

识别欺诈索赔 – 机器学习方法

Dr. Behrang Jalali, Gen Re, 科隆

保险公司通过出色的理赔管理可以展现其稳健性与高品质。这一保险环节不仅对于保险公司的盈利能力至关重要，而且也能根本性地影响客户的满意程度。有效且高效地识别正当索赔与欺诈索赔是确保保险公司财务实力的重要途径，从而为客户提供最优的保险补偿与财务支持。

举例来说，健康险欺诈给全球保险公司带来巨大的保险损失与额外工作，不仅延误对客户的索赔处理与赔款给付，而且未被识别的欺诈也导致保险成本上升，使诚实的客户承担保费上涨的不利后果。

各种新型和不同规模的欺诈行为随着科技进步而产生并衍变。为了加强欺诈识别与防范相关保险损失，我们应该与理赔专业人员合作，采用多种数据分析策略，密切监测索赔数据并完善业务规则，助力识别可疑索赔。

根据欧洲保险与职业养老金监管局（EIOPA）的报告，理赔管理（包括欺诈预防）仅次于定价与核保，是保险价值链上最大的领域之一，分析应用程序尤其是机器学习对于理赔管理可能大有益处。

传感器与移动电话使用日益普及，这意味着从不同来源收集与连接数据正在成为常态，而这种数字化发展也更便于被保险人频繁提起索赔。大数据技术与现代方法论可优化理赔处理系统，为理赔部门提供支持。

本文介绍分析学在健康险上的运用。我们为健康险理赔业务下的欺诈索赔建模。由于建模对象少见，在这一分类任务中，算法的选择、参数的整定以及挑选最优阈值至关重要。我们将说明为何这一任务应该是一个迭代过程，需要分析学专家与理赔管理人员之间的充分沟通。

数据与业务任务

在我们的数据中，有将近一百万条索赔记录以及20多个变量，其中有：

- 个人属性，如年龄、索赔年龄及性别
- 索赔特征，如索赔史（多次索赔）、微小或重大索赔、单个或多个产品索赔以及索赔额

目录

数据与业务任务	1
分析学方法	2
模型评估与比较	2
模型与阈值选择的经济影响	3
结语	4

本刊简介

Risk Insights是Gen Re面向全球寿险及健康险管理人员推出的一份专业期刊。期刊文章聚焦于精算、核保、理赔、医学以及风险管理方面的问题。重点讨论的产品包括寿险、健康险、失能收入险、长期护理险和重大疾病险。

- 保单信息，包括直接登记或代理登记、保险计划类型（选项从基础保障到最复杂的保障）
- 医院相关信息，如入院原因（高阶编码）、住院时间及在院状态（例如院前正常/可疑经历）

理赔专业人员对索赔进行评估，并将索赔标记为正常或警示标记为欺诈可能。警示索赔的理由包括：可疑保单信息或恶意代理、恶意索赔或医院相关欺诈行为。在该项目中，警示索赔非常少见，仅占索赔总数的0.3%。这使得模型构建极具挑战性，因为反应分布非常不平衡，就如大海捞针一般。

我们的业务目标是识别尽可能多的警示索赔，同时又不会错误地将大量正常索赔预测为警示索赔，因为这将不必要地增加调查费用。

记住，在这一极端情形下，只是因为正常索赔的数量庞大，任何弱模型都会预测大多数索赔为正常索赔；但也可能轻易漏掉大多数（如果不是全部的话）警示索赔，因为它们的数量太少。

分析学方法

我们计划开发出一个机器学习模型，即所谓的二元分类器，它能够尽可能正确地识别两个标记。由于数据已被标记，这是一个监督式学习方法。这种解决方案可以作为推荐系统被整合到业务流程中去，它能自动快速处理许多索赔，为理赔管理提供一种机器学习驱动的补充方法。

需要注意的是，对于每次观测（在该项目上是每件索赔），二元分类器的直接输出是介于0与1之间的一个概率。取决于设定的切点阈值，预测概率或模型概率高于（或低于）此阈值的索赔将被标记为警示（或正常）索赔。若您欲了解关于二元分类器的更多信息，请参考Risk Insights期刊上由Louis Rossouw撰写的文章“Classification Model Performance（分类模型表现）”。

在整个分析中，我们给正常索赔赋值为0（阴性），给警示索赔赋值为1（阳性）。选择太高的阈值意味着大多数索赔将被预测为正常索赔（因为它们的预测概率将小于这个高阈值）。这就意味着我们会漏掉一些或大多数应被警示标记的索赔。在另一个极端，选择零阈值意味着大多数索赔将被预测为警示索赔（因为很可能大多数预测概率将大于零），而我们知道事实并非如此。因此，我们打算找到一个“好的”低阈值。在下一章节，我们将进一步选择一个最优阈值，并探讨最佳模型的经济效益。

在构建机器学习模型时，通常的做法是选择一个统计指标或测度来优化模型，但是在极端情形下，就如在该项目上，选择统计指标来帮助克服少见的欺诈（或“警示”）索赔识别困难，这并不是一个简单直观的过程。从分析学角度而言，对于此类极不平衡的数据，通过优化典型指标来发现好的模型并不理想，如“准确性”（一个分类指标，指阳性与阴性正确预测的数量除以正确与错误预测之和）或“AUC”（另一个分类指标，指ROC曲线以下的区域）。

这是因为“准确性”和“AUC”对于同时完美识别两种标记不敏感。因此，我们寻求其他指标，如已识别的警示索赔数量（用专业术语来说与“查全率”相关），或者我们监测二元模型所有四个可能结果的情况，它们是称之为混淆矩阵的分量，可适用于特定的业务案例。这就是我们在当前项目中要探讨的内容。

模型评估与比较

我们采用三种不同的算法：广义线性模型（GLM）、梯度推进机（GBM，一种决策树系综）以及神经网络（NN，一种多层体系结构）。我们在每一种算法内比较多个模型的表现，以期找到最优模型。

我们使用H2O（一个先进的机器学习平台）构建了多种模型网络，并使用了定制开源工具。为了确保分析与解决方案可扩展性的高性能，我们在本地的超级计算机上来执行分析。

在图1的左上方，可以看到GLM模型（红色短虚线）没有正确识别任何警示索赔，在所有阈值上基本都是零真阳性率（TPR）。左下方的蓝色短虚线也显示大量的假阴性率（FNR），即索赔被错误识别为正常，即使设定的阈值非常小。因此，GLM模型太过简单，无法解决这个问题，很可能由于其线性与简单附加属性，GLM无法学到必要的模式为这种富有挑战的不平衡数据正确建模。

通常GBM模型（决策树系综）足够复杂，通过增大其复杂性（如增加树的数量，并让每颗树延展得更深），它们在学习精细模式方面表现上佳。在这个项目上，也与我们预期相符，GBM在正确识别警示索赔上有所改善（真阳性识别），如图1红色点虚线所示。

为了更好地比较几个模型，表1列出阈值为0.1时四个模型的表现。我们看到，GBM模型的表现仍然差强人意，仅仅识别出约26%的警示索赔。即使对于决策树系综，这个挑战看起来也十分巨大。

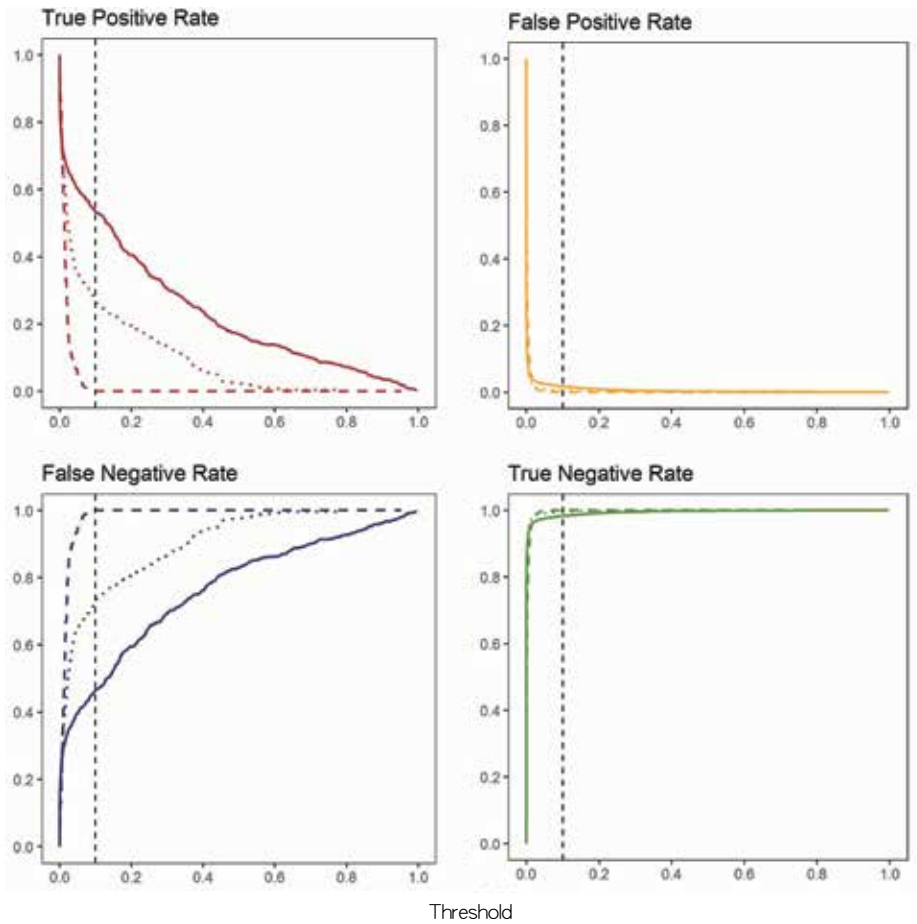
我们增大算法的复杂性，采用了神经网络。简单地说，神经网络（NN）是一种多层体系结构，每层有许多神经元，可以将输入以非线性方式转换。当信息从最初输入层向前流动时，错误被用来调整权重，预测朝着最终输出层被修正。

这种结构的优势在于，无需预先假设就能自动识别精细模式，包括众多变量之间的非线性交互。因此，在复杂数据应用方面，NN优于其他算法。实现NN的一个关键条件是可使用强大的计算硬件，以便执行巨量的计算。

首先我们使用默认参数设置的NN模型。如表1所示，出人意料的是，相较于GBM，我们看到结果并无显著改善。接下来，我们发现只是通过调试详细的参数，以及采用在不平衡数据情形下众所周知的一些技术，如在训练步骤对小类过采样，结果就能改善至可接受的程度。图1也显示出这一点，即左上方的红色实线。

模型的真阳性率较高，同时真阴性率（TNR）也不低，这两点很重要。如表1以及后面的表2所示，TNR即使只轻微降低，也能导致额外增加大量的假阳性结果，这将极大地增加调查费用。

图1 – 三种模型的混淆矩阵分量。GLM以短虚线表示，GBM以点虚线表示，优化NN以实线表示。



来源：Gen Re

表1 – 阈值为0.1时各模型的表现

模型	真阳性率	真阴性率
GLM	0	1
GBM	0.26	0.996
默认NN	0.28	0.993
优化NN	0.53	0.983

表2 – 选定优化神经网络（阈值为0.1）经济论证的两项

名称	实际	预测	计数	总数 (单位)	
真阳性	1	1	290	165,320	索赔总额
假阳性	0	1	2,980	59,600	调查费用

来源：Gen Re

模型与阈值选择的经济影响

到目前为止，我们只是以主观假设低阈值（0.1）为例，比较了这些模型的表现。我们应该怎样去挑选一个最佳阈值呢？

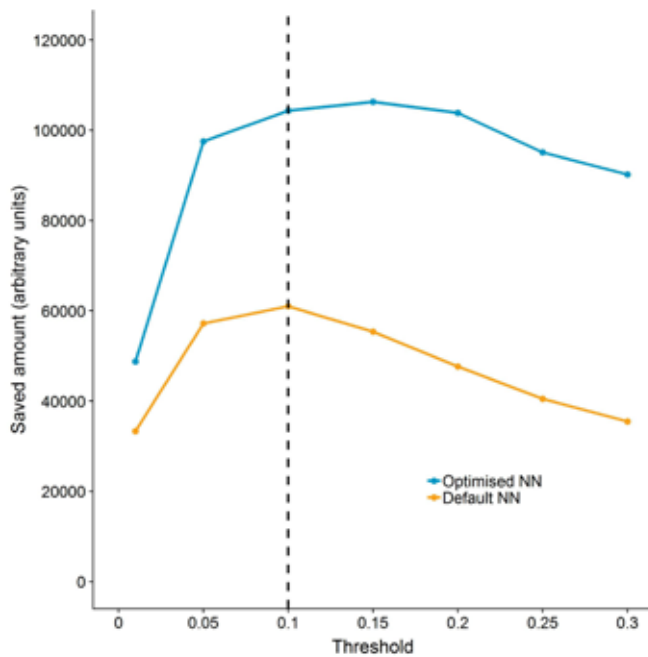
为此，分析团队需要与业务专家合作来确定满意的结果。通过充分交流，我们形成了以下的经济论证：我们应该评估机器学习模型生成的数值来预测警示索赔，并只对这些索赔进行更深入的调查。

从好的一面来看，我们有节约（未给付）索赔总额，这些索赔被正确识别为警示索赔（真阳性）。我们需要从索赔总额中减去错误识别为警示索赔（假阳性或FPs）的调查费用，即假阳性索赔数量乘以调查费用。为了说明这一点，以阈值为0.1的优化NN模型为例，表2列示出这一论证的每一项，假设FPs的一小时调查费用为20个单位。

为找到一个最佳阈值，图2显示这一论证随小范围阈值变化的表现。根据蓝线的表现，假设FPs的一小时调查费用为20个单位，节约索赔总额在阈值为0.15左右时达到最高值（0.1和0.2为次优阈值）。

我们看到，分析与财务需求之间的相互作用实际上可以被用于约束阈值。举例来说，阈值小于0.1在财务上不理想，因为节约索赔总额下降（由于FPs数量增加，因此以上论证中的第二项增加）。另一方面，阈值大于0.1在分析学上不理想，因为真阳性索赔数量减少（因此会漏掉一些警示索赔并错误地做出理赔给付），但是如果差别导致节约索赔总额足够大，则从业务角度而言也是合理的。

图2 – 两个NN模型的净节约索赔总额以及真阳性与假阳性识别率之间的权衡



Source: Gen Re

最后，图2以默认参数设置的NN为例，评估了以上经济论证。如橙色线所示，与优化NN模型相比，默认NN模型的节约索赔总额低得多，几乎只有前者的一半。这与表1的信息是一致的，在表1中，NN模型的TPR比优化NN模型的TPR低得多。

这表明构建模型网格来发现更强模型的重要性，因为它们在经济上获益更大。将开发出的更好模型整合到业务流程中可以提高其效率，更准确地将索赔分类至相关级别。

结语

本文阐述了二元分类器模型在理赔应用上的实例。这一任务可以被轻松扩展，以改善保险价值链上的其他领域，例如核保以及多种分类用例，如将核保决定分类为标准、次标准以及拒保，或者客户细分应用的多种分类，在这些场景下，定价或激励机制应根据不同客户群组做出调整，调整方式是可持续的、数据驱动的并且是非歧视的。

总之，我们谨慎评估了表现最佳的模型，审视了它们的经济影响以优化业务驱动价值。我们看到，选择最优阈值来分类正常索赔与警示索赔需要业务团队与分析团队之间的深入沟通。

构建与评估许多模型是分析项目中的重要环节，尤其是需对众多参数进行实验的复杂模型，如此才能开发出可以进入生产阶段的模型。

Gen Re已开发出有多种可解释性方法的机器学习模型，我们可以对模型表现进行验证，并更好地理解模型的输出，以及将分析步骤的发现与特定业务需求最好地连接起来。

此外，我们在本地以及云端环境都有适宜的基础设施，可以无缝开发出模型网格，并大规模地分析海量数据项目。我们希望与客户共同开展此类项目，并根据客户需求量身定制项目。

来源

European Insurance and Occupational Pensions Authority (2019), Big Data Analytics in Motor and Health Insurance https://www.eiopa.europa.eu/sites/default/files/publications/eiopa_bigdataanalytics_thematicreview_april2019.pdf

Rossouw, L. (2018). Classification Model Performance. Risk Insights Gen Re <http://www.genre.com/knowledge/publications/ri18-1-en.html>

H2O.ai, an open source machine learning platform <https://www.h2o.ai/>

作者简介

Dr. Behrang Jalali 负责 Gen Re 国际寿险/健康险业务的高级分析能力建立与开发。他是数据分析团队负责人，与业务部门合作管理分析项目，包括实验设计与现代技术的实施，为客户创造附加价值。在加入 Gen Re 之前，他是一名计算天体物理学研究员。

Tel. +49 221 9738 799

behrang.jalali@genre.com



The difference is...the quality of the promise.



genre.com | genre.com/perspective | Twitter: @Gen_Re

General Reinsurance AG
Theodor-Heuss-Ring 11
50668 Cologne, Germany
Tel. +49 221 9738 0
Fax +49 221 9738 494

Editors:
Ulrich Pasdika, ulrich.pasdika@genre.com
Ross Campbell, ross_campbell@genre.com

Photos: © getty images – TABoomer, Gen Re, Artem Peretiatko

© General Reinsurance AG 2020

This information was compiled by Gen Re and is intended to provide background information to our clients, as well as to our professional staff. The information is time sensitive and may need to be revised and updated periodically. It is not intended to be legal or medical advice. You should consult with your own appropriate professional advisors before relying on it.