

承诺书

我们完全清楚，在竞赛开始后参赛队员不能以任何方式，包括电话、电子邮件、“贴吧”、QQ 群、微信群等，与队外的任何人（包括指导教师）交流、讨论与赛题有关的问题；无论主动参与讨论还是被动接收讨论信息都是严重违反竞赛纪律的行为。

我们以中国大学生名誉和诚信郑重承诺，严格遵守竞赛章程和参赛规则，以保证竞赛的公正、公平性。如有违反竞赛章程和参赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

(指导教师签名意味着对参赛队的行为和论文的真实性负责)

日期: 2020 年 09 月 13 日

(请勿改动此页内容和格式。此承诺书打印签名后作为纸质论文的封面，注意电子版论文中不得出现此页。以上内容请仔细核对，如填写错误，论文可能被取消评奖资格。)

赛区评阅编号：_____
(由赛区填写)

全国评阅编号：_____
(全国组委会填写)

2020 高教社杯全国大学生数学建模竞赛

编 号 专 用 页

赛区评阅记录（可供赛区评阅时使用）：

评 阅 人						
备 注						

送全国评阅统一编号：
(赛区组委会填写)

(请勿改动此页内容和格式。此编号专用页仅供赛区和全国评阅使用，参赛队打印后装订到纸质论文的第二页上。注意电子版论文中不得出现此页。)

全国大学生数学建模竞赛编写的 L^AT_EX 模板

摘要

本文根据中小微企业票据信息建立全连接神经网络模型预测企业违约率，完成信贷风险的量化分析，然后建立最优化模型分别求出每个企业的贷款额度和年利率，最后通过层次分析法模型完成企业实力的量化分析，对突发因素下的信贷策略进行修正。

针对问题一，要求对信贷风险进行量化分析并给出银行的信贷策略。我们首先对附件一的发票数据进行了预处理，提取出 1200000 个 8×5 的特征矩阵并附上是否违约的标签，作为全连接网络的输入数据实现二分类模型，附件一中的 27 个违约企业我们成功预测了其中 13 个。对于信贷策略的制定实际上是确定每一个企业的信贷额度和年利率，首先根据信贷风险按比例划分银行的信贷总额求出每一个企业的信贷额度，然后确定年利率的优化目标是使银行年收益更大和客户流失更少，据此我们根据企业的信用等级求得该等级下的基准年利率，然后按照企业违约率进行修正从而得到最终的年利率。

针对问题二，我们运用问题一中的数据预处理方式和全连接网络模型得到每一个企业的预测违约率，然后根据附件一中的信用等级与违约率的关系建立模型，从而确定附件二 302 家企业的信用等级，最后按照与问题一相同的方法制定信贷策略。

针对问题三，要求我们对突发因素下的信贷策略进行调整，实际上只需对之前模型的违约率进行人工微调即可。首先根据突发因素的影响行业确定违约率的修正方向，然后根据所处行业，票据信息的统计量，企业信贷风险建立基于层次分析法的企业实力量化模型，接着根据企业实力和突发因素的剧烈程度确定违约率的修正范围，最后按照与问题二相同的方法制定信贷策略。

本文针对数据预处理时的数据分布不均衡和数值超过程序处理范围的问题，创新性地提出了重采样和归一化的方法，并运用全连接神经网络去挖掘票据信息中的隐藏特征，避免了人工分析的繁琐和精度低问题。本文针对违约率的二分类预测模型通过推广更有现实意义。

关键字： 全连接网络 最优化 层次分析法 数据预处理

一、背景介绍

众所周知,信息不透明、信息不对称是中小企业融资难的基本原因。与大企业相比,中小企业由于难以提供经审核合格的财务资料 and 经营记录,在向银行贷款时面临着更高的信贷风险。为了防止信息不对称造成的逆向选择和道德风险,银行往往会实施信贷配给政策,有选择地对有信贷需求的企业进行授信,拒绝不能满足信息要求的借款人的申请。国外大量的经验证据表明,信贷配给的主要对象正是信息不透明的中小企业。严重的信息不对称,使中小企业产生逆向选择和道德风险的问题,商业银行在贷款时处于更加不利的地位,而且在此过程中,商业银行为了尽可能的减少信息不对称,会加大监督执行成本,而且在收集和处理信息方面较之中小企业而言是规模而不经济的,直接增加商业银行对中小企业的贷款成本,商业银行不愿对缺乏相应信息的企业提供信贷支持而产生了“惜贷”的现象。

为了有效缓解金融交易中的信息问题,银行根据不同类型的信息制定了多种贷款技术。美国经济学家 Berger 等人将其分为四大类:基于财务报表的贷款,抵押担保贷款,信用评级技术,抵押担保贷款。由于中小企业自身的特有属性以及其存在的发展潜力,信用评级技术在中小企业信贷策略的制定中应用极为广泛。这是利用现代数理统计模型和信息技术,对客户的信用记录进行测算和分析,从而做出决策的新技术。它具有成本低、效率高的特点,可以在很大程度上缓解信息问题,从 20 世纪 80 年代开始,美国金融机构就开始使用这种技术发放小额贷款。但由于该技术的复杂性,对信息系统和数据积累的要求较高,其应用受到限制。

考虑到本题可供分析的数据比较单一且常用的信用评级技术较为复杂,为银行提供一个简单有效的信贷策略制定模型显得尤为重要,因此我们借鉴了信用评级技术的核心思想,充分挖掘票据信息等已有数据中的关系,致力于让银行以较低的贷款成本完成对中小微企业信贷策略的制定。

二、问题重述

在实际中,由于中小微企业规模相对较小,也缺少抵押资产,因此银行通常是依据信贷政策、企业的交易票据信息和上下游企业的影响力,向实力强、供求关系稳定的企业提供贷款,并可以对信誉高、信贷风险小的企业给予利率优惠。银行首先根据中小微企业的实力、信誉对其信贷风险做出评估,然后依据信贷风险等因素来确定 [1] 是否放贷及贷款额度、[2] 利率和 [3] 期限等信贷策略。

某银行对确定要放贷企业的贷款额度为 10 100 万元;年利率为 4% 15%;贷款期限为 1 年。已经给出了 123 家有信贷记录企业的相关数据、302 家无信贷记录企业的相关

数据和贷款利率与客户流失率关系的 2019 年统计数据。该银行希望根据实际和已给的数据信息，通过建立数学模型研究对中小微企业的信贷策略，主要解决下列问题：

- 123 家有信贷记录企业的信贷风险进行量化分析，给出该银行在年度信贷总额固定时对这些企业的信贷策略。
- 在问题 1 的基础上，对 302 家无信贷记录企业的信贷风险进行量化分析，并给出该银行在年度信贷总额为 1 亿元时对这些企业的信贷策略。
- 企业的生产经营和经济效益可能会受到一些突发因素影响，而且突发因素往往对不同行业、不同类别的企业会有不同的影响。综合各企业的信贷风险和可能的突发因素（例如：新冠病毒疫情）对各企业的影响，给出该银行在年度信贷总额为 1 亿元时的信贷调整策略。

三、 问题分析

3.1 问题 1 的分析

问题一首先需要我们对 123 家有信贷记录企业的信贷风险进行量化分析。由于信贷风险主要集中在中小微企业可能发生的违约上，所以对信贷风险的量化分析实际上可以转化为对企业违约率的分析上。由于附件一已经给出了 123 家有信贷记录企业是否违约的数据，因此我们考虑抽取出进项发票信息和销项发票信息等数据特征，并结合是否违约的标签训练一个二分类器，从而预测企业的违约率。由于原始数据分布不均匀，数值差异化很大，因此需要完善有效的数据预处理方案。考虑到输入数据的可解释性差，所以基于线性模型人工建模的方案可能不太适用，为了充分挖掘输入数据中的隐藏信息，我们选择了全连接神经网络。

问题一其次需要我们对这些企业制定在年度信贷总额固定时的信贷策略。由于题目已经确定贷款期限为 1 年，所以我们不必考虑信贷策略中的贷款期限因素，后续我们的银行收益计算也是以一年为限度。我们将会给出的信贷策略包含两个方面：贷款额度、年利率。根据题目信息和生活经验，贷款风险低信用评级高的企业应该分配更大的信贷额度，并为其贷款年利率给予一定的优惠。考虑到了信用评级为 D 的企业不予贷款，因此我们需要对剩下的企业按照他们的预测违约率对银行的贷款总额进行划分，满足违约率低的企业将会得到更多的贷款额度。然后在每家企业贷款额度确定的情况下寻找最优的年利率，这可以转换成关于年利率的最优化问题，最优化目标是企业的年收益最大和客户流失率最小，并且我们仅考虑影响年收益和客户流失率的众多影响因素中的年利率设置因素。求得每一个企业的贷款额度和年利率即为最终的贷款策略。

3.2 问题 2 的分析

问题二相较于问题一的差别主要是给出的 302 家企业没有贷款记录，并且信贷总额被限制在 1 亿元，基本思路和问题一保持一致。首先运用问题一中的数据预处理方式得到模型的输入特征矩阵和对应的标签，喂入全连接网络模型后得到每一个企业的预测违约率。与问题一不同的是，这 302 家企业的信誉评级数据缺失，所以我们先要根据附件一中的信誉等级与违约率的关系建立模型，然后对这 302 家企业的信誉评级进行补全，最后按照与问题一相同的方法确定每一个企业的贷款额度和年利率即为最终的贷款策略。

3.3 问题 3 的分析

首先补充说明我们后续会用到的对企业实力的量化分析方案，我们考虑进项和销项发票信息中价税合计的统计量（最大值，均值，和），企业信贷风险以及企业是否处于社会热门行业共 7 个因素，并建立层次分析法模型得到 302 家企业的量化实力权重向量，向量值越大的企业代表相对实力越强。

问题三要求对突发因素下的信贷策略进行调整，使模型更加接近于真实情况，实际上只需对之前模型的预测违约率进行人工微调即可。我们可以首先根据突发因素的影响行业确定违约率的修正方向，不同类型的企业修正方向是不同的，如疫情的爆发等对医疗有关企业有促进作用，而对食品有关企业有抑制作用；然后根据企业实力和突发因素的剧烈程度确定违约率的修正范围，一般来说，规模大(综合实力强)的企业受突发因素的影响较小，规模小（综合实力弱）的企业受突发因素的影响较大，另外如果突发因素客观上非常剧烈，对社会影响很大，那么对应的违约率的修正范围也会相应的增大。最后使用微调过的预测违约率按照与问题二相同的方法确定最终的信贷策略。

四、 模型假设以及符号说明

4.1 模型假设

- 假设附件一和附件二的数据分布相同，以便我们根据使用附件一训练的模型有效预测附件二企业的违约率。
- 假设银行给与中小微企业的信贷额度都能被企业完全使用，以便我们计算信贷策略确定时的银行年收益。
- 假设客户流失率仅仅与银行的年利率设置有关，不考虑更多因素的影响

4.2 变量及常量符号说明

变量或常量符号	变量或常量符号解释
P_i	企业 i 的违约率
Q_i	企业 i 的信誉评级，可选取值为 A，B，C，D
β_k	信用评级为 k 的企业的基准年利率
ρ_i	企业 i 的年利率
M_i	企业 i 的信贷额度
θ_i	企业 i 的实力权重
N	银行每年平摊在每笔贷款上的固定成本
K	银行的信贷总额
W	银行从借贷中获得的总收益
α	突发因素对违约率的基准影响程度

表 1 文中出现变量与常量符号及其解释

五、模型的建立与求解

5.1 问题一数据预处理

数据预处理的目的是从附件一抽取出进项发票信息和销项发票信息中的数据特征，构造数量足够的特征矩阵，并为每一个矩阵贴上是否违约的标签，作为输入数据训练一个二分类器。对于附件二没有信贷记录的企业的预处理只需提取特征矩阵即可。

我们面临的挑战包括：数据分布不均衡，即附件一中没有违约的企业数量远远大于有违约记录的企业数量；数值差异巨大，对于金额，税额，价税合计的取值差异很大，以至于均值归一化后计算机无法表示（只能表示为 NaN）。对此我们运用了重采样和正态归一化的方法。

六、其它小功能

6.1 脚注

利用 `\footnote{具体内容}` 可以生成脚注¹。

6.2 无序列表与有序列表

无序列表是这样的：

- one
- two
- ...

有序列表是这样子的：

1. one
2. two
3. ...

6.3 字体加粗与斜体

如果想强调部分内容,可以使用加粗的手段来实现。加粗字体可以用 `\textbf{加粗}` 来实现。例如：**这是加粗的字体。This is bold fonts**。

中文字体没有斜体设计，但是英文字体有。斜体 *Italics*。

七、参考文献与引用

参考文献对于一篇正式的论文来说是必不可少的，在建模中重要的参考文献当然应该列出。 \LaTeX 在这方面的功能也是十分强大的，下面介绍一个比较简单的参考文献制作方法。有兴趣的可以学习 `bibtex` 或 `biblatex` 的使用。

\LaTeX 的入门书籍可以看《 \LaTeX 入门》[1]。这是一个简单的引用，用 `\cite{bibkey}` 来完成。要引用成功，当然要维护好 `bibitem` 了。下面是个简单的例子。

参考文献

[1] 刘海洋. \LaTeX 入门[J]. 电子工业出版社, 北京, 2013.

[2] 全国大学生数学建模竞赛论文格式规范 (2020 年 8 月 25 日修改).

¹脚注可以补充说明一些东西

[1] <https://www.latexstudio.net>

附录 A 模板所用的宏包

表 2 宏包罗列

模板中已经加载的宏包				
amsbsy	amsfonts	amsgen	amsmath	amsopn
amssymb	amstext	appendix	array	atbegshi
atveryend	auxhook	bigdelim	bigintcalc	bigstrut
bitset	bm	booktabs	calc	caption
caption3	CJKfntef	cprotect	ctex	ctexhook
ctexpatch	enumitem	etexcmds	etoolbox	everysel
expl3	fix-cm	fontenc	fontspec	fontspec-xetex
geometry	getttitlestring	graphics	graphicx	hobsub
hobsub-generic	hobsub-hyperref	hopatch	hxdetex	hycolor
hyperref	ifluatex	ifpdf	ifthen	ifvtex
ifxetex	indentfirst	inwarerr	intcalc	keyval
kvdefinekeys	kvoptions	kvsetkeys	l3keys2e	letltxmacro
listings	longtable	lstmisc	ltxcaption	ltxcmds
multirow	nameref	pdfescape	pdftexcmds	refcount
rerunfilecheck	stringenc	suffix	titletoc	tocloft
trig	ulem	uniquecounter	url	xcolor
xcolor-patch	xeCJK	xeCJKfntef	xeCJK-listings	xparse
xtemplate	zhnumber			

以上宏包都已经加载过了，不要重复加载它们。

附录 B 排队算法–matlab 源程序

```

kk=2; [mdd,ndd]=size(dd);
while ~isempty(V)
    [tmpd,j]=min(W(i,V)); tmpj=V(j);
    for k=2:ndd
        [tmp1,jj]=min(dd(1,k)+W(dd(2,k),V));
        tmp2=V(jj); tt(k-1,:)= [tmp1,tmp2,jj];
    end
    tmp=[tmpd,tmpj,j;tt]; [tmp3,tmp4]=min(tmp(:,1));
    if tmp3==tmpd, ss(1:2, kk)=[i;tmp(tmp4,2)];
    else, tmp5=find(ss(:,tmp4)~=0); tmp6=length(tmp5);
    if dd(2,tmp4)==ss(tmp6,tmp4)
        ss(1:tmp6+1, kk)=[ss(tmp5,tmp4);tmp(tmp4,2)];
    else, ss(1:3, kk)=[i;dd(2,tmp4);tmp(tmp4,2)];
    end; end
    dd=[dd, [tmp3;tmp(tmp4,2)]]; V(tmp(tmp4,3))=[];
    [mdd,ndd]=size(dd); kk=kk+1;
end; S=ss; D=dd(1,:);

```

附录 C 规划解决程序—lingo 源代码

```

kk=2;
[mdd,ndd]=size(dd);
while ~isempty(V)
    [tmpd,j]=min(W(i,V)); tmpj=V(j);
    for k=2:ndd
        [tmp1,jj]=min(dd(1,k)+W(dd(2,k),V));
        tmp2=V(jj); tt(k-1,:)= [tmp1,tmp2,jj];
    end
    tmp=[tmpd,tmpj,j;tt]; [tmp3,tmp4]=min(tmp(:,1));
    if tmp3==tmpd, ss(1:2, kk)=[i;tmp(tmp4,2)];
    else, tmp5=find(ss(:,tmp4)~=0); tmp6=length(tmp5);
    if dd(2,tmp4)==ss(tmp6,tmp4)
        ss(1:tmp6+1, kk)=[ss(tmp5,tmp4);tmp(tmp4,2)];
    else, ss(1:3, kk)=[i;dd(2,tmp4);tmp(tmp4,2)];
    end;
end
dd=[dd, [tmp3;tmp(tmp4,2)]]; V(tmp(tmp4,3))=[];
[mdd,ndd]=size(dd);
kk=kk+1;
end;
S=ss;
D=dd(1,:);

```