Research Article
Evaluating Connection in Internet of Things Using Big Data Fusion Pattern

Xu Hong and Hu Ruo

Guangdong Polytechnic Normal University, 510665 Guangzhou, People's Republic of China

Abstract

Objective: The purpose of the study is to evaluate the relationships among nodes in the network. **Materials and Methods:** The nodes of the internet of things like mix-linker currently are one of successful internet services just after large websites such as Baidu, Tencent and trading website like TaoBao. Internet of things is widely accepted gradually. **Results:** Hence, the main finding is that various patterns are exploited for big data evaluating, which the best method was performed to combine all of possible conditions and relationships along with using intelligent feedback model. The results shows that this study is efficiently to assess the relationship of a node with a new node using big data on mix-linker node. **Conclusion:** In this study, it got about 86% effective evaluating about the conclusion, which are get by using big data fusion patterns.

Key words: Internet of things, big data fusion, relationship evaluating, pattern, mix-linker node

INTRODUCTION

In 2005, Friendster and My Space internet of things nodes hold top spots with 6 and 5 millions of users, respectively. In the same year facebook node was launched.

The year of emerging rules for internet of things was 2006 when a large amount of node's information was in hands of these networks. Therefore, specific rules for keeping the information secret and types of relationships were regulated. Eventually year 2007 was the year of rapid increase of users and visitors of internet of things.

Internet of things is a structure consists of node or organizational groups. These groups are connected to each other by one or some dependencies[1].

For the 1st time, in 1961 after that the first internet of things node was set up to the internet address "Mixlinker.com". But after year 2003, business explosion in nodes of internet of things such as LinkedIn, Orkut and etc., made a great evolution in this area and prospered internet of things.

Cyberspace provides the possibility of the formation of new societies for users. All philosophers of social and cultural sciences have mentioned "being face-to-face", "number limitation" and "being based on emotional relationships but not rational" as fundamental features of the society.

Relationship evaluating is a subfield of assessing internet of things in which one should deduces or estimates a series of links that are not directly observable or do not exist about existing observations and links.

The networks are instantaneously growing by joining new individuals or by creation of a new connection between existing ones in the network. One of the major focuses in assessing these types of networks is the evaluating of relationships between individuals within the network[2]. Today, the popularity of internet of things among node is undeniable. The internet of things provide users with a lot of facilities for communications.

In April 2009, facebook introduced its key web pages in different languages which propelled the site to 152% growth relative to past year.

Generally, assessing relationships includes the following fields:

- Regression of relations
- Categorization and classification of relationships, which means what kind of relationship exists between nodes
- Relationship assessing existence. That is to evaluate whether exists a connection between two arbitrary nodes

MATERIALS AND METHODS

Proposed pattern: In the beginning for recognizing the type of data and the circumstances of ROC rating, this study generated a series of data and get conclusion as this study mention here. In this study, it exploited ROC curves to compute the validity of evaluating values. The ROC is a strong simulation tool, which is used in medical decision making, psychology, communications and whenever need for threshold values is concerned[3].

In the learning data, friend A is related with friend B, so it could be assess that friend B has relationship with friend 1 as demonstrated in Fig. 1. Exploring in learning test data, this study recognized that about 400 numbers of data exist in the test data which satisfy this condition. Then in continue, this study presumed that "Friend A" represents column-1, "Friend B" represents column-2 and evaluating conclusion of relationship between friend 1 and 2 forms the 3rd column and this study knows that training data of the 1st and the 2nd column certainly related to each other.

Another pattern this study used in assessing relationships is based on the following principle.

With the condition that if the paired nodes (a, b) and (a, c) are related and there exists a relationship between them, then nodes (b, c) are related as well. This means that friend 1 has relationship with friend B and C has relationship with friend A, so one can conclude that it is probable that friend B has relationship with friend A, too. Using this pattern ROC conclusion was get 0.43. If there is a relationship between paired nodes (a, b) and (a, c) and also there exists a relationship between them, then the probability of existence of a relationship between nodes (b, c) in future is very high. This is depicted in the Fig. 2.

As demonstrated in Fig. 3, if relationships (a, b) and (b, c) exist, then this study creates the relationship (a, c) as well. That is friend 1 has connection with friend 2 and friend 2 has connection with friend 3, so it could be said that it is likely that friend 1 and friend 3 have relationship. With this pattern ROC conclusion was 0.485. In the next model this study actually finds a route which includes only three nodes and then this study surmises that the starting point and ending point has direct relationship with each other.
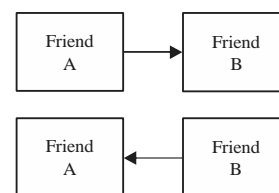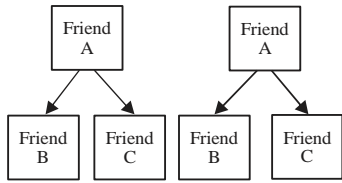

Fig. 1: Friendship relation
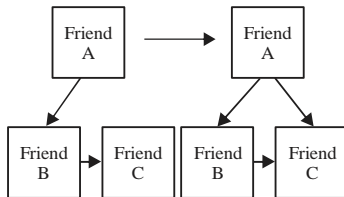
Fig. 2: Relation between nodes

Fig. 3: Friend connection

In the next step, this study considered the all patterns had realized and puts the all relations evaluating all ones together and as a consequence recorded the final ROC which get 0.77. About Fig. 4 in the followed step, supposing the relationships between paired nodes (a, b), (b, c) and (c, d), then it is highly probable that a connection exists between paired nodes (a, d) and (a, c). Using this pattern this study gets 0.525 for ROC conclusion.

Data evaluating: Whenever this study wants to specify a relationship within a collection of X's with a dependent variable Y and is confronting a multivariate problem. In assessing such focus different mathematical patterns are exploited. Intelligent feedback is a mathematical model which could be used for describing relationship between a No. of variables X's and a two-state or multi-state dependent variable as Y, naming two-state variable is a variable that have just two answers like dying or surviving, being present or absent and having relationship or not having relationship. Often binary codes are used for such variables. Code "1" is used for positive state (success) if that feature and code "0" for negative state (failure).

Today, in most studies, this study is seeking a specific aim using several factors such that yields to the optimum value.

In statistics, such studies are realized by different regression patterns and conclusion will be analyzed. In regression response variable is estimated by using.

Independent variables and this variable is the key objective of the most study[4]. As described earlier, regression patterns are applied based on type of the factors in studies. Intelligent feedback is a particular type of regression which is used in cases that response variable is double-choice or
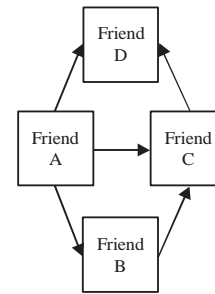

Fig. 4: Relation a node with other nodes

multiple-choice, that is only two or a few different states exist for response variable. This case is often used in medical and sociological study[5].

The essential subject in regression topic is finding a relation between response variable Y and a set of evaluating variables such as $X_1, X_2, \cdots$ and X_k. Actually regression technique is seeking to make a relationship like $Y = f(X_1, X_2, \cdots, X_k)$ between observations of Y and observation of $X_1, X_2, \cdots, X_k$. The simplest solution that one could imagine is the following linear Eq. 1:

$$Y = \alpha + \beta_1 X_1 + \ldots \tag{1}$$

As could be considered, because of these conditions the linear model is not always effective and for different data proper models should be used.

In this case, it is not possible to use linear model in (1), because left hand side of Eq. 1 arbitrates only values 0 and 1, whilst the right hand side could theoretically has any values form $-\infty$ to $+\infty$. Logical regression is a proper solution for these kinds of situations. Where β_0 coefficients are usually estimated by using one sample and with the aid of an estimation pattern like MMSE1 pattern. Although this requires applying some conditions on response and evaluating variables. For example, presumptions of model linearity, observations independency, normality of response variable distribution and stability of response variable variance must be hold. Sometimes, response variable is a two-state variable. On the other hand, evaluating variables that their effect on response variable could be evaluated are quantitative. In this pattern, the left-hand side of the equality is converted to a quantitative variable. This is carried out in three steps:

- Instead of using direct probability (Pr), its equivalent concept, "Odds ratio" is used. Note that probability p = 0.9 could be expressed in form of 9-1 or OR = p/p–1 = 0.9/0.1 = 9. It's obvious that if p = 0 then OR-0 and if p = 0.5 then OR equals one

- Deriving the natural logarithm from new response variable OR in order to range of new response variable varies from $-\infty$ to $+\infty$. In fact, $\ln(0) = -\infty$, $\ln(1) = 0$ and $\ln(+\infty) = +\infty$. It worth to mention that $\ln(p/1-p)$ is abstractly called logit (p)
- Substituting Y with the term Pr [Y = 1] in Eq. 1. Apparently, this probability could have any value between 0 and 1

In this case the new model given by Eq. 2:

$$\text{logit}\left(\frac{p_r[Y=1]}{1-p_r[Y=1]}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \qquad (2)$$

To estimate coefficients in (2), a random sample of long "n" is selected and for those values of response variable and evaluating variables are evaluated. This study gets ROC = 0.76 for all of previously evaluating models. This study used binary intelligent feedback for un-evaluating states and the ones this study could not able to make a condition, for which this study gets ROC = 0.88.

Hence for the sum of n observations of evaluating variable, there exist J different patterns (j = 1, 2, ⋯, J) so that for the j'th pattern of evaluating variables, there are mj No. of observations and the probability that j'th pattern contains Y = 1 equals Eq. 3:

$$\pi_j = \frac{e^{\beta + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj}}}{1 + e^{\beta + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj}}} \qquad (3)$$

Thus the logarithm of cofficient's likelihood function $\beta = (\beta_0, \beta_1, ..., \beta_k)$ is given by Eq. 4:

$$L(\beta) = \ln L(\beta) = \sum_{j=1}^{J} yj \ln(\pi j) + (mj - yj) \ln(1 - \pi j) \qquad (4)$$

In which yj denotes sum of observations for the j'th pattern. To find the maximum likelihood that get by maximization of equity 4 with respect to β, this study should solve the following Eq. 5, involves k+1 equation and k+1 variable, with respect to β:

$$\sum_{j=1}^{J} xji(yi - mj\pi j) = 0 \qquad (5)$$

Equation 5 are nonlinear with respect to $\beta_0, \beta_1, ..., \beta_k$, therefore, iterative numerical patterns are used to solve them.

RESULTS

Fliker is a huge internet of things having 33 millions of users and 32 billion of photos. This site is full of relationship data, including node's comments, group memberships, friend suggestions, clicking on favorite's photo and restricting the visit to some of the friends and families.

The file is "Social_Test.zip" which includes 8856 records and three columns, like the first file, it has two columns of 1st and 2nd node and 3rd column is the evaluating column which represents whether or not the first node and the second node are friends. These columns are filled with 0 and 1, value 1 in the case of relationship existence and value 0 otherwise. These data have been collected from December, 2011 to January, 2012.

The 1st file, "Social_Train.zip", contains of 6,357,612 records with two columns of first node and second node. These columns are filled numerically which denotes node unique number that assigned to a node within the whole data. There are 2,543,674 different individuals in the data. Each column shows that the first node is friend with the second node.

The other studied was that, it's probable to get better ROC by swapping evaluating value (0-1 and 1-0) in the evaluating column not having good ROC. This study assumed one of earlier evaluating in which ROC was get 0.476 and about new conditions per theories, this study evaluating that ROC conclusion will be improved by inversing data. So, by submitting the inverted data, ROC conclusion increases to 0.552, which this conclusion supports our hypothesis. This conclusion in ROC represents that the 2 Evaluating columns are complements and the sum of 2 ROCs equals one. Considering the above contents this study can say that of the 8582 test data, 4542 data are 1 and there exist 4246 data with 0 value.

Firstly the evaluating column is loaded with a series of random numbers between 0 and 1. The ROC was got about 0.436 using these numbers. Then in the 2nd step, this study filled half of the evaluating columns of the test file with one and the other half with zero, which concluded in ROC of about 0.48. In the 3rd step, this study filled all the data with zeros and after submitting it, ROC was get equal to 0.6. In step 4, all data of evaluating columns were filled with ones and again ROC = 0.6 was get after submission. About the last 2 steps, it can be concluded that the No. of one's or zeros, i.e., having relationship relation between two individuals are equal in the whole test data.

Up to this point this study has gathered some information about type of data and their way of implementation, as this study mentioned above. In the following this study presents a few models which could be helpful in identification relationship between two individuals.

DISCUSSION

Zheng and Bundell[6] have presented a pattern in mobile internet of things, manner in which it spoke of the relationship between the numbers of mobile users in the physical world to discover internet of things. Nowadays most works are focused on studying and assessing internet of things graphs. Many efforts have carried out to solve the problem of exclusive evaluating of internet of things[7,8]. Although most primary studies on Internet of things are done by social sciences scholars and the psychologists, numerous efforts by computer scientists are performed recently[9].

In the present study, supervised patterns are used to relationship evaluating. Gao, J. and Shih detailed the challenges that a relationship exists in the evaluating system[6] were analyzed. This study discussed imbalance focus and proposing to treat evaluating separately for different classes of potential friends[10].

One of the other study done by Hu[11] is introducing an item called sub-graph relationship. A small sub-graph is the best relationship in internet of things. They also proposed an effective algorithm based on electronic rules, which finds sub-graphs connections in large internet of things. Number of sub-graphs could be used in computation of various values for solving the evaluating problem of internet of things, especially when the networks are very large[11].

The target of this study is to evaluating relationship with high probability. This evaluating helps internet of things nodes a lot in finding out the existence of a relation between two individuals. To do so, this study used the data of Kaggle competition site which have been collected from fliker social site[12].

CONCLUSION

From total number of 8582 data this study have correctly evaluating about 86% of data and this model gives us the best conclusion for evaluating of zeros and ones value. In this study, this study exploited various patterns for big data evaluating, which the best conclusion was to combine all of possible conditions and relationships along with using intelligent feedback model.

To resolve this one may put weights to presented models in a way that the more likely the model is correct, the more it is weight is assigned and eventually sum of all represents the effective evaluating probability of relationship relation. Conclusion of model X: Z = weight of X* conclusion of model X. About get data from aggregation of previously presented models and considering number of ones and conclusion yielded from playing with numbers, the number of assessed ones is more than expected in relationship.

ACKNOWLEDGEMENT

REFERENCES

1.   Amid, A. and S. Moradi, 2013. A hybrid evaluation framework of CMM and COBIT for improving the software development quality. J. Software Eng. Applic., Vol. 6. 10.4236/jsea.2013.65035.

2.   Binder, R.V., 1994. Design for testability in object-oriented systems. Commun. ACM., 37: 87-101.

3.   Gao, J. and M.C. Shih, 2005. A component feasibility method for verification and measurement. Proceedings of the 29th Annual International Computer Information System and Applications Conference, July 26-28, 2005, Edinburgh, pp: 211-218.

4.   Lee, M.C., 2014. Information system efficiency elements and information system efficiency rules to enhance information system efficiency assurance. Br. J. Applied Sci. Technol., 4: 3069-3095.

5.   Fu, J.P. and M.Y. Lu, 2009. Request-oriented method of information system feasibility measurement. Proceedings of the ITCS 2009 International Conference on Information Technology and Computer Science, July 25-26, 2009, Kiev, pp: 77-80.

6.   Zheng, W.Q. and G. Bundell, 2008. Contract based information system component measurement with UML methods. Proceedings of the International Symposium on Computer Science and its Applications, October 13-15, 2008, Hobart, Australia, pp: 83-102.

7.   Huda, M., Y.D.S. Arya and M.H. Khan, 2015. Evaluating effectiveness factor of object oriented design: A testability perspective. Int. J. Inform. Syst. Eng. Applic., 6: 41-49.

8.   Esposito, D., 2008. Design your classes for ascertainable. http://dotnetslackers.com/articles/nnet/Design-Your-Classes-for-Ascertainable.aspx

9. Huda, M., Y.D.S. Arya and M.H. Khan, 2015. Quantifying reusability of object oriented design: A testability perspective. J. Inform. Syst. Eng. Applic., 8: 175-183.

10. Hu, R., 2012. Channel access controlling in wireless sensor network using smart grid system. Applied. Math. Inform. Sci., 6: 813-820.

11. Hu, R., 2012. Stability analysis of wireless sensor network service via data stream methods. Applied Math. Inform. Sci., 6: 793-798.

12. Hu, R. and J.H. Guo, 2013. New network access control method using intelligence agent technology. Applied Math. Inform. Sci., 7: 44-48.