Research Journal of Applied Sciences, Engineering and Technology 6(18): 3447-3450, 2013
DOI:10.19026/rjaset.6.3664
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: January 22, 2013 Accepted: February 22, 2013 Published: October 10, 2013

Research Article

Entropy-Based Credit Evaluation for Mobile Telephone Customers

[1]Yang Zong-Chang and [2]Kuang Hong
[1]School of Information and Electronical Engineering,
[2]School of Arts and Industrial Design, Hunan University of Science and Technology, Xiangtan 411201, China

Abstract: The arrears problem puzzled most mobile communication corporations in China. In information theory, the Shannon entropy is a measure of the uncertainty in a signal or random event. Motivated by entropy theory especially the Shannon Entropy, in this study, one called customer information entropy is defined and proposed to credit evaluation for arrearage customers of cellular telephone. The proposed customer information entropy is based on customer's behavior attributes. Arrearage customers often include malevolent ones and non-malevolent ones. 52364 arrearage customers among a total number of 400000 ones in a mobile communication corporation are chosen for experiment. The proposed measure yields good results in its application of credit evaluation for 52364 arrearage customers in August and September. Its correct evaluation rates for malevolent and non-malevolent of the 52364 arrearage ones both are over 90.0%: among the 52364 arrearage customers, 90.75% of the non-malevolent ones, whose entropy changes is less than zero, while for the entropy changes of the malevolent ones, 95.36% is equal to zero and 1.57% is more than zero. The experimental results indicate that the entropy changes of the non-malevolent ones could be considered as negative and nonnegative for the malevolent ones. The proposed show its potential practicality.

Keywords: Arrearage, credit evaluation, customer information entropy, entropy, mobile communications

INTRODUCTION

One well-known idea involved in the concept of entropy is that nature tends from order to disorder in one isolated system. Entropy first introduced in classical thermodynamics, is a very import physical concept for it offers a quantitative basis for the common observation that a particular direction is enclosed in naturally occurring processes. Beyond the concept of entropy originally defined in a thermodynamic construct, it has been used outside of the classical thermodynamics, other fields including psychodynamics and information theory, thermo-economics and evolution also adapt the entropy concept. In statistical thermodynamics, entropy was described as a measure of the number of microstates in one system. In the communication information theory, entropy is a definition of measure of information (Perrot, 1998; Clausius, 1850; Laidler, 1995; Shannon, 1948).

In the statistical thermodynamics (Perrot, 1998; Clausius, 1850; Laidler, 1995), entropy is defined as a measure of a system's energy that is unavailable for work or of the degree of the system's disorder. When a system is added with heat held at constant temperature, the change in entropy is related to the change in energy, which is depicted by the changes of the pressure and the temperature and the volume. Its magnitude varies from zero to the total amount of energy in a system. In popular, non-technical use, entropy can be regarded as a measurement for the chaos or randomness of a system.

The concept of entropy in information theory (Shannon, 1948) describes how much randomness (or alternatively, "uncertainty") there exits in a signal or random event. An alternative way to see it is to talk about how much information is carried by the signal. For a discrete random variable X with possible states (or outcomes) $x_1 \ldots x_n$, its (Shannon) information entropy is defined by:

$$H(X) = \sum_{i=1}^{n} p(x_i)\log_2 \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \quad (1)$$

where, $p(x_i) = \Pr(X = x_i)$ is the probability of the i^{th} outcome of X.

Shannon indicates that any definition of entropy satisfying his assumptions will be of the form:

$$-K\sum_{i=1}^{n} p(x_i)\log p(x_i) \quad (2)$$

Corresponding Author: Yang Zong-Chang, School of Information and Electronical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

where, K is a constant (and is really just a choice of measurement units).

In other words, in information theory, the term of entropy is defined as a measure of the uncertainty in a random variable or event. In this connection, comparing to normal customers, arrearage customers especially malevolent arrearage ones should show some uncertainty in one defined measure based on the entropy.

Motivated by entropy theory especially the Shannon Entropy, in this study, we try to introduce a called entropy-based measure for credit evaluation for the mobile telephone customers.

Problem definition: In the year of 2001, it was reported (Yang and De-Qi, 2005) that in the telecommunication industry of China, a sum of ¥200,000,000 (about $25,000,000) was falling into arrears. In 2005, it showed that arrearage rates in the industry of Chinese mobile communication services still remained high, which were often approximately or over 10.0%. As to the global telecommunications service industry, this phenomenon is not unique that reported by a Cambridge investigate corporation, the loss of global communication services industry caused by delinquent customer was about more than billion pounds. As a conservative estimate, the industry of Chinese mobile communication services suffers more than 5.0% of loss in its total annual income (Yang and De-Qi, 2005). Take the Guangdong province of China for example, whose arrears in the mobile communication services industry are on average with ten millions Yuans each year (Yang and De-Qi, 2005). With the booming of China's mobile telephony industry, its amount of customers is now sharply increasing, while the arrears problem will be more shaped as well.

In China's mobile telephone industry, the typical countermeasures (Telecom-World, 2001) being taken for the arrears problem is simply setting a margin for stopping service. That is, when one's bill exceeds a margin that is set by the service provider, the customer will be stopped providing services. However, it can keep arrears within limits, but should bring customers much inconvenience and cause them discontent even leave off. People have now increasingly recognized that (Ji-Sheng *et al*., 2002; Jun-Yan *et al*., 2005; Yueh-Min *et al*., 2006; Jonathan *et al*., 2006; Cheng-Lung *et al*., 2007; Nan-Chen and Lun-Ping, 2010; Amir *et al*., 2010; Baesens *et al*., 2003; Cao *et al*., 2009; Zong-Chang and Hong, 2008) it is necessary to provide scientific analysis for the customers' credits.

The basic idea of the credit evaluation is to find the elements that can decide the customer's credits, which may be enclosed in the attributes of their behaviors. To one individual, there seems to be no apparent relationship between his credit and behavior attributes. However, according to our statistical analysis, it is found that there are certain relations between client behavior attributes and customer's credits on the whole, thus the customer's credit might be forecasted and evaluated by necessary mathematical algorithms with quantitative attributes.

ENTROPY-BASED CREDIT EVALUATION FOR MOBILE TELEPHONE CUSTOMERS

The well-known concept of entropy in information theory describes how much randomness (or alternatively, "uncertainty") there exits in a signal or random event. In this context, the term usually refers to the Shannon entropy. In other words, in information theory, the term of entropy is defined as a measure of the uncertainty in a random variable or event. In this connection, comparing to normal customers, arrearage customers especially malevolent arrearage ones should show some uncertainty in one defined measure based on the entropy.

Motivated by the entropy concepts, especially the concept of entropy in information theory, a definition of entropy for mobile phone customers is presented as follows.

In the credit evaluation for mobile telephone customers, each one of customers is identified by its behavior attributes, which is usually presented as an attribute (feature) vector. Suppose each customer with n attributes and W_i denotes one customer' attribute vector in i^{th} month (credit is usually monthly evaluated):

$$W_i = \{w_{i1}, w_{i2}, ..., w_{in}\} \tag{3}$$

Now, we define the entropy E_i for the customer in i^{th} month called Customer Information Entropy as follows:

$$E_i = -\sum_{k=1}^{n} \frac{w_{ik}}{W_{a_i}} \ln\left(\frac{w_{ik}}{W_{a_i}}\right) \tag{4}$$

where, $W_{a_i} = -\sum_{j=1}^{n} w_{ij}$

Then, we can measure the changes of one customer information entropy from i^{th} month to j^{th} month ($j>i$), denoted as ΔE_{ji}:

$$\Delta E_{ji} = E_j - E_i \tag{5}$$

Case study: Our experimental data is obtained from a mobile communication corporation in the Guangdong province of China. This corporation services about 400,000 customers. With both computational and evaluation effectiveness considered, based on our statistics analysis, from dozens of customer attributes, we chose and quantify 18 attributes as the attributes for the evaluation. For example, these attributes like the customer's "name", "sex", "birth date", since their statistical rates between the normal customers and the arrearage ones are approximately equal to 1.0, are not selected for evaluation.

Each customer is identified by 18 attributes as follows:

- Native or not
- International roaming function
- IDD (International Direct Dial) function
- Number of call transfer
- Number of new services
- Important group symbol
- Personal credit terms
- Mobile data service symbol
- Form of payment
- Monthly tab of new service
- Stop-service history
- Monthly tab of SMS (Short Message Service)
- Phone-number selection fee
- Monthly tab of monthly long-distance phone
- Enrollment date
- VIP customer symbol
- The corresponding province in the identification card
- Consumption level (over the past 3 months)

Among the total 400,000 customers, there are 52364 arrearage ones with complete records in August and September. We will recognize the 52364 arrearage customers into two categories: one is the customers whose states are falling into arrearage in both the two month that are defined as malevolent arrearage ones, the other is the customers whose states are falling into arrearage in August but are normal in September, which are defined as non-malevolent ones.

We then employ the proposed customer information entropy to measure the entropy changes of the 52364 arrearage customers in August and September, which include 5028 non-malevolent ones and 47336 malevolent ones.

Results are depicted in Fig. 1 and 2 and listed in Table 1 and 2, respectively.

It is found that among the 52364 arrearage customers, 90.75% of the non-malevolent ones, whose entropy changes is less than zero, i.e., whose entropy changes are negative; while for the entropy changes of the malevolent ones, 95.36% is equal to zero and 1.57% is more than zero, i.e., 96.93% are nonnegative. Thus, we might get that as to the arrearage customers, the entropy changes of the non-malevolent ones could be considered as negative, while nonnegative for the malevolent ones. In this way, correct evaluation rate of the proposed entropy-based method for the 52364 arrearage customers is over 90.0% (90.75% for the no malevolent ones and 96.93% for the malevolent ones).

The result having significant differences for the non-malevolent arrange customers and non-malevolent arrange ones, as a reference, is acceptable for the telephone service provider. The result also indicates that addressing the arrange problem, further management based on the credit evaluation is of workability.

Table 1: Results of the proposed entropy changes for the customers (ΔE_{ji}, j = 9, i = 8)

Arrearage customer	Non-malevolent	Malevolent
	5028	47336
ΔE_{ji}<0	4563	1455
ΔE_{ji} = 0	111	45140
ΔE_{ji}>0	354	741

Table 2: Results of the proposed entropy changes for the customers (ΔE_{ji}, j = 9, i = 8)

Arrearage customer	Non-malevolent	Malevolent
	5028	47336
ΔE_{ji}<0	90.75%	3.07%
ΔE_{ji} = 0	2.21%	95.36%
ΔE_{ji}>0	7.04%	1.57%

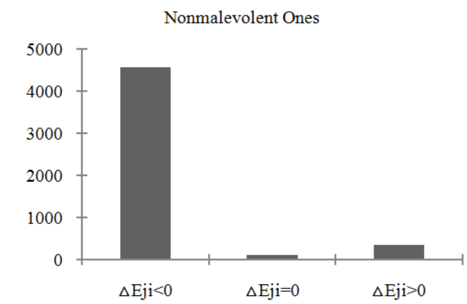

Fig. 1: Illustrations of the entropy changes for the non-malevolent arrange customers (ΔE_{ji}, j = 9, i = 8)

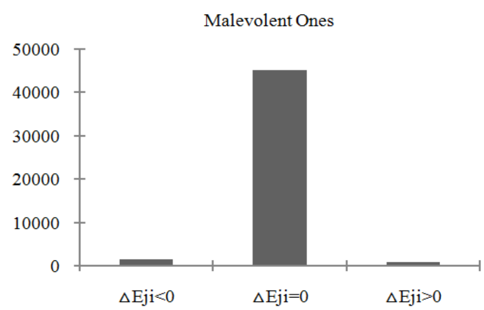

Fig. 2: Illustrations of the entropy changes for the malevolent arrearage customers (ΔE_{ji}, j = 9, i = 8)

CONCLUSION

One well-known idea involved in the concept of entropy is that nature tends from order to disorder in one isolated system. In information theory, entropy is a measure of the uncertainty in a signal or random event. In this connection, comparing to normal customers, arrearage customers especially malevolent arrearage ones should show some uncertainty in one defined measure based on the entropy.

Motivated by the entropy concepts, especially the concept of entropy in information theory, a definition of entropy for mobile phone customers is presented in this study. Experimental results show that for the 52364 arrearage ones among the total 400,000 customers, the correct evaluation rates of the proposed entropy-based method for both the malevolent ones and non-malevolent ones among the 52364 arrearage ones are over 90.0%. It indicates workability of the proposed measure based on entropy theory for credit evaluation. And others could benefit from ours for its potential practicality.

However, extra experiments with much more datasets employing the entropy-based credit measure for mobile telephone customers are still included in our further study. Additionally, to further extend and improve the proposed entropy-based measure model will also be considerably emphasized in our future work.

ACKNOWLEDGMENT

The research is supported by Scientific Research Fund of Hunan Provincial Education Department (12B044) and (09C399) and research fund of Hunan University of Science and Technology (E50811).

REFERENCES

Amir, E.K., K.J. Adlar and L.W. Andrew, 2010. Consumer credit-risk models via machine-learning algorithms. J. Bank. Financ., 34(11): 2767-2787.

Baesens, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens and J. Vanthienen, 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. J. Oper. Res. Soc., 54(6): 627-635.

Cao, R., J.M. Vilar, A. Devia, N. Veraverbeke, J.P. Boucher and J. Beran, 2009. Modelling consumer credit risk via survival analysis. Stat. Oper. Res. T., 33(1): 3-30.

Cheng-Lung, H., C.H. Mu-Chen and W. Chieh-Jen, 2007. Credit scoring with a data mining approach based on support vector machines. Expert. Syst. Appl., 33(4): 847-856.

Clausius, R., 1850. On the motive power of heat and on the laws which can be deduced from it for the theory of heat. Ann. Der. Phys. Chem., 79(1850): 368-524.

Ji-Sheng, X., L. Zhan, W. Daoheng and L. Peng, 2002. An approach to the forecast and evaluation of the credits for mobile clients. Chin. Telecommun. Sci., 18(8): 17-20.

Jonathan, N.C., D.B. Edelman and L.C. Thomas, 2006. Developments in consumer credit risk assessment. Eur. J. Oper. Res., 183(3): 1447-1465.

Jun-Yan, C.H., X. Ji-Sheng and L. Zhan, 2005. An application of BP neural network technology on the credit analysis for mobile clients. J. Wuhan Univ. Nat. Sci. Ed., 51(1): 114-118.

Laidler, K.J., 1995. The Physical World of Chemistry. Oxford University Press, UK.

Nan-Chen, H. and H. Lun-Ping, 2010. A data driven ensemble classifier for credit scoring analysis. Expert. Syst. Appl., 37(1): 534-545.

Perrot, P., 1998. A to Z of Thermodynamics. Oxford University Press, UK.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J., 27: 379-656.

Telecom-World, C.H.N., 2001. Note on China's mobile telephone industry. Chin. Telecommun. World, 6: 17.

Yang, L. and Y. De-Qi, 2005. Report on the arrearage problem. Chin. Comput. World, 24: E20-E20.

Yueh-Min, H., H. Chun-Min and J.C. Hewijin, 2006. Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. Nonlinear. Anal. Real., 7(4): 720-747.

Zong-Chang, Y. and K. Hong, 2008. Credit evaluation for mobile customers using artificial immune algorithms. Proceedings of the IEEE 7th World Congress on Intelligent Control and Automation, Chongqing, China, July 25-27, pp: 7021-7025.