ARTICLE IN PRESS

Engineering Applications of Artificial Intelligence ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Multiple-negative survey method for enhancing the accuracy of negative survey-based cloud data privacy: Applications and extensions ☆

Ran Liu [a,b], Jinghui Peng [a], Shanyu Tang [a,b,*]

[a] School of Computer Science, China University of Geoscience, Wuhan, Hubei 430074, China
[b] Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 9 January 2016
Received in revised form
10 May 2016
Accepted 6 June 2016

Keywords:
Artificial immune system
Cloud data privacy
Multiple-negative survey
Confidence level
Bayes method
Anonymity vote model

ABSTRACT

Cloud computing brings convenience to people's lives because of its high efficiency, usability, accessibility and affordability. But the privacy of cloud data faces severe challenges. Although negative survey, which is inspired by Artificial Immune System (AIS), can protect users' privacy data with high efficiency and degree of privacy protection, its accuracy is influenced by the number of client terminals, and insufficient client terminals may lead to large errors. This study focuses on a multiple-negative survey method of remedying this weakness. Compared with the traditional negative survey method, the multiple-negative survey method collects each user's multiple different negative categories rather than only one negative category. Two key scientific problems (accuracy and confidence level) are analyzed, and an application (anonymity vote model) is then proposed based on the multiple-negative survey method.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Inspired by negative selection principle (Hofmeyr and Forrest, 2000), which is an essential mechanism of Artificial Immune System (AIS), the negative selection algorithm (Forrest et al., 1994) has been proposed and used in network security and virus detection. The negative selection algorithm can generate a set of detectors unmatched by itself. If a sample is matched by a detector, the sample is marked as "nonself", otherwise it is marked as "self". Negative representation (Esponda, 2008), which is inspired by the negative selection algorithm and Artificial Immune System (AIS), is a new kind of information representation. Different from the general information representation, negative representation stores the contents not consistent with the real information. Existing work (Esponda et al., 2004, 2005, 2007) showed that reversing a negative representation (such as negative databases) to get the original information equals to solve a SAT formula. Based on this kind of property, negative representation can be used for information security (Esponda et al., 2007). As shown in Zhao et al., negative representation can be used for iris recognition.

In recent years, users face cloud data privacy protection problems with the advent of cloud computing and intelligent computing techniques. Cloud data privacy protection affects the high efficiency of cloud computing to a certain degree. So reducing the amount of background calculations to the cloud data privacy protection in client terminal is an urgent issue. Negative survey (Esponda, 2006; Esponda and Guerrero, 2009), which is inspired by the negative representation of information, could protect the privacy of participants effectively while collecting information. In Horey et al. (2007), negative survey is used for anonymous collection of traffic behavior. Furthermore, negative survey method (Esponda et al., 2016) allows each participant to select different number of negative categories to customize its own privacy degree. Negative surveys only collect parts of the negative categories, so this method can increase operating speeds by saving the tedious encryption process. Meanwhile, how to enhance the accuracy (Bao et al., 2013) of converting negative survey results into positive survey results is one of the key issues in negative surveys.

The main research (Esponda and Guerrero, 2009; Horey et al., 2007; Bao et al., 2013, 2014; Xie et al., 2011; Lu et al., 2014; Liu et al., 2015) of present work focuses on the traditional negative survey (symbolized as 1-NS). Only limited work (Esponda et al., 2016; Bao et al., 2014) discusses the multiple-negative survey. In consequence, the content of this study is enhancing the accuracy of negative survey-based cloud data privacy by multiple-negative survey, i.e. each participant selects multiple different negative

☆An earlier abbreviated version of this paper was presented as the ICSI3 2015, Taormina, Italy, 17–18 July 2015.

* Corresponding author at: School of Computer Science, China University of Geoscience, Wuhan, Hubei 430074, China.
E-mail address: shanyu.tang@gmail.com (S. Tang).

http://dx.doi.org/10.1016/j.engappai.2016.06.002
0952-1976/© 2016 Elsevier Ltd. All rights reserved.

ARTICLE IN PRESS

categories randomly.

In the remainder of this study, Section 2 introduces the related work of this study. Section 3 describes the multiple-negative survey and analyzes the accuracy of the positive survey reconstructed by the multiple-negative survey. Section 4 approximately calculates the confidence level of the multiple-negative survey with Bayes method. Based on the multiple-negative survey, Section 5 proposes an anonymity vote model to protect the privacy of each voter. Section 6 summarizes the conclusions and future work.

2. Related work

In this section, the related work of negative survey (Esponda, 2006; Esponda and Guerrero, 2009) is introduced. Some definitions are described in Fig. 1 for convenience.

Define n to be the number of users who use negative survey method to send their privacy data, and c to be the number of categories. The results of the privacy data collected in cloud platform are $R = (r_1, r_2, \ldots, r_c)$, where r_i $(1 \le i \le c,\ \ c \ge 3)$ represents the total number of users who send the i-th category to the cloud platform. Similarly, the real privacy data is $T = (t_1, t_2, \ldots, t_c)$, and $n = \sum_{i=1}^{c} r_i = \sum_{i=1}^{c} t_i$. In Esponda (2006) and Esponda and Guerrero (2009), the reconstructed positive survey of privacy data can be calculated by the following formula:

$$\hat{t}_j = n - (c-1)r_j \tag{1}$$

Although $\hat{t}_j = E(t_j)$, it can be observed that $\hat{t}_i < 0$ when $r_i > n/(c-1)$. Therefore, this traditional method is not practical sometimes, and two methods (Bao et al., 2013) were proposed to solve the negative value problem.

In Bao et al. (2014), it discusses the confidence level of multiple-negative survey. But it only gives the complicated formulas by generation function method by analyzing the relationship between multiple-negative survey and traditional negative survey.

This study analyzes the accuracy of multiple-negative survey. It indicates that the multiple-negative survey method can enhance the accuracy of negative survey-based cloud data with less number of client terminals. The confidence level is calculated using Bayes method and an anonymity vote model is proposed to protect the privacy of each voter.

3. Accuracy of multiple-negative survey method

The accuracy is regarded as a key scientific problem of multiple-negative survey method. Obviously, a client terminal sends k repeatable negative categories (from $c-1$ negative categories) which can be treated as each client terminal (k client terminals in all) sends a negative category independently. So this section focuses on each client terminal sending multiple different negative categories. Section 3.1 describes the definition of multiple-negative survey, and Theorem 3.1 in Section 5.1 analyzes the accuracy of data reconstructed by multiple-negative survey, i.e. each client terminal sends multiple different negative categories.

3.1. Derivation of formulas

In a multiple-negative survey, each client terminal sends k $(k \le c-2)$ different negative categories is defined to be k-NS, and the traditional negative survey is 1-NS. In this study, a positive survey with n interviewees and c categories is written as $PS(n, P)$, where $P = (p_1, p_2, \ldots, p_c)$ is the proportion vector of categories. And the corresponding multiple-negative survey (of $PS(n_k, P)$) is written as k-NS(n_k), and $Q_k = (q_{k,1}, q_{k,2}, \ldots, q_{k,c})$ is the proportion vector

n : the number of users (or client terminals)

c : the number of categories in surveys

r_i : the number of users sending negative category i

t_i : the original number of users in positive category i

$\hat{t}_i$: the estimated value of t_i

R : the negative data vector, $i.e.$ $R = (r_1, r_2, \cdots, r_c)$

T : the user data vector, $i.e.$ $T = (t_1, t_2, \cdots, t_c)$

p_i : the proportion of positive category i, i.e. $p_i = t_i/n$

$\hat{p}_i$: the estimated value of p_i

Fig. 1. The definitions of traditional negative survey (1-NS).

of the k-NS. Different from 1-NS, $q_{k,i}$ in k-NS is $r_{k,i}/kn_k$ to make $\sum_{i=1}^{c} q_{k,i} = 1$, because each user sends k different negative categories so that $\sum_{i=1}^{c} r_{k,i} = kn_k$. Function A(k-NS) is used to define the accuracy of positive survey reconstructed by k-NS. For convenience, some definitions are in Fig. 2.

Similar with Esponda (2006) and Esponda and Guerrero (2009), the reconstructed positive survey by k-NS can be calculated by Formula (2). Because the accuracy of each reconstructed positive category is influenced by $q_{k,j}$, the accuracy of k-NS can be measured by the vector Q_k:

$$\begin{cases} \hat{t}_j &= n_k - (c-1)r_{k,j}/k \\ \hat{p}_j &= 1 - (c-1)q_{k,j} \end{cases} \tag{2}$$

In this subsection, the accuracy of multiple-negative survey method is analyzed in Theorem 3.1. It proves that the accuracy of 1-NS (n) is no higher than that of k-NS (n/k).

Theorem 3.1. *For the original positive survey $PS(n, P)$, the accuracy of the 1-NS (n) is no higher than that of the k-NS (n/k). It can be quantified as*

$$A(1 - NS(n)) \le A(k - NS(n/k)) \tag{3}$$

n_k : the number of users sending privacy data

c : the number of categories in surveys

k : the number of negative categories to be sent

$r_{k,i}$: the number of users sending negative category i

R_k : the vector of user number in each category, i.e. $R_k = (r_{k,1}, r_{k,2}, \cdots, r_{k,c})$

$q_{k,i}$: $q_{k,i} = r_{k,i}/kn_k$, $\sum_{i=1}^{c} q_{k,i} = 1$

k-NS(n_k): each user (n_k total users) sends k categories in the multiple-negative survey

Q_k : the negative vector of k-NS, i.e. $Q_k = (q_{k,1}, q_{k,2}, \cdots, q_{k,c})$

A(k-NS): the accuracy of the positive survey reconstructed by k-NS

Fig. 2. The definitions in multiple-negative survey.

Please cite this article as: Liu, R., et al., Multiple-negative survey method for enhancing the accuracy of negative survey-based cloud data privacy: Applications and extensions. Eng. Appl. Artif. Intel. (2016), http://dx.doi.org/10.1016/j.engappai.2016.06.002

Download English Version:

https://daneshyari.com/en/article/4942757

Download Persian Version:

https://daneshyari.com/article/4942757

Daneshyari.com