Mining Fraudsters and Fraudulent Strategies in Large-Scale Mobile Social Networks

ABSTRACT

The rapid development of modern communication technologies—in particular, (mobile) phone communications—has largely facilitated human social interactions and information exchange. However, the emergence of telemarketing frauds can

significantly dissipate individual fortune and social wealth, resulting in potential slow down or damage to economics. In this work, we propose to spot telemarketing frauds, with an emphasis on unveiling the “precise fraud” phenomenon and the strategies that are used by fraudsters to precisely select targets. To study this problem, we employ a one-month complete dataset of telecommunication metadata in Shanghai with 54 million users and 698 million call logs. Through our study, we find that user’s information might has been seriously leaked, and fraudsters have preference over the target user’s age and activity in mobile network. We further propose a novel semi-supervised learning framework to distinguish fraudsters from non-fraudsters. Experimental results on a real-world data show that our approach outperforms several state-of-the-art algorithms in accuracy of detecting fraudsters (e.g., +0:278 in terms of F1 on average). We believe that our study can potentially inform policymaking for government and mobile service providers.

**EXISTING SYSTEM**

In many literatures, fraud detection is formulated as a binary classification problem. That is, given a set of phone numbers, predict whether each number is normal or fraudulent. For example,Weatherford et al. [23] utilize user profiles that store long-term information and train neural networks to differentiate fraud behavior and normal one. Instead of neural networks, Yusoff [4] propose a model based on Gaussian mixed model (GMM) as the classifier. Dominik uses a threshold-type classification algorithm [24]. The major limitation of classifier-based methods is that, its performance is heavily influenced by annotations, and will be hurt when the label is sparse. In this work, we propose a semi-supervised learning framework to further utilize unknown labels and improve the performance.

Disadvantages

1) An existing system doesn’t predict fraudulent strategy explicitly.

2) An existing is absent in Distinguish Fraudsters from Others.

**PROPOSED SYSTEM**

Based on our discoveries, we design a novel factor graph based model, FFD, to distinguish fraudsters. More specifically, our model incorporates fraudsters’ structural information and preference on choosing targets. We further propose a semi-supervised learning framework to utilize both the known and unknown labels and address the label sparsity challenge. According to our experiments, we see that our model achieves an improvement on F1 of 0.278 comparing with several state-of-the-art methods.

It is worthwhile to highlight our contributions as follows:

\_ Based on a real phone-communication data, we disclose how fraudsters and non-fraudsters behave differently in mobile network.

\_ The system also studies the “precise fraudulent strategy” and appeal to everyone to make sure the protection of personal information has been brought to the forefront.

\_ The system proposes a novel framework to distinguish fraudsters from others in a given mobile network.

\_ The system validates the effectiveness of our model on a larges scale mobile network in real world.

**Advantages**

* According to a user’s calling logs, we define energy dispersion, the proportion of a user’s energy invested in relationship with one of her contacts, as the proportion of times she calls a particular contact.
* The system resists Fraudsters which choose target not randomly, but have preference over users’ age and activity in mobile network.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).