{34}) + \frac{6}{99}I(s_{15},s_{25},s_{35}) = 1.391$

Therefore, the information gain of the division was:

Gain (food safety information sources) = $I(s_1, s_2, s_3)$-E (food safety information sources) = 1.469-1.391 = 0.078.

ID3 algorithm selected the maximum attribute information gain as nodes, at this time, choosed "food safety information sources" as the root node and its five attribute values were divided into five branches, the training set was divided into five categories, the branches contained only a risk perception level would be marked categories, the branches contained various risk perception levels would continue to recursive calls contribution algorithm, got the decision tree shown in Fig. 2.

EXPERIMENTAL RESULTS AND ANALYSIS

- No matter which the source of information to obtain food safety information, The higher prevalence of food safety risk perception level of women, this maybe women are responsible for housework, with more time to prepare the family food, thus they will take more attention to some of the food safety information and lead to higher level of food safety risk perception.
- Consumers access to food safety information through "newspapers or relevant books", their perceived risk levels are higher, mainly because the majority of the mainstream presses generally release the events of food safety problems, the media that participate in the investigation show they have limited exposure to the food safety, can't be comprehensive reports. While men compared with women, men's perceived risk levels are lower, may be the male consumers see the newspapers are often concerned about political news, while the female consumers more biased in favor of life news.
- Consumers' food safety risk perception levels are high when they mainly to obtain the food safety information through "government or expert advices". That shows food safety information obtained through this channel has shook consumer's dependence for information, confusion and panic for the food safety.
- Consumers through the "relatives and friends" and "TV broadcast network media" access to

information, their food safety risk perception levels are lower than the rest of the two channels. Word of mouth between friends and relatives belonging to a word of mouth information, consumers generally feel a sense of trust for their familiar people around, believe that their passing reputation information does not have the commercial color, which have higher credibility than advertising.

CONCLUSION

In general, consumers' food safety risk levels are universally high now a days, ultimately, the crux of the problem is the public with distrust and worry about food safety, even fear. The food safety judgment depends on accessing to food safety information. To reduce consumers' perceived risk level, there are a few suggestions:

- As can be seen from the results of the survey, along with the network as a kind of new media platform obtained the rapid development, the network has become the main channel for people to get information, all wind sways grass concerning food safety can easily spread through the network and rapidly swell. So the government should increase the mass media including network supervision, purify network environment and the network media must adhere to the principle of seeking truth from facts of food safety information disclosure, use of user-friendly, rich and colorful form for consumers to expand the popularity of propaganda of the food safety knowledge, enhance the identification ability of consumers on the issue of products.
- Strengthen the supervision of departments and disclose information timely. Government's chief executives of all levels are not the industry experts, they can command how to timely deal with the issues after them happened, but they can't predict what food have safety problems, the key to this study is to be supervised by functional departments. If the functional departments are lack of regulations, when the food safety issues happen, it will make the government into a passive situation.

ACKNOWLEDGMENT

This study was supported by Young of National Natural Science Foundation of China (No. 71203171).

REFERENCES

Cha Hua, 2009. The basic channels of consumers access to food safety information [J]. Legal Syst. Soc., 9: 219-221. (In Chinese)

Jody, L.F., 2012. Study on analysis process of employee satisfaction based on ID3 algorithm [J]. Hum. Resour. Manage., 4: 105. (In Chinese)

Mahon, D. and C. Cowan, 2004. Irish consumers' perception of food safety risk in minced beef [J]. Brit. Food J., 106(4): 301-312.

Rutsaert, P., A. Regan, Z. Pieniak Á. McConnon, A. Moss, P. Wall and W. Verbeke, 2013. The use of social media in food risk and benefit communication [J]. Trends Food Sci. Tech., 30(1): 84-91.

Vainio, A., J.P. Mäkiniemi and R. Paloniemi, 2013. System justification and the perception of food risks [J]. Group Process. Interg., Doi: 10.1177/1368430213503502.

Wu, L.H. and L.L. Xu, 2009. Food safety risk perception and consumers' behaviors, based onanalysising of consumers in Jiangsu Province [J]. Consum. Econ., 25(2): 42-44. (In Chinese)

Zhang, L. and G. Liu, 2010. Empirical consumers for fresh food quality and safety information search behavior analysis-Shanghai city fresh food consumption survey based on agriculture technical economy [J]. 2: 97-103. (In Chinese)