

INTERNET FINANCE

互联网金融



栏目编辑 贾红宇
jiah@pbcfs.tsinghua.edu.cn

● P2P行业的全球发展



大数据征信应用与启示 ——以美国互联网金融公司 ZestFinance为例

文/刘新海 丁伟 本文编辑/贾红宇

近期，美国互联网金融公司ZestFinance受到国内互联网金融专业人士的热捧，其基于大数据的信用评估模型也越来越受到关注和效仿。本文结合美国的金融环境，对ZestFinance进行简要介绍，分析大数据征信产生的背景，剖析大数据征信技术，并全面客观地阐述了大数据征信技术对于中国互联网金融和征信业未来发展的借鉴意义。



INTRODUCTION

刘新海

中国人民银行征信中心博士后



ZestFinance简介

ZestFinance, 原名 ZestCash, 是美国一家新兴的互联网金融公司, 2009年9月成立于洛杉矶, 由互联网巨头谷歌(Google)的前信息总监道格拉斯·梅瑞尔(Douglas Merrill)和金融机构Capital One的信贷部高级主管肖恩·卜德(Shawn Budde)(曾管理过收益超过10亿美元的次级信贷业务)联合创办。ZestFinance的研发团队主要由数学家和计算机科学家组成, 前期的业务主要通过ZestCash平台提供放贷服务, 后来专注于提供信用评估服务, 旨在利用大数据技术重塑审贷过程, 为难以获得传统金融服务(Underbanked)的个人创造可用的信用, 降低他们的借贷成本。

ZestFinance起初是为传统的发薪日贷款(Payday Loans)提供在线替代的产品。发薪日贷款因借款人承诺在发薪日还款而得名。由于美国传统的信用风险评估体系无法覆盖全部的人群, 大约15%的人因没有信用评分而被银行排斥在外, 无法获得基本的信贷需求。除了解决传统信用评估体系无法解决的无信用评分借贷问题, ZestFinance还主要面向传统信用评估解决不好的领域, 将信用分数低而借贷成本高的人群视为服务对象, 利用大数据技术降低他们的信贷成本。与传统信贷管理业务比较, ZestFinance的处理效率提高了将近90%, 风险控制方面, ZestFinance的模型相比于传统信用评估模型性能提高了40%。

ZestFinance目前也正在向信用风险管理的其他领域纵深扩展, 2014年2月ZestFinance宣布推出基于大数据分析的收债评分(Collection Score), 旨在为汽车金融、学生贷款、医疗贷款提供一种新的评分系统。ZestFinance的未来发展方向是希望把其在这种发款日的贷款上的优势继续拓展到其他贷款领域, 包括信用卡、汽车的贷款, 甚至包括房屋的贷款, 在未来的10~15年, 这一方法将取代现行指标, 成为申请信贷的唯一评估标准。

ZestFinance引起国内的关注始于2013年7月, 当时全球第三方支付平台PayPal联合创始人、美国知名投资人彼得·泰尔(Peter Thiel)领投了ZestFinance的2000万美元融资。

为什么要进行大数据评估传统的信用评估服务无法覆盖全体人群, 特别是弱势群体

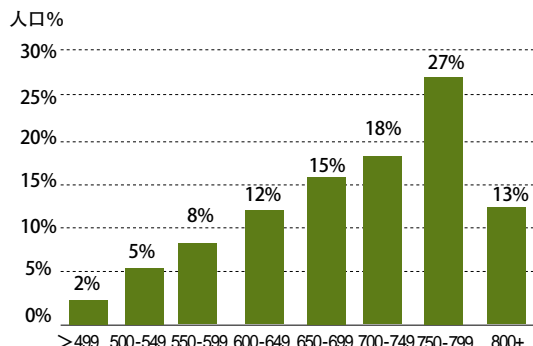
图1展示了美国FICO评分与其对应的人口分布情况, 初始每个

表1: FICO评分人群和相关的金融服务机构

FICO 评分人群	金融服务机构
>650分	美国大银行 (花旗、大通、富国、美国银行等)
600分~720分	互联网P2P网贷Lending club等
500分~700分	小贷公司Spring leaf等
<500分	新兴的互联网金融公司ZestFinance、Turbofinance、Thinkfinance等

人的分值基数为850分, 信用评分模型利用征信数据从多个评分因素考察消费者的信用风险, 从850分中减分。大致来看, 美国个人消费者信用评分人群分布状况呈现两头小中间大的形态, 信用分数处于750~850的人群有40%之多, 其中信用分数在800~850大约占总人数的13%, 在750~799超过总人数的25%, 这是整个信用社会的中间阶层, 对应于美国的中产阶级。其中, 美国个人消费者的平均FICO评分为678。从图1可以看出, 还有大量的人群远低于平均的678分, 如FICO评分在550~549的占8%, 在500~549的占5%, 小于499分的占2%。根据FICO的标准, 如果人们未能如期还款, 或者缺乏借贷经历, 他们就会自动被视为风险人士, 他们的贷款也就会被惩罚性地给以更高的利率。还有一种可能, 那就是他们的贷款申请会被拒, 无论是否事出有因。比如, 遇到了医疗紧急事故, 或者最近刚刚移民美国。表1根据FICO评分将服务人群

图1: 美国FICO评分及相应的人群分布





INTRODUCTION

丁伟

中国联通网络技术研究院高级工程师

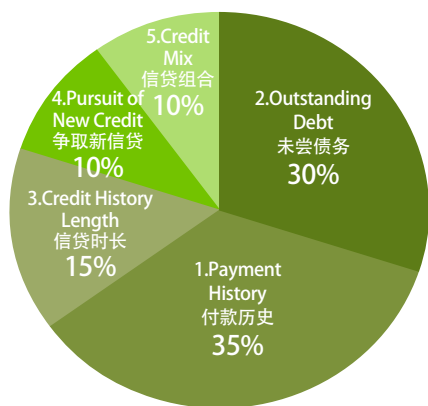


分为四个区间，并对应不同的金融服务机构。信用记录不完整或者不够完善的个人消费者，依据传统信用评估体系（FICO评分），往往很难被传统金融服务机构所覆盖，即使在金融体系发达的美国也无法获得常规的金融服务，或者需要付出很大的代价才能获得常规的金融服务。

传统信用评估模型信息维度比较单一

传统的FICO评分模型的基本思想是比较借款人信用历史资料与数据库中的全体借款人的信用习惯，检查借款人的发展趋势跟经常违约、随意透支，甚至申请破产等各种陷入财务困境的借款人的发展趋势是否相似。如图2所示，它主要从五个方面考察用户的信贷资质。但随着信贷业务的进一步开展，FICO信用评分由于单一的标准、严苛的门槛和片面的评估结果而饱受诟病。

图2：传统信用评分模型的参考要素



传统的信用评估模型虽然在进行信用风险管理过程中发挥了很大的作用，如曾经促进了美国房贷市场的飞速发展。但在大数据背景下个人消费者出现许多信息维度，如电子商务、社交网络和搜索行为等，传统信用评估模型解决问题的能力越来越受限。

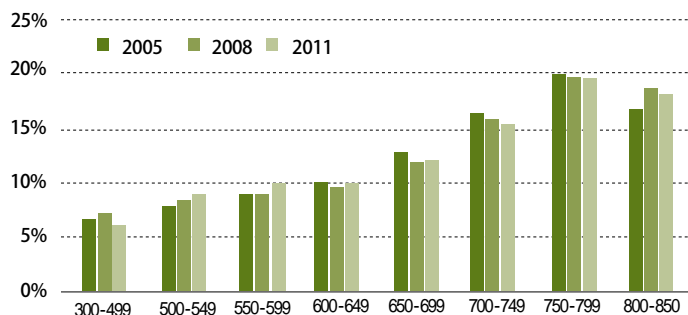
传统信用评估模型时间上比较滞后

虽然FICO评分仍然体现风险排序，但其预测绝对风险的能力和在2008年金融危机中的表现饱受指责，图3展示，FICO分数从2005年到2011年在美国人口中的分布基本

上没有大的变化，这和2008年金融危机爆发之后出现大量坏账的现实严重不符。

由于传统的基于FICO评分的信用评估模型覆盖人群窄、信息维度单一、时间上滞后，所以，在大数据时代，需要探索信用评估的新思路。国外三大征信机构和FICO公司都已经开始了如何利用大数据技术来完善传统信用评估体系的前瞻性研究，如益百利（Experian）投入研究团队关注社交网络数据对信用评分的影响，FICO公司多年前就开始了在线评估的信息工具和基于互联网的信用评估系统的项目研究。

图3：FICO评分从2005年到2011年在美国人口中的分布情况



	2005	2006	2007	2008	2009	2010	2011
300-499	6.6	6.5	7.1	7.2	7.3	6.9	6.3
500-549	8	8	8	8.2	8.7	9	8.7
550-599	9	8.8	8.7	8.7	9.1	9.6	9.9
600-649	10.2	10.2	9.7	9.6	9.5	9.5	9.8
650-699	12.8	12.5	12.1	12	11.9	11.9	12.1
700-749	16.4	16.3	16.2	16	15.9	15.7	15.5
750-799	20.1	19.8	19.8	19.6	19.4	19.5	19.6
800-850	16.9	17.9	18.4	18.7	18.2	17.9	18.1
合计	100	100	100	100	100	100	100

来源：FICO 银行分析博客@ 2011 Fair Isaac 公司

ZestFinance大数据信用评估的实践

ZestFinance的基本理念是认为一切数据都是和信用有关，在能够获取的数据中尽可能地挖掘信用信息。ZestFinance对大数据技术的应用主要从大数据采集和大数据分析两个层面为缺乏信用记录的人挖掘出信用。

大数据采集技术

ZestFinance以大数据技术为基础采集多源数据，一方面继承了传统征信体系的决策变量，重视深度挖掘授信对象的信贷历史。另一方面，将能够影响用户信贷水平的其他因素也考虑在内，如社交网络信息、用户申请信息等，从而实现了深度和广度的高度融合。

ZestFinance的数据来源十分丰富，依赖于结构化数据的同时也

导入了大量的非结构化数据。另外，它还包括大量的非传统数据，如借款人的房租缴纳记录、典当行记录、网络数据信息等，甚至将借款人填写表格时使用大小写的习惯、在线提交申请之前是否阅读文字说明等极边缘的信息作为信用评价的考量因素。类似地，非常规数据是客观世界的传感器，反映了借款人真实的状态，是客户真实的社会网络的映射。只有充分考察借款人借款行为背后的线索及线索间的关联性，才能提供深度、有效的数据分析服务，降低贷款违约率。

如图4所示，ZestFinance的数据来源的多元化体现在：首先，对于ZestFinance进行信用评估最重要的数据还是通过购买或者交换来自于第三方的数据，既包含银行和信用卡数据，也包括法律记录、搬

家次数等非传统数据。

再次是网络数据，如IP地址、浏览器版本甚至电脑的屏幕分辨率，这些数据可以挖掘出用户的位置信息、性格和行为特征，有利于评估信贷风险。此外社交网络数据也是大数据征信的重要数据源。

最后，直接询问用户。为了证明自己的还款能力，用户会有详细、准确回答的激励，另外用户还会提交相关的公共记录的凭证，如水电气账单、手机账单等。

多维度的征信大数据可以使得ZestFinance能够不完全依赖于传统的征信体系，对个人消费者从不同的角度进行描述和进一步深入地量化信用评估。

大数据分析模型

图5展示了ZestFinance的信用评估分析原理，融合多源信息，采用了先进机器学习的预测模型和集成学习的策略，进行大数据挖掘。首先，数千种来源于第三方（如电话账单和租赁历史等）和借贷者的原始数据将被输入系统。其次，寻找数据间的关联性并对数据进行转换。再次，在关联性的基础上将变量重新整合成较大的测量指标，每一种变量反映借款人的某一方面特点，如诈骗概率、长期和短期内的信用风险和偿还能力等。然后将这些较大的变量输入到不同的数据分析模型中去。最后，将每一个模型输出的结论按照模型投票的原则，形成最终的信用分数。

其中，ZestFinance开发了10个基于机器学习的分析模型，对每位信贷申请人的超过1万条数据信

图4：ZestFinance的大数据来源

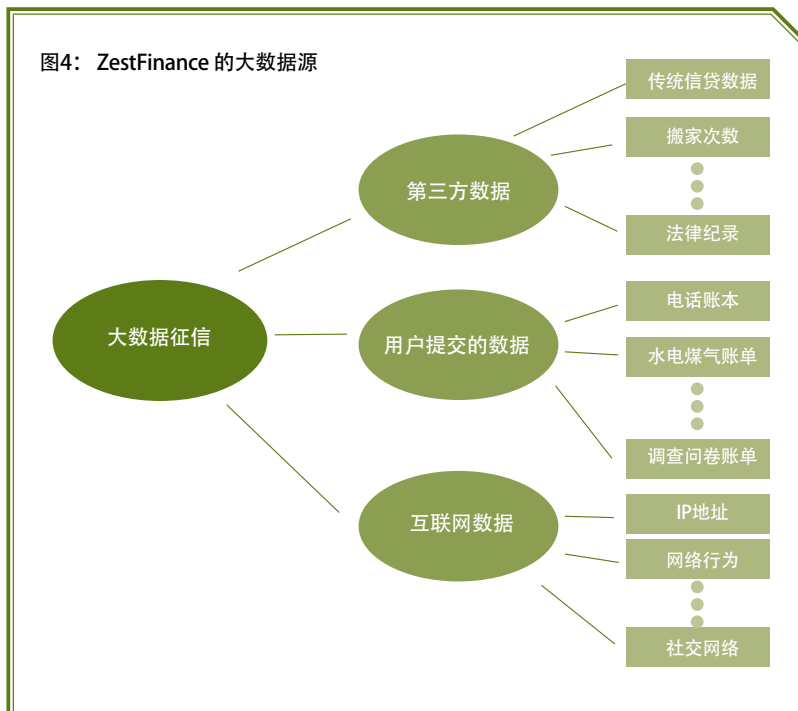


图5: ZestFinance的信用评估分析原理 (来自于ZestFinance官方网站)



息进行分析, 并得出超过7万个可对其行为做出测量的指标, 在5秒钟内就能全部完成。这10个模型以如下的方式进行投票: 让你最聪明的10个朋友坐在一张桌子旁, 然后询问他们对某一件事情的意见。这种机制的决策性能远远好于业界的平均水平。

近年来, 这种基于大数据的信用风险评估框架(远不能称为主流的信用评估方法)被国内外多家互联网金融机构采用, 如德国的Kreditech、美国的Kabbage, 以及国内最近获得IDG公司A首轮4000万元投资的闪银(Wecash)等, 对传统的信用体系形成了冲击。

如表2所示, 将这种将基于大数据技术的信用评估体系和传统信用评估(以美国的征信体系为例)相比, 发现主要的区别有以下几个方面。

(1) 从服务的人群来说, 新的信用评估体系可以服务没有被传统征信体系覆盖的人群, 即没有征信记录的人群(美国的征信体系能够覆盖85%的人群, 覆盖不到15%

的人群)。

(2) 从数据源来说, 这种新的信用风险评估体系大量采用非传统的信用数据, 包括互联网上的行为数据和关系数据, 传统的信用数据(银行信贷数据)的比重仅占到了40%, 甚至完全不用传统的信贷信用数据进行风险评估。

(3) 从关注的侧重点来看, 传统的信用评估模型更关注授信对象的历史信息, 致力于深度挖掘。而新的信用评估体系更看重用户现

在的信息, 致力于横向拓展。

(4) 信用量化评估的方式也发生了改变, 新的信用评估体系抛弃了只用很少变量的FICO信用评分模型, 基于大数据技术, 不仅采用机器学习的模型, 而且使用更多变量, 一方面可以使信用评估的决策效率提高, 另一方面还明显降低了风险违约率。

对中国互联网金融和信用评估的启示

利用大数据技术的信用评估方法在现实中有着很大的市场需求, 如国内快速发展的互联网金融中的风险管理问题。目前互联网金融处于快速的发展过程中, 根据银监会的统计, 目前国内可查的P2P网贷公司已经达到1200家。信用风险评估是P2P网贷的核心问题, 存在很多挑战, 如很多信贷客户没有或者是缺乏银行的信贷记录。在应对风险控制的挑战时, ZestFinance 受到了互联网金融机构的热捧, 目

表2: 传统的信用风险评估体系和基于大数据的信用评估体系的比较 (以美国为例)

	传统信用风险评估体系	基于大数据的信用风险评估
代表企业	FICO	ZestFinance
服务人群	有丰富信贷记录的(约占85%)	缺乏或无信贷记录的(约占15%)
数据格式	结构化数据	结构化数据+大量非结构化数据
数据类型	信贷数据	信贷数据、网络数据、社交数据
理论基础	逻辑回归	机器学习
变量特征	还款记录, 金额, 贷款类别	传统数据、IP地址、邮箱姓名、填写习惯等网络行为
数据来源	银行提交给第三方的数据和银行当地数据	第三方(如电话账单和租赁历史等)和借贷者本身提供的数据
变量个数	15~30(变量库400~1000)	多达几千到一万个

前国内多家互联网金融机构正在和ZestFinance洽谈合作,认为这种利用大数据技术的信用评估方法是解决国内互联网金融和普惠金融的信用风险管理问题的灵丹妙药。然而对于ZestFinance的大数据征信技术,还需要有全面的认识。

(1) ZestFinance的大数据征信是完善和更新传统征信系统的积极尝试,而不是替代品。美国的金融体系比较健全,而且信用体系也比较健全,这是ZestFinance赖以生存的土壤。ZestFinance的服务人群定位比较清晰,并且有完善的征信体系做支撑,ZestFinance并不是完全摆脱传统征信体系,在ZestFinance进行信用评估时,传统征信数据要占到至少30%。中国的金融生态环境和美国还是有一定的差别,ZestFinance的经验不能直接照抄照搬,需要进行消化吸收,结合中国的实际情况来进行大数据征信。另外,生活在互联网时代,面对大数据技术的发展,美国三大征信机构以及FICO也已经开始大数据征信方面的研发(作为常规的数据源更新的一种方式),但截至目前,尚未形成独立的信用评估手段。

(2) ZestFinance的体量不大,目前仅为10万美国人提供了服务,在美国的影响力有限,真实的效果目前还很难总体评价。截至2013年7月,ZestFinance的C轮融资达到2亿美元。

(3) ZestFinance的大数据分析模型也给信用风险管理带来复杂性的挑战。传统的基于FICO的信用

评估方法,处理的变量比较少,对每一个变量进行细致地处理,并且可以给出合适的解释,模型的透明性可以方便地在银行的不同部门之间进行沟通,而且便于个人消费者对分数的理解。ZestFinance的基于大数据的数以千计的变量规模和多模型使得数据的处理和模型的解释变得比较复杂,在实际应用中会带来许多麻烦。

(4) 大数据的应用要注意个人的隐私保护。ZestFinance在利用个人消费者的大数据进行信用评估时,很多数据会涉及个人隐私,如对于个人社交网络的数据(微信朋友圈)和电商交易的数据、通话记录、微博的数据等应用,美国对个人隐私的保护是有明确的边界的。而国内关于个人隐私方面的保护目前处于空白,已经出现国内一些互联网金融公司为了进行信用评估,忽视个人消费者的知情权和隐私保护。因此在利用大数据进行信用评估的时候,要考虑使用个人隐私的合规性前提。

虽然以ZestFinance为代表的新兴信用评估体系还不够成熟,但是为征信业的变革注入了活力,特别是对于中国的征信体系的建设会有一定的启示作用。

(1) 普惠金融需要挖掘更多人的信用。国内目前真正发挥作用的征信体系主要是央行的征信系统,所覆盖的人群还是非常有限,远远低于美国征信体系对人口的85%的覆盖。目前我国个人有征信记录的仅有约3.2亿人,约占13.5亿人口中的23.7%。国内数量庞大没

有被传统征信体系覆盖的人群同样也需要信用服务,享受金融普惠,这就需要探索征信的新思路。

(2) 互联网上的海量信息可以成为征信体系的新数据源。ZestFinance的大数据实践的重要方面就是大量地利用互联网上的数据作为征信的数据源。中国目前是世界上互联网人口最大的国家,截至2013年12月,中国网民规模达6.18亿人,互联网普及率为45.8%,其中手机网民规模达5亿人,继续稳定增长。2013年中国网络购物用户规模达3.02亿人,网络使用率达到48.9%。截至2013年12月,我国使用网上支付的用户规模达到2.6亿人。这些海量而且丰富的互联网数据资源可以被国内征信体系建设很好地利用,通过分析互联网上这些信用主体的基本信息、交易行为信息和金融或经济关系信息,同样可以挖掘出这些信用主体的信用模式。

(3) 大数据技术可以使得“一切数据皆信用”成为可能。以大数据为代表的IT新技术的应用,给征信体系建设带来了新的思路,原来海量庞杂、看似无用的数据,经过清洗、匹配、整合和挖掘,可以转换成信用数据,而且信用评估的效率和准确性也得到了一定程度的提升。新的信用风险体系的一个颠覆性的基本思想是一切数据皆信用,这是需要大数据技术来支撑的。国内征信体系的建设应当关注大数据技术的应用和发展,并加大投入,勇于实践。❶(本文仅代表作者个人观点,与所在单位无关。)

45.8%

截至2013年12月,中国网民规模达6.18亿人,互联网普及率为45.8%,其中手机网民规模达5亿人,继续稳定增长。