KDD 2018 论文阅读报告

任恒　计算机科学与技术 16-3 班　2016212063

-- Who Stole the Postage? Fraud Detection in Return-Freight Insurance Claims

Chen Liang　Ziqi Liu　Bin Liu

Ant Financial　Ant Financial　Ant Financial

Hangzhou, China　Hangzhou, China　Hangzhou, China

lc155190@antfin.com ziqiliu@antfin.com lb88701@alibaba-inc.com

Jun Zhou　Xiaolong Li

Ant Financial　Ant Financial

Beijing, China　Seattle, USA

jun.zhoujun@antfin.com　xl.li@antfin.com

摘要

网上购物的出现已经导致了一种被称为运费险的新型保险。它提供了退货邮费赔偿机制来解决电子商务中因退货导致的买家与卖家之间的争执。然而，故意滥用保险政策可能会导致沉重的经济损失。为了预防欺诈性的保险索赔，我们开发了一个全新的数据驱动的程序识别欺诈性账户来帮助预防在索赔阶段经济损失。

在这篇论文中，我们介绍了一个索赔人之间的设备共享网络，然后开发了基于图像学习算法的欺诈检测自动化解决方案。这种解决方法已经应用在阿里巴巴并且在人类专家调查后确认与之前的基于规则的分类器相比精度超过 80%，同时覆盖了超过 44%的可以账户。我们的方法能够很容易的推广到其他类型的保险上。

关键词

欺诈检测，图像学习，网络学习，保险欺诈

1 介绍

如果你买了一条连衣裙但是发现实际的产品颜色和屏幕上的产品有很大的颜色上的区别会怎么办？如果你在购买笔记本后发现了一件更便宜的同样产品会怎么办？如果你花了大价钱事后却对你的冲动购买行为感到后悔会怎么办？在网上购物的时候退回一件未使用过的物品会使卖家和买家之间因为模糊不清

的责任划分而产生很大的争执。令人惊喜的是大多数的争执焦点不是在是否这件未受损的产品时候应该被退回，而是在谁应该支付退货导致的邮寄费用。解决这类纠纷耗费了大量的时间以及精力，对于阿里巴巴[1]这种拥有数以百万的卖家以及多种多样的退货政策的公司来讲更是如此。为了解决争执以及维护购买者后悔的权力，一个新兴的保险组成已经出现了。

退货运费险，主要是用来再退货的时候赔偿给买家邮寄费用，已经有了数十亿美元的收入。但是诈骗造成的损失并非是微不足道的。根据阿里巴巴保险专家的评估，成千上万的可能是诈骗索赔在以往的基于规则的欺诈监测系统中是没有被发现的。对于一个更智能、灵活的欺诈监测解决方法的需求越来越重要。

1.1 我们的欺诈监测问题

在保险索赔上的欺诈监测可以被视为一个有监督的二分类问题。我们将保险账户分为两类：欺骗和定期。训练集上的账户标签是从早先部署在基于规则的系统中获得的，其中部分是不充分可信的。我们致力于在保持高准确率的基础上比基于规则的系统发现更多的欺骗性账户。

Networks[2]在为描述和建模同伙（合作欺诈人员）间的复杂关系提供了直接的信息。为了挖掘这种信息，我们提供了一个共享图工具，一个事务图还有一个友谊图去解释这种关系，并且用两个图学习方法，一个基于 node2Vec[8]，另一个基于 GeniePath[9]。我们指导拓展的实验来比较这些方法并且描述我们实现设备共享图和 GeniePath 的完整的诈骗检测解法。

1.2 诈骗监测中的挑战

我们发现阻碍欺诈监测系统的表现的因素包括**概念漂移、标签不确定性以及过度的人力。**

概念漂移是指在欺诈监测中新的欺诈类型随着时间的推移对欺诈检测系统[1]来说越来越不可预测的现象。账户的非固定行为、从保险索赔历史、退货历史以及购物历史中提取信息是导致概念漂移的主要原因。一些系统通过对这些非固定行为使用自适应学习算法[21]解决概念漂移问题；我们通过添加更多固定化的关系来解决这个问题。诈骗合作者之间的关系使用一个共享设备图以及一个使用图像学习算法建模更自然地阐述。

标签不确定是指因为规则产生的标签的使用而导致的。传统的部署的基于规

则的欺诈监测系统为每一个账户输出一个风险等级，可以为“高风险”，“低风险”，或者是“没有监测到风险”。我们对“高风险”的账户可以很确定为欺诈账户，到那时对“没有监测到风险”的账户是否处于风险之中是不清楚的。换句话说，在这些标签中由少量的确定有风险的标签一个大量的不确定的标签组成。为了建立我们的训练集，我们随机的从“没有监测到风险”的类中抽取样本，这将在数据准备章节中进行解释。

大量的人力是来自于传统诈保监测设置中的标签任务以及评估任务。作为一个金融科技公司，蚂蚁金融[3]注重于自动的风险控制而人力的工作可以忽略不计。我们的方法要求在一个周期的评估中（每周或者甚至一个月）除了保险专家的指导样本以及为损失估计评估分类结果之外没有人工的干预。

2 相关工作

诈保监测方法可以被一般的分为监督学习，无监督学习以及两者都有的方法。比较受欢迎的监督算法例如神经网络，支持向量机，回归模型，Bayesian 网络以及决策树已经被应用或者整合使用[2-4,7,10-16,19]，无监督方法例如聚类分析、异常值检测以及尖峰监测也已经被应用[5,18]，混合监督算法和无监督算法也已经在研究之中并且无监督方法已经被用于将保险数据划分为集群应用监督学习方法[6,17]。

我们的两种方法分别属于监督学习以及两者的混合。我们的方法不同，因为他们用图标表示数据，图标是最自然的表示数据的方法并且允许在没有简化数据的情况下进行复杂的分析。

3 构造图表

为了解决概念漂移同时发现有组织的欺诈团伙，我们借助图表的力量帮助揭示账户之间的强大关系。在这一节，我们构造并且比较不同类型图表，其中包括一张设备共享表，一张事务表以及一张友谊表。我们解释什么是好的图表并且适应我们的需要，最终集成设备共享图到我们的欺诈监测解决方案中。

3.1 好的图表

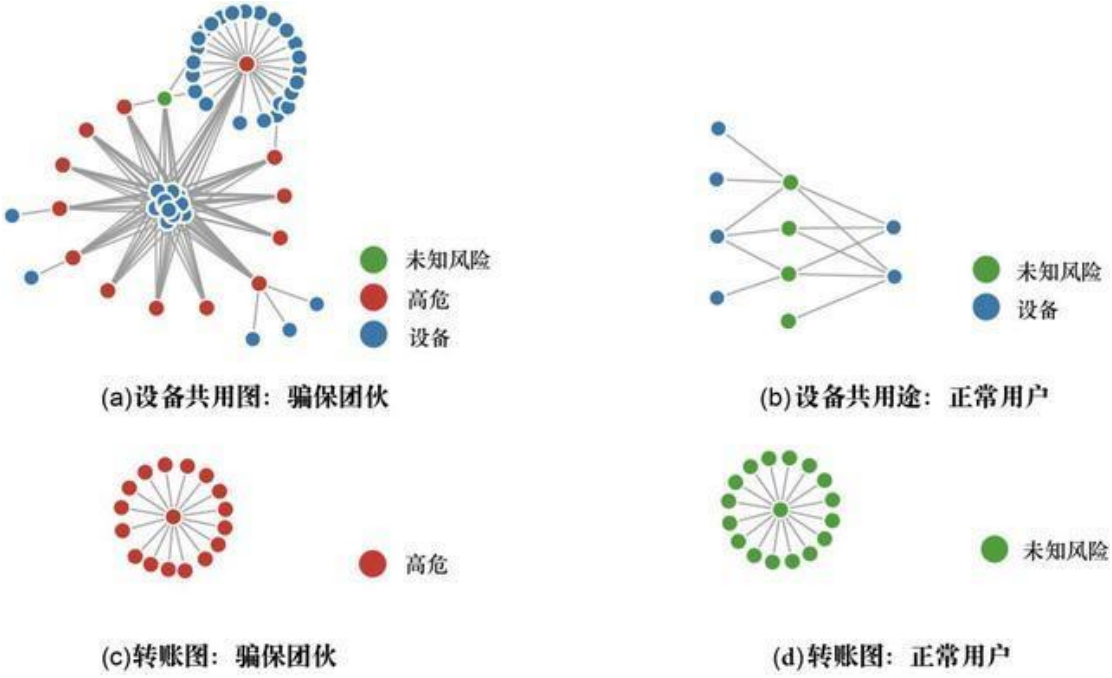
我们期望的图表应该能够将诈骗通过下列属性区分正常的账户。

（1） 距离聚合：距离比较近的两个节点应该具有相似的标签。

（2） 结构化不同：有组织的欺诈者的结构应该和正常账户的结构不同。

表一：对比图

图表	|V|	|E|	节点	边
设备共享	2M	2M	账户/UMID	设备使用
事务	2M	2M	账户	资金交易
友谊	8M	11M	账户	关系



图一：在设备共享图，事务图，友谊图中对典型诈骗伙伴以及正常用户进行可视化

3.2 三个图形

设备共享图揭示了账户之间共享一个设备的关系。一个顶点或者是一个设备（用户机器 ID，UMID[4]）或者是一个账户。边只存在于一个设备顶点和一个 UMID 顶点之间，表示历史记录之间的登录活动。**（个人理解为边其实是账户与设备之间的连线）。**

事务图展示了账户之间的资金交易关系。一个顶点是一个账户，一条边表示账户之间存在转账行为。

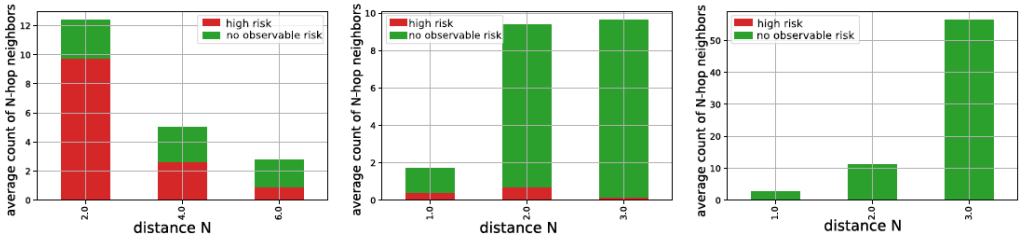
友谊图是建立在支付宝朋友关系的基础上，这是蚂蚁金融网络社交的产物。

我们预处理了这些表来删除孤立的账户。在事务图和关系图，节点度数（连接到该节点的边数）为零的将会被删除。在设备共享图中，账户节点，将删除没有和其他账户共享 UMID 的节点，并且他们的邻居节点也会被删除。

3.3 图表对比

典型的诈骗组织以及常规的用户的子图在图一中总结和可视化。诈骗团伙的设备共享图和友谊图与常规用户的展示是鲜明的对比。但是事务图在展示这些特性上是失败的。

此外，一个有效的图表需要区分欺诈账户与常规账户。特别是，相邻的节点应该有类似的标签。我们根据与欺诈节点的距离来衡量使用节点分布汇总欺诈帐户的能力。这种分布在表二中展示。诈骗账户在设备共享图中相互之间聚集，暗示着它更适合于账户分类任务。



(a) Device-sharing (b) Transaction (c) Friendship

Figure 2: Average number of N-hop neighbors around fraudulent accounts.

4 图学习算法

我们分别为诈保监测介绍两种基于节点嵌入和图神经网络算法的图学习算法。

4.1 基于节点嵌入的方法

4.1.1 节点嵌入：节点嵌入是一种低维度向量来表示图节点。节点被映射到嵌入，使得嵌入空间中的相似性近似于图中的相似性。

4.12 Node2vec：Node2vec[8]为每一个节点指定了一个独一无二的嵌入向量。它通过一个随机开始的节点$v_0 \epsilon V$明确诱导边的随机遍历，并且重复的取样一个边$(v_i, v_{i+1}) \epsilon \mathcal{E}[v_i]$。边的取样在 node2vec 中是有偏差的并且它可以在局部和全局视图中权衡。一个随机散步的自己被用来在同一个上下文中学习相似的嵌入节点。

4.1.3 我们的方法：Node2vec 是一个只可以使用图结构信息的无监督的算法。我们连接图嵌入法以及账户特征并且使用梯度提升决策树提供新的特征向量到下流分类任务（GBDT）[20].

4.2 图神经网络方法

4.2.1 图神经网络方法（GNN）：是一个使用神经网络来聚集邻居节点信息的

深度学习结构的集合。在神经网络中，一个较深的层次距离更远的节点并且第 K 层中的节点 V 是

$$h_v^k = \sigma(W_k \cdot AGG\left(h_v^{k-1}, \forall u \in \mathrm{N}(v) \cup \{v\}\right))$$

初始化嵌入层$h_v^0 = X_v$是账户的特征，σ是一个非线性的函数，AGG 是一个聚合函数，和 GNN 不同的是它跨越不同的层以及邻居。

4.2.2 GeniePath:我们使用的欺诈监测方法是基于 GeniePath 的[9]，简单地在图中堆叠自适应路径图层以进行宽度和深度探索。对于广度探索，它对邻居的聚类为

$$\mathrm{AGG}\left(h_v^k\right) = \sum_{u \in N(v) \cup \{v\}} softmax(W^T \tanh(W_s h_v^k + W_d h_u^k)) \cdot h_u^k$$

这种宽度搜索函数强调具有相似账户特征的邻居的重要性。

给出这些隐藏单元$(h_u^0, h_u^1, \ldots, h_u^k)$，一个深度搜索函数被增加到进一步提取和过滤不同深度的信号。这个嵌入结果被送到最终的 sftmax 或者 sigmoid 层为了下一层的欺诈账户的分类任务。

5.1 数据准备

我们的所有训练数据和测试数据都至少提出索赔 30 天。我们每周训练一次并且报告在后面数据中的分类措施。

对于每一个账户，我们搜集了 50 个特征（一个月内提交的申诉数量，作为一名顾客的持续时间，等等），来自保险索赔历史，运输历史以及购物历史。设备使用历史也被凑集来构建图结构。

被用来分类的标签是基于规则的“风险等级”指示符。我们视“高风险”账户为欺诈账户，“未监测到风险”的账户为正常账户。

然而，这个数据集经历了标签的不确定性—基于规则的风险指示符对于“高风险”的账户判断为欺诈账户比“未监测到风险”判断为正常账户更有把握。为了说明这个问题，这个在训练集中的“正常”类是随机被选择为样本的。下采样帮助降低被分类为“未监测到风险”的账户未欺诈账户的风险，正如下面修改的目标函数：

$$\mathcal{L}(\mathrm{w}) = \min_w(\sum_{v \in V_{fraudulent}} \ell(f(X_v; w), fraudulent)$$

$$+\sum_{v\in samole(Vregular)} \ell(f(X_v;w),regular)$$

我们的目标是最小化因为错误分类造成的损失。可以通过新目标函数的下采样率来调整惩罚误报的可能性。

5.2 底线

我们针对 GBDT 分类器评估了两种图学习方法。 它使用帐户功能作为输入，没有任何图形结构信息。 对于所有方法，我们计算测试数据集中每个帐户存在风险的概率，然后使用它们计算 F1 得分 5。

5.3 实验开始

我们使用的实现的 GeniePath，node2vec，以及 GBNT（Parameter Server-based Scalable Multiple Additive Regression Tree[20]）作为我们实现的人工智能蚂蚁金融平台的组成部分)。

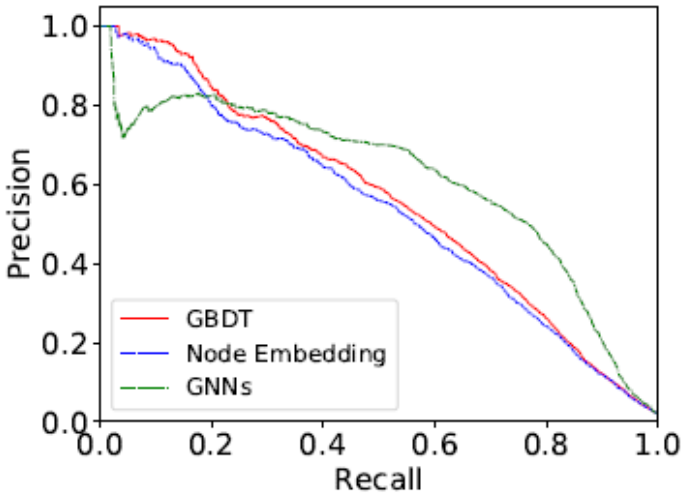
对于在这次实验中用到的所有的 GBDT 模块，我们设置这个超参数是相同的：500 个树，每棵树的最大深度未 5，数据样本率未 0.6 并且特征样本率未 0.7 来避免过拟合，学习速率为 0.009.我们随机选取 25%的“未监测到风险”的账户作为训练集的消极样本。

5.4 结论和讨论

我们的结论，在表 2 中总结并且在图 3 中绘制，结果显示基于 GNN 的图像学习算法的表现优于其他的。报告扩展（RE），定义为$\frac{FP+TP+FN}{TP+FN}$,表明检测欺诈账户的能力。F1 分数和 REs 被使用混淆矩阵计算出来，他的“基本事实标签”是建立在基于规则的账户风险指示符。我们所有的方法都将欺诈性帐户检测的覆盖率提高了 40％以上，而基于 GNN 的方法在大多数时间内具有更高的精确度和召回率。

表二　基于规则的标签结果

	GBDT	NODE	EMBEDDING GNNS
F1	0.547	0.535	0.623
RE	1.47	1.44	1.44



GDBDT 算法略好于节点嵌入算法。这个结果表明嵌入学习完全来自图的信息而不是账户的特征。我们找出最具有价值的特征来子购物历史-如果一个用户已经花费了很多在过去的一年里，那么我们很确信他不是一个诈骗人员。

6 应用

我们退货诈保监测的工作流程如图 4 所示。它搜集账户账户在过去几个月里已经提交的请求并且以批处理的模式对他们每天进行更新。分类的结果周期性的被保险专家评估以及监督。他们随机的抽样我们报告的诈骗账户并且最近的报告显示相比于以往的基于规则的分类器我们已经实现了超过 80%的准确率同时覆盖超过 44%的可疑账户。这个系统预计每月可以节省超过 1w 美元。

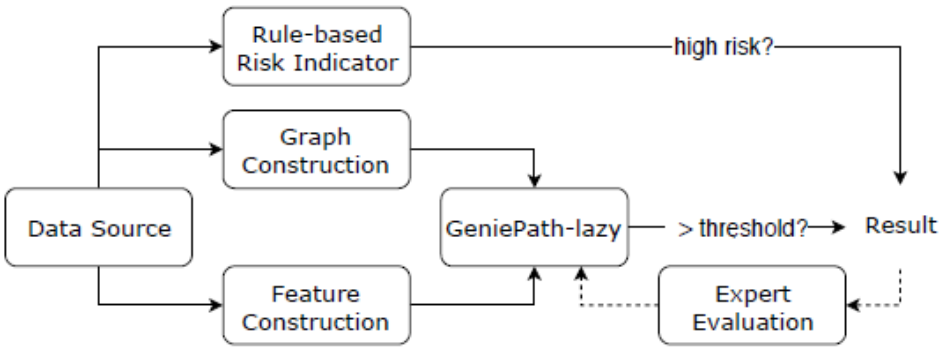


Figure 4: Workflow for fraud detection.

这个同样的设备共享图和类似的图学习算法已经被应用到其他多种在线保险例如蚂蚁金融健康保险。这个结果是可以保证的并且目前在进一步完善中。

7 结论

这篇论文提出设备共享图和图形基于学习的方法去解决在退货保险索赔中的欺

诈监测问题。这是在文献里第一次利用图标的强大表现力介绍一个现实世界中的诈保监测系统的论文。图的能力已经在多种在线保险领域被证明。设备共享图提供了更好的可视化和区分好的和坏的的能力。基于 GNN 的 GeniePath-lazy 方法通过为信息聚合选择更多有意义的接受途径比其他的表现更好。有了恰当的图，特征以及算法，我们通过自动化方法已经实现了超过 80%的精度和覆盖超过 44%的可以账户在一个诈保检测领域。

8 参考文献

[1] Aisha Abdallah, Mohd Aizaini Maarof, and Anazida Zainal. 2016. Fraud detection system: A survey. Journal of Network and Computer Applications 68 (2016),90–113.

[2] Manuel Artís, Mercedes Ayuso, and Montserrat Guillén. 2002. Detection of automobile insurance fraud with discrete choice models and misclassified claims.Journal of Risk and Insurance 69, 3 (2002), 325–340.

[3] Manuel ArtıÌĄs, Mercedes Ayuso, and Montserrat Guillen. 1999. Modelling different types of automobile insurance fraud behaviour in the Spanish market. Insurance: Mathematics and Economics 24, 1-2 (1999), 67–81.

[4] El Bachir Belhadji, George Dionne, and Faouzi Tarkhani. 2000. A model for the detection of insurance fraud. The Geneva Papers on Risk and Insurance-Issues and Practice 25, 4 (2000), 517–538.

[5] Patrick L Brockett, Richard A Derrig, Linda L Golden, Arnold Levine, and Mark Alpert. 2002. Fraud classification using principal component analysis of RIDITs.Journal of Risk and Insurance 69, 3 (2002), 341–371.

[6] Patrick L Brockett, Xiaohua Xia, and Richard A Derrig. 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud.Journal of Risk and Insurance (1998), 245–274.

[7] Štefan Furlan and Marko Bajec. 2008. Holistic approach to fraud management in health insurance. Journal of Information and Organizational Sciences 32, 2 (2008),99–114.

[8] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on

Knowledge discovery and data mining. ACM, 855–864.

[9] Ziqi Liu, Chaochao Chen, Longfei Li, Jun Zhou, Xiaolong Li, and Le Song. 2018. GeniePath: Graph Neural Networks with Adaptive Receptive Paths. arXiv preprint arXiv:1802.00910 (2018).

[10] Lindsay CJ Mercer. 1990. Fraud detection via regression analysis. Computers & Security 9, 4 (1990), 331–338.

[11] Thomas Ormerod, Nicola Morley, Linden Ball, Charles Langley, and Clive Spenser.2003. Using ethnography to design a Mass Detection Tool (MDT) for the early discovery of insurance fraud. In CHI'03 Extended Abstracts on Human Factors in Computing Systems. ACM, 650–651.

[12] Jesús M Pérez, Javier Muguerza, Olatz Arbelaitz, Ibai Gurrutxaga, and José I Martín. 2005. Consolidated tree classifier learning in a car insurance fraud detection domain with class imbalance. In International Conference on Pattern Recognition and Image Analysis. Springer, 381–389.

[13] Clifton Phua, Damminda Alahakoon, and Vincent Lee. 2004. Minority report in fraud detection: classification of skewed data. Acm sigkdd explorations newsletter 6, 1 (2004), 50–59.

[14] Stijn Viaene, Richard A Derrig, Bart Baesens, and Guido Dedene. 2002. A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection. Journal of Risk and Insurance 69, 3 (2002), 373–421.

[15] Stijn Viaene, Richard A Derrig, and Guido Dedene. 2004. A case study of applying boosting Naive Bayes to claim fraud diagnosis. IEEE Transactions on Knowledge and Data Engineering 16, 5 (2004), 612–620.

[16] Herbert I Weisberg and Richard A Derrig. 1998. Quantitative methods for detecting fraudulent automobile bodily injury claims. Risques 35, July–September (1998), 75–99.

[17] Graham J Williams and Zhexue Huang. 1997. Mining the knowledge mine. In Australian Joint Conference on Artificial Intelligence. Springer, 340–348.

[18] Kenji Yamanishi, Jun-Ichi Takeuchi, Graham Williams, and Peter Milne. 2004. On-

line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Mining and Knowledge Discovery 8, 3 (2004), 275–300.

[19] Wan-Shiou Yang and San-Yih Hwang. 2006. A process-mining framework for the detection of healthcare fraud and abuse. Expert Systems with Applications 31, 1 (2006), 56–68.

[20] Jun Zhou, Qing Cui, Xiaolong Li, Peilin Zhao, Shenquan Qu, and Jun Huang. 2017. PSMART: parameter server based multiple additive regression trees system. In Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 879–880.

[21] Indr˙e Žliobait ˙ e. 2010. Learning under concept drift: an overview. arXiv preprint arXiv:1010.4784 (2010).5

心得体会

下图是全国部分省市的退运险，每一笔的费用都在 10 元上下，从单笔交易来说费用不是很高，但是对于阿里金融这种体量的公司来说对于诈保行为的损失却是巨大的，因此他们开发的系统的作用是巨大的。

发货地	收件地	赔付金额	发货地	收件地	赔付金额	发货地	收件地	赔付金额
上海	上海	5	安徽	上海	7	河北	上海	10
上海	云南	11	安徽	云南	11	河北	云南	12
上海	内蒙古	15	安徽	内蒙古	15	河北	内蒙古	12
上海	北京	9	安徽	北京	12	河北	北京	8
上海	吉林	12	安徽	吉林	15	河北	吉林	12
上海	四川	10	安徽	四川	11	河北	四川	10
上海	天津	10	安徽	天津	12	河北	天津	8
上海	宁夏	15	安徽	宁夏	15	河北	宁夏	15
上海	安徽	8	安徽	安徽	8	河北	安徽	11
上海	山东	9	安徽	山东	12	河北	山东	10
上海	山西	11	安徽	山西	12	河北	山西	11
上海	广东	9	安徽	广东	12	河北	广东	10
上海	广西	12	安徽	广西	12	河北	广西	12
上海	新疆	18	安徽	新疆	15	河北	新疆	15
上海	江苏	6	安徽	江苏	8	河北	江苏	9
上海	江西	9	安徽	江西	12	河北	江西	11
上海	河北	9	安徽	河北	10	河北	河北	7
上海	河南	9	安徽	河南	11	河北	河南	9
上海	浙江	6	安徽	浙江	8	河北	浙江	10
上海	海南	12	安徽	海南	12	河北	海南	12
上海	湖北	9	安徽	湖北	10	河北	湖北	10
上海	湖南	10	安徽	湖南	12	河北	湖南	10
上海	甘肃	12	安徽	甘肃	12	河北	甘肃	12

图 1 各地退运险

在该论文中讲述了训练数据的标记以及使用的模型，以及使用的算法。

训练数据的标记：

对于每个账户，文中提取 50 个 feature，例如，超过一个月的投诉提交数量，作为客户的持续时间，从保险索赔历史，邮寄历史和购物历史中提取。还收集了设备使用历史记录，以便进行图形构建。将“高风险”账户视为欺诈性账户，而“无明显风险”账户则视为常规账户。

由于数据集过于庞大，因此使用人工为每一个数据打标签是不可能的，因此在该模型中使用的数据标签都是由基于规则的诈保检测模型给出的，但是该模型对于高风险的账户可以准确地判断，但是对没有检测到风险的账户无法确定该账户是否存在诈保行为，此处他们给出的解决方法为随机抽取样本数据，这样导致的存在的错误是随机的，降低了因错误标签产后严重后果的可能性。

使用的模型：

在模型部分建立了三个 graph，分别是 device-sharinggraph, transaction graph 和 friendship graph，分别代表着用户使用的设备之间的关系，用户的交易之间的关系以及用户的人际关系，并给出了评判三个表的好坏标准：

1 相距较近的节点应当具有相似的标签；

2 fraudulent accounts 与 regular accounts 所构成图的结构应当是有明显不同的。

其中 regular account 和 fraudulent account 在 device 和 transaction graph 上有着明显的区分（图一），因此证明了上述三幅图的正确性。

使用的算法：

该论文中运用的欺诈检测模型是基于 GeniePath 的。简单地将自适应路径层叠加在图中的宽度和深度探测上。对于该论文中的广度探索，它强调具有相似帐户功能的邻居的重要性。

文中的广度搜索邻居聚合模型：

$$\mathrm{AGG}\left(h_v^k\right) = \sum_{u \in N(v)\cup\{v\}} softmax(W^T \tanh(W_s h_v^k + W_d h_u^k)) \cdot ,_u \qquad \text{公式 1}$$

文中的目标函数：

$$\mathcal{L}(\mathrm{w}) = \min_{w}(\sum_{v \in V_{fraudulent}} \ell(f(X_v; w), fraudulent) \qquad \text{公式 2}$$

$$+\sum_{v\in samole(Vregular)}\ell(f(X_v;w),regular)$$

思考：

该论文中解决的一个很大的问题就是训练数据的问题，这其中包括数据模型的建立，文中给出的解决方法为使用了三个图；对于数据的特征，文中给出的是选取了五十个特征可以极大范围的抽取出数据具有的特征信息，这与上面的三个图的结构相得益彰，还有就是关于训练数据的标签的问题，由于文中的训练数据标签是采用基于规则的诈保监测方法，即使是采取了随机选取诈保数据的方法，已经在很大程度上降低了由于部分标签错误造成的损失，但是毕竟依然存在着先天的缺陷。

在结论部分中我们看到采用该论文提出的方法经过保险专家的评估具有很好的准确度，要远远的高于基于规则的退运险诈保检测系统的结果，此外为了解决概念漂移的问题（随着时间的推移而发生的新型欺诈行为，并使欺诈检测系统变得越来越不可预测）每隔一段时间（30 天）将会更新用户的信息，因此在后面的训练过程中我们的训练数据的标签可以不再是基于规则的退运险诈保监测方法获得，而是基于该系统以往的检测结果，由于训练数据的准确度的提升，最终的处理结果也应该会有所提高，长此以往将会形成一个良性循环。

在文中考虑的是三个图对账户性质的判断，但是对于 transaction graph 和 friendship graph 他们之间有某种联系，因此在后面的过程中是否应该考虑到交易事务与人际关系之间的联系，因为任意两个不认识的人都可以通过不超过六个人来联系起来，也就是说任意两个人之间的交易都可能存在 friendship 中的关系相对应，相信这对最终结果应该也会有一定的影响。

注：本文为自己翻译，知识水平有限，部分专有词汇可能翻译不准确，正在增强中。