工业企业财务稳健性与风险预警识别统计分析案例

**摘要**

财务稳健性是企业可持续经营的必要条件。本案例使用全国36万家工业企业87个维度的财务数据进行统计分析，实现从数据中挖掘蕴含信息的目标。通过数据可视化展现了数据潜在规律和价值；通过对数据缺失值和异常值的处理，剔除了干扰信息对建模的影响，实现了一般和特殊情况统一分析；根据分析目标构建了指标体系，结合经典统计分析方法，将描述统计中发现的规律定量化描述；同时针对本数据量较大，财务变量采集全面，分析中适时引入深度学习方法；此外文中对比了经典方法和深度学习方法估计效果。总体而言，本案例数据为统计方法和实际应用分析提供了良好的数据基础，即可以采用标准的分析流程探寻一般规律，也满足多角度，个性化数据挖掘条件，满足统计学教学案例需要。

**关键词：工业企业 财务分析** **统计学习 推断统计 描述统计**

# 引言

本案例库是对中国工业企业财务数据进行深入统计分析和研究，对工业企业不同行业地域分布，发展演变规律进行深入挖掘，并利用工业企业财务数据对工业企业财务稳健性和风险进行预警以及识别。本案例数据来源为中国工业企业数据库的统计是基于国家统计局进行的“规模以上工业统计报表统计”取得的资料整理而成。数据库的统计对象为规模以上工业法人企业，包括全部国有和年主营业务收入500万元及以上的非国有工业法人企业，与《中国统计年鉴》的工业部分和《中国工业统计年鉴》中的覆盖范围一致。区别是本数据库是企业层面的原始数据，而“年鉴”是按不同维度得到的加总数据。因此在前期进行大量的数据处理与挖掘的工作，并为后期数据分析提供良好的数据基础。然后，在对数据进行探索性分析的基础上，根据研究目的，确定数据变量和研究方案，并根据实际研究需要，选择可行的统计模型方法。

在内蒙古财经大学研究生院与统计与数学学院的大力支持下，利用中国工业企业真实财务数据，按照统计分析规范，逐一演示了数据预处理、基本数据分析和深入数据研究等商业数据统计分析流程，展示了从数据到结论的统计学分析框架。

# 相关背景及意义

## 研究背景

一般而言，企业所面临的风险分为系统性风险以及非系统性风险。所谓系统风险，也称市场风险，用β风险系数衡量，如政策风险，宏观风险，汇率风险，自然灾害等。是指由于企业外部、不为企业所预计和控制的因素造成的风险。政策风险，指因政府制定相关政策而带来的行业影响所造成的风险。宏观风险，譬如全球通胀加剧的风险以及产业链供应链中断风险。流动性风险，根据2009年银监会印发的《商业银行流动性风险管理指引》中将流动性风险定义为商业银行无法及时获得充足资金或无法以合理成本及时获得充足资金以应对资产增长或支付到期债务的风险。这里将流动性风险定义为工业企业在正常存续阶段无法及时获得充足资金或无法以合理成本及时获得充足资金以应对资产增长或支付到期债务的风险。企业的汇率风险，一般指对外贸易企业所面临本币与外币的兑换而导致的风险。自然灾害，是指给人类生存，经济发展，社会进步等各方面带来危害或损害人类生活环境的自然现象，包括[干旱](https://baike.so.com/doc/1671386-1767150.html" \t "https://baike.so.com/doc/_blank)、[高温](https://baike.so.com/doc/6194374-6407633.html" \t "https://baike.so.com/doc/_blank)、[低温](https://baike.so.com/doc/6425150-6638822.html" \t "https://baike.so.com/doc/_blank)、[寒潮](https://baike.so.com/doc/10038466-10516605.html" \t "https://baike.so.com/doc/_blank)、[洪涝](https://baike.so.com/doc/6025169-6238166.html" \t "https://baike.so.com/doc/_blank)、[山洪](https://baike.so.com/doc/6180464-6393710.html" \t "https://baike.so.com/doc/_blank)、[台风](https://baike.so.com/doc/7189817-7414015.html" \t "https://baike.so.com/doc/_blank)、[龙卷风](https://baike.so.com/doc/5407698-5645630.html" \t "https://baike.so.com/doc/_blank)、瘟疫流行病等，或一些不可预见或难以预防的重大灾难性事件，会对一个国家经济建设，社会发展带来重大影响甚至破坏。一般来说，系统风险是指宏观层面的风险，对于系统性风险的预防以及处理都比较困难，后果带有普遍性。

非系统性风险，也称α风险。是指发生于个别企业的特有事件造成的风险，包括财务风险、企业战略布局、信用风险、偶然事件风险等。非系统性风险只存在于相对独立的范围内，或个别行业中，它来自于企业内部的微观因素，与整个金融市场或国家经济状况不发生系统性的联系，对于企业自身运营状况和战略发展危害性较大，一般对于国家整体经济状况影响较小，但一些处于国家经济命脉的龙头企业如若发生财务风险等，也会对国家整体经济发展造成不可估量的损失。财务风险，一般涉及企业运营过程中资金的流动性出现问题，现金流的断裂或高的金融杠杆导致长期或短期的债务问题发生。

财务危机，在国外通常称为财务破产、经营失败。企业财务危机的发生会

对企业形象、企业利益带来严重打击，同时还会引发巨大的舆论压力，因此国内外学者对其愈发关注。国外的学者在上世纪60年代就对财务危机问题展开了研究，大多数学者对于财务危机的界定一般分为以下两种。以从企业出现明显的迹象表明其有严重的财务风险如拖欠优先股股利、拖欠债务等都可以认定为产生了财务危机，包括从资金管理上的技术性失败到破产性失败，以及处于两者之间发生的各式情况。技术性失败主要指企业缺乏足够的短期净现金流来支付流动负债，破产性失败是指企业到期净资产为负，资不抵债的情况而另一部分国外研究学者直接将宣告破产的企业认为是发生了财务危机的企业，即认为财务危机等同于企业发生破产行为。由于国外企业主动申请破产的行为真实准确，可以明确记录，方便确定研究样本；其次企业的破产更能引起企业管理者及相关利益方的重视，可以促进后续财务危机预测模型的应用及发展，因而近年来国外研究者在学术研究中直接将企业陷入财务危机定义为企业破产。

从宏观角度来说，企业所面临的系统性风险一般难以预测，且在应对风险过程中也十分棘手，是对企业管理层决策是否合理，企业内外部风控体系是否完善，企业战略资源配置是否恰当，企业财务方面如：资金流是否充沛，企业金融杠杆是否合理，长短期偿债能力是否具备等多个方面的综合考验。但从微观角度来说，企业在运营过程中所面临的非系统性风险可以在统计学、经济学理论支称下，依靠企业每年的财务数据进行合理而有效的分析以及预测，提出合理恰当的参考方案，使企业管理层有效的监控企业财务状况，防范可能会出现的风险，确保企业在市场中不被淘汰且建立优势，有效规避或转移风险，保护企业自身，保护法人、最终受益人以及其他关联方利益，避免财务危机的发生。

## 研究意义

### 宏观层面

一方面，对于我过资本市场、金融市场来说，建立行之有效的财务预测系统可以促进我国资本市场的平稳发展和有效运作,也可以一帮助银行等金融机构识别企业的质量和偿债能力,减少不良贷款的发生,提高资金的利用效率,避免贷款风险防止坏账发生，减少，甚至阻断个别行业龙头企业破产而导致的社会上的系统性风险。

另一方面，对于我国监督机构来说，财务预测系统可以帮助政府监管者把握市场风险,确定监管重点,有效地进行资源配置,对监控企业质量和减少证券市场风险有着重要的现实意义。基于传统工业企业财务稳健性评价的研究方案，局限于小规模样本数据和已有知识背景下的指标体系。深度学习方法可以通过对超大规模样本的数据分析，找到此前未知的关系，并在此基础上探究合理的财务指标信息与企业盈利，经营能力的关系，对于企业健康发展，避免财务危机有着重要作用。

### 微观层面

一方面，对于企业自身来说，企业财务预测系统的建立可以帮助市场经济主体降低运营风险,保障企业的正常运行，企业的管理层可以根据预警信号掌握企业当前的财务状态、预知企业潜在的风险和危机,从而在战略、企业管理和内部控制、市场定位等诸多方面进行调整,扭转企业经营状况恶化的趋势,以避免沦为的破产风险行列。另外,企业越早获得财务危机信号,越可以减少其在会计、审计、律师等方面所支付的费用。

另一方面，对于审计人员来说,财务危机预测系统能帮助审计人员确定审计范围,制定必要的审计程序,并帮助判断被审计公司能否维持经营或保持良好的前景,进而提高审计人员的评估能力,降低审计风险,避免因未能正确披露其经营失败而招致的法律诉讼。特别是当审计人员与客户发生意见分歧时,有客观分析模型加以支持,客户往往会更容易接受审计意见。

总之，风险管理是企业管理的非常重要的组成部分，风险管理的好坏直接关系到企业的运营生存和发展。而财务风险管理在企业风险管理中具有重要地位，财务风险预警是企业财务风险管理的核心和重点。而工业制造业在国民经济中占有重要的地位，是国民经济的支柱产业，直接影响着国计民生。因此对于工业企业财务分析预警的研究就显得尤为重要。

# 主题内容

## 数据概览与数据转化

### 工业企业地域分布描述性分析

本文利用工业企业财务数据中所涉及的地域性数据以及产值数据进行探索性分析，利用地图演变规律呈现不同行业发展规律。数据中包含了40个行业企业5年发展变动数据，以此数据真实反映了我国工业企业在全国的地理分布，集聚演变路径，下文以通用设备制造业、电力生产和供应业、黑色金属冶炼和压延加工业、医药制造业等为例，展示各地区产业和内蒙古地区发展区别，反应我国个别行业的发展变动。

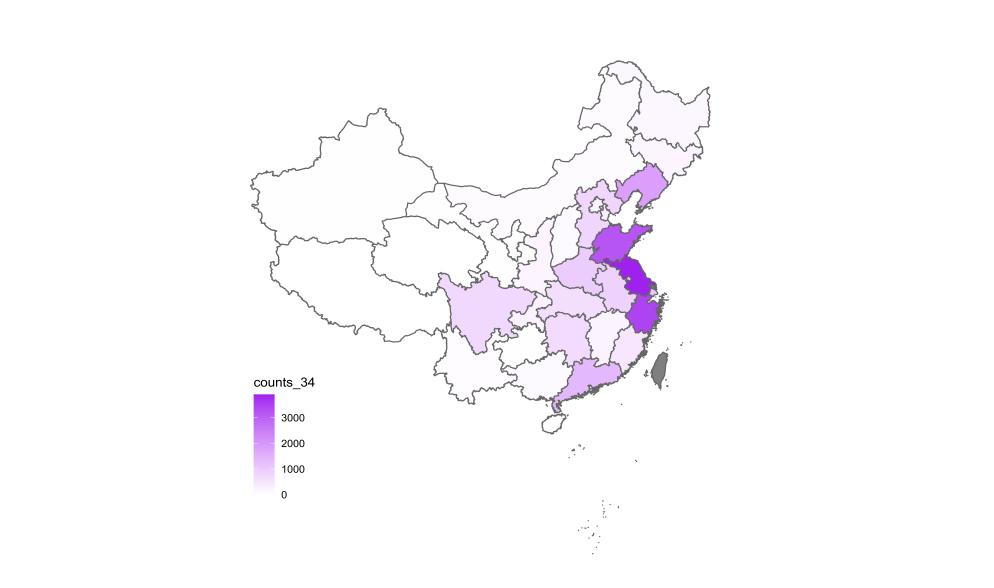
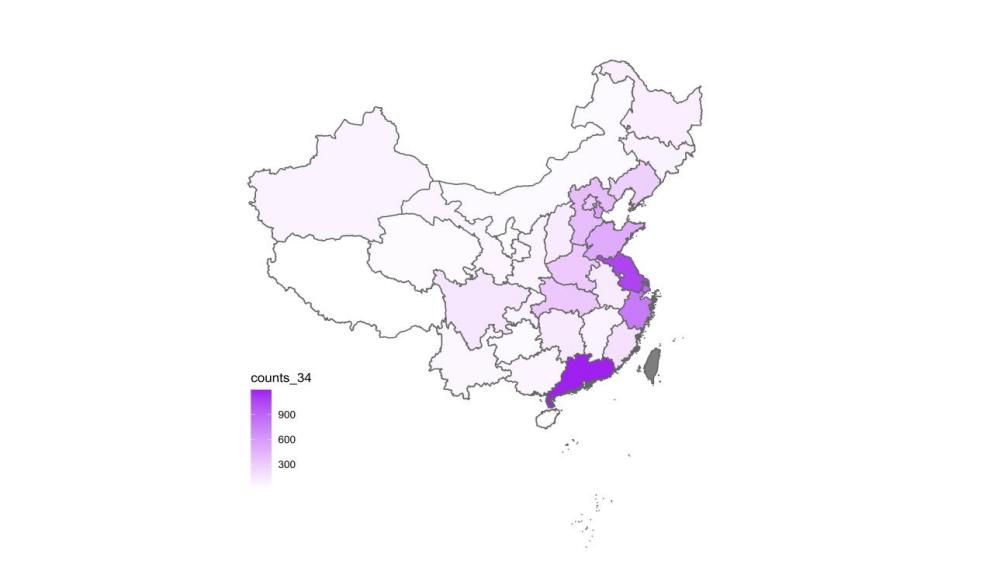


图 1 通用设备制造业演变规律

（1）通用设备制造业

上图为我国通用设备制造业（行业代码23）1998年（左）与2013年（右）企业数量对比图，众所周知，我国制造业的快速发展，逐渐已经成为了“世界工厂”。生产设备是一国制造业的基础,它决定了制造业发展的上限。通用设备是制造业生产设备中不可或缺的重要组成部分。通用设备包括机械基础件、机床工具、制冷空调设备、磨具磨料和内燃机等。

从上图我们不难看出我国通用制造业工厂总体规模持续增加，沿海地区且拥有众多劳动力的省份成为了主要生产地区。且通用制造业中心已经发生转移，由1998年以广东省、江苏省为中心，变为以江浙两地及山东省为中心的延东海地区。推测东部地区生产的通用设备更易运输到其需求地，或者主要贸易国从东海出海更节约资本，亦或者东部地区离生产所需原料原产地近，更适和建厂。也有可能是由于广东省的生产重心发生变化，不以通用设备制造业为主。

（2）电力生产和供应业

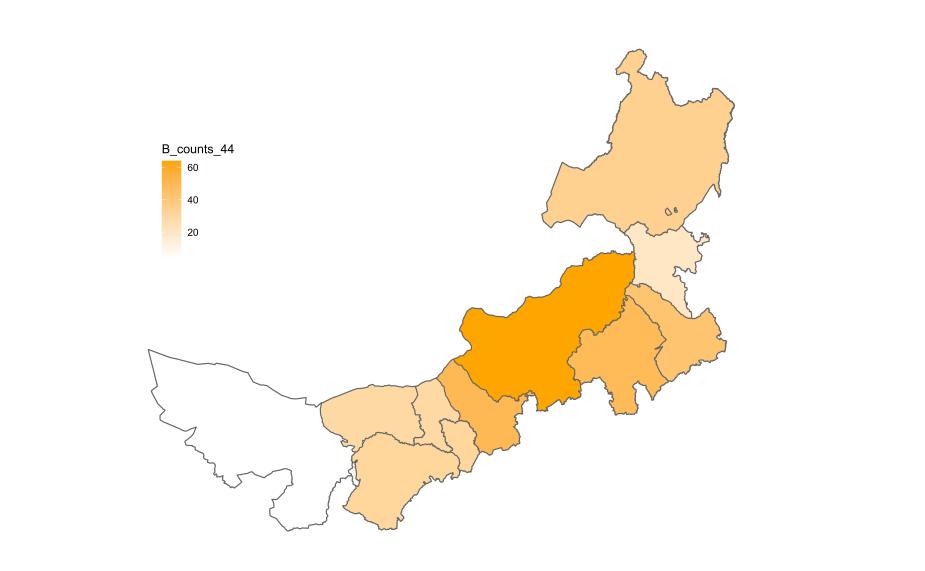
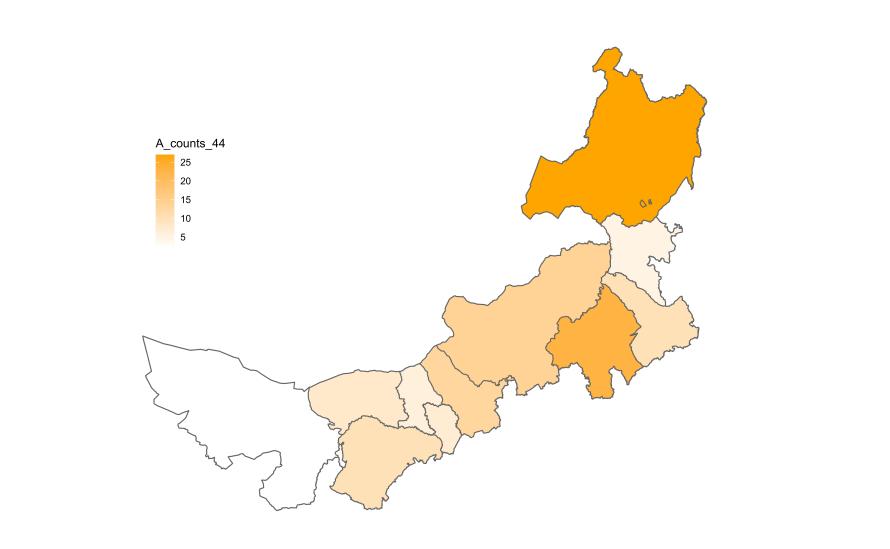


图 2 电力生产和供应业

上图为内蒙古自治区电力生产和供应业（行业代码44）1998年（左）与2013年（右）企业数量对比图，内蒙古作为我国西电东送的主要省份之一，将源源不断的电力输送到京津冀地区。内蒙古地区依托其地理优势，拥有火力、电力、水力、太阳能发电等多种发电类型。

内蒙古煤炭资源丰富，故储煤丰富的西部地区主要采用火力发电，但火力发电因为其自身的局限性，以及一些人为因素，大力的发展也对当地环境造成了不可小视的破坏。故随着社会现代化建设，可持续发展观念深入人心，西部地区并未成为内蒙古自治区发展的主力军。内蒙古自治区风能储量2.7亿千瓦小时，占中国总储量的1/5，居中国首位；全区年平均风速3.7米/秒，大部地区年平均有效风能功率密度为150-200瓦/平方米，其中以赤峰市及锡林郭勒盟、乌兰察布市风速为最。故随着风力发电技术的不断精进，上述三地成为内蒙古自治区发电的主要地区。且由于锡林郭勒盟的地广人稀、土地平旷，其发电公司数为最。

内蒙古自治区海拔较高，地处中纬度高海拔地区，以温带大陆性气候为主，全年降水较少，多晴朗天气，云量低，日照时间较长，日照时数在2300~3400h之间。太阳能辐射较强，全区年辐射总量仅次于青藏高原，居全国第二。其有着极佳的优势发展太阳能发电，但由于太阳能发电技术的不成熟，内蒙古自治区太阳能发电还在不断将进行摸索和尝试，故与发电量相较于南方水利资源丰富地区的水力发电不纳入问题分析。

同时呼伦贝尔市作为内蒙古自治区1998年发电量最高的区，到2013年企业数量几乎未有变化，推测由于东北老工业基地的没落及化石燃料的减少，限制了呼伦贝尔的发展。同时呼伦贝尔的风速也不低，但发电公司却未达到锡林郭勒盟的水平，推测其成因可能是由于呼伦贝尔市纬度过高，冬季及其漫长且寒冷，导致发电或储电受限，或呼伦贝尔地价较其他地区贵，导致缺少电力公司。

（3）黑色金属冶炼和压延加工业

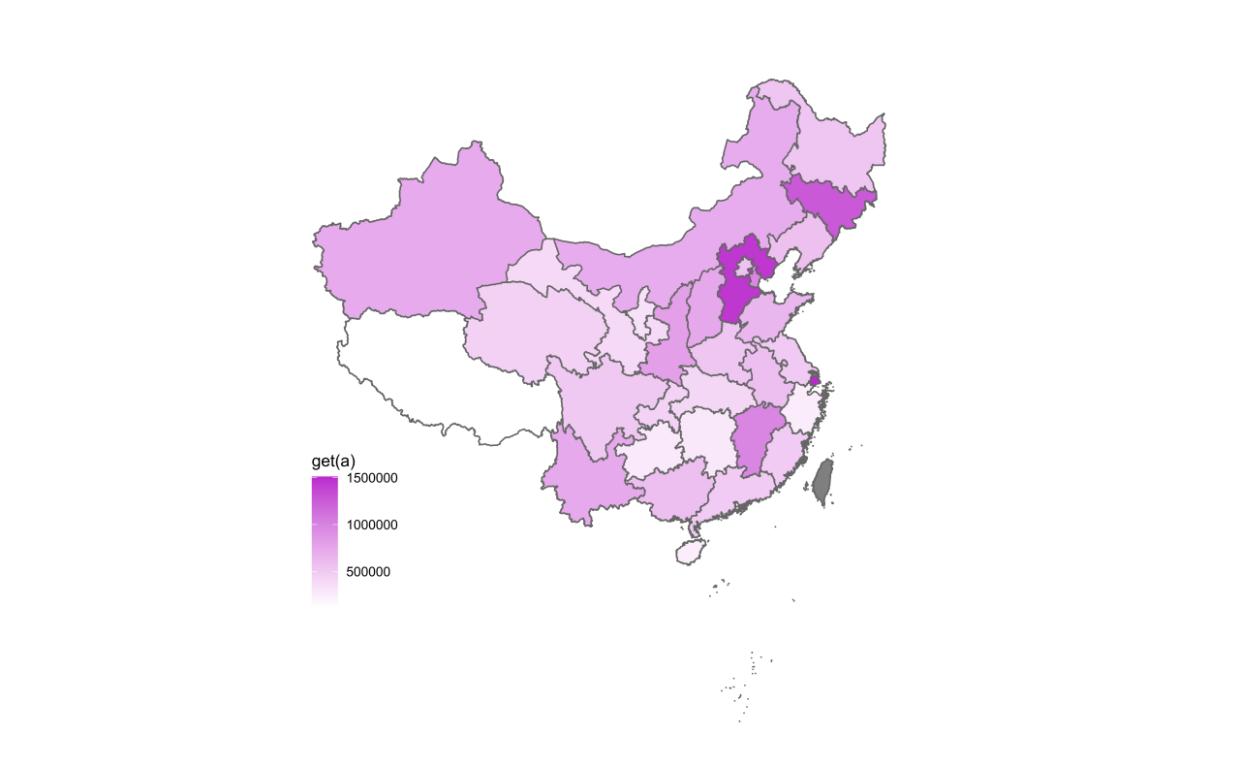
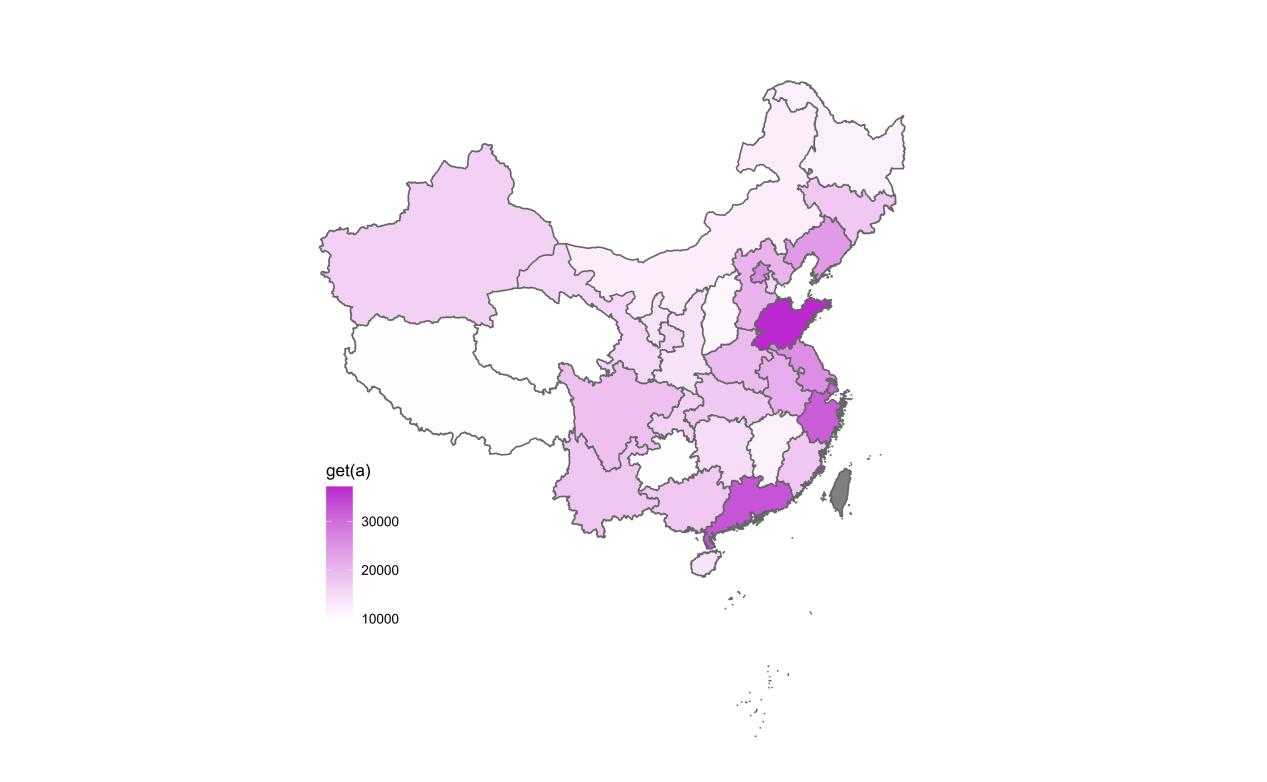


图 3 黑色金属冶炼和压延加工业

上图为我国黑色金属冶炼和压延加工业（行业代码31）1998年（左）与2013年（右）平均产值（总产值除以总企业数）对比图。[黑色金属材料](https://baike.baidu.com/item/%E9%BB%91%E8%89%B2%E9%87%91%E5%B1%9E%E6%9D%90%E6%96%99/1878750" \t "https://baike.baidu.com/item/%E9%BB%91%E8%89%B2%E9%87%91%E5%B1%9E/_blank)乃工业上对[铁](https://baike.baidu.com/item/%E9%93%81" \t "https://baike.baidu.com/item/%E9%BB%91%E8%89%B2%E9%87%91%E5%B1%9E/_blank)、[铬](https://baike.baidu.com/item/%E9%93%AC" \t "https://baike.baidu.com/item/%E9%BB%91%E8%89%B2%E9%87%91%E5%B1%9E/_blank)和锰、钒矿和钛矿等的统称。亦包括其几种金属的合金，尤其是合金黑色金属钢及钢铁。黑色金属的分类是因为这三种金属都是冶炼钢铁的主要原料，而钢铁在国民经济中占有极其重要的地位，亦是衡量一国家国力的重要标志。黑色金属的产量约占世界金属总产量的95%。

1998年我们企业的平均产值较高的地区为广东省、浙江省、山东省三地，2013年平均产值较高的地区为吉林省和河北省。分析其成因在于1998年时，我国拥有高超的冶炼黑色金属技术的地区在于发达的沿海地区，且沿海地区拥有发达的交通，便于将黑色金属原料运输如企业，同时也便于将冶炼好的金属输送去制造业工厂加工。随着我国交通运输业的发展，及西部等拥有丰富矿产资源的地区技术的进步，我国的黑色金属冶炼和压延加工业的企业平均产值整体提高，且峰值大都分布于拥有丰富矿产资源且人口密度低的地区。由于冶炼钢铁等需要排出大量的有害气体，故除了少部分发展多年的大型企业，大部分企业都远离东部沿海地区，整体西移。

表 1 不同行业企业数目（1）

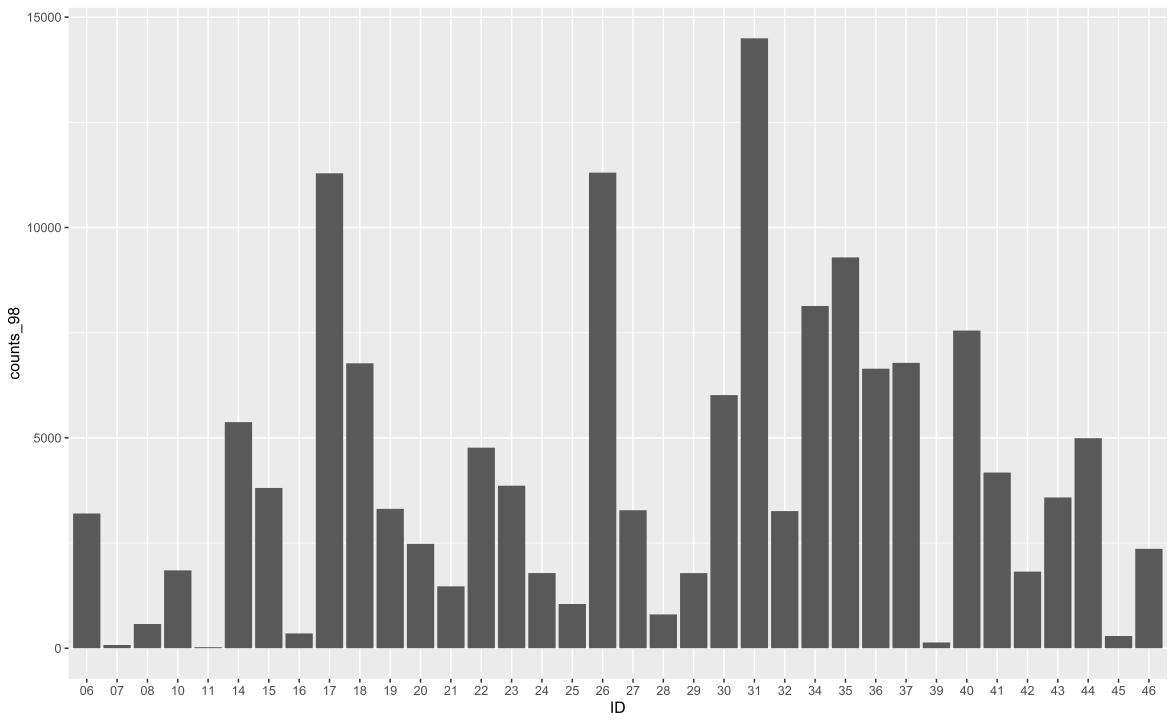
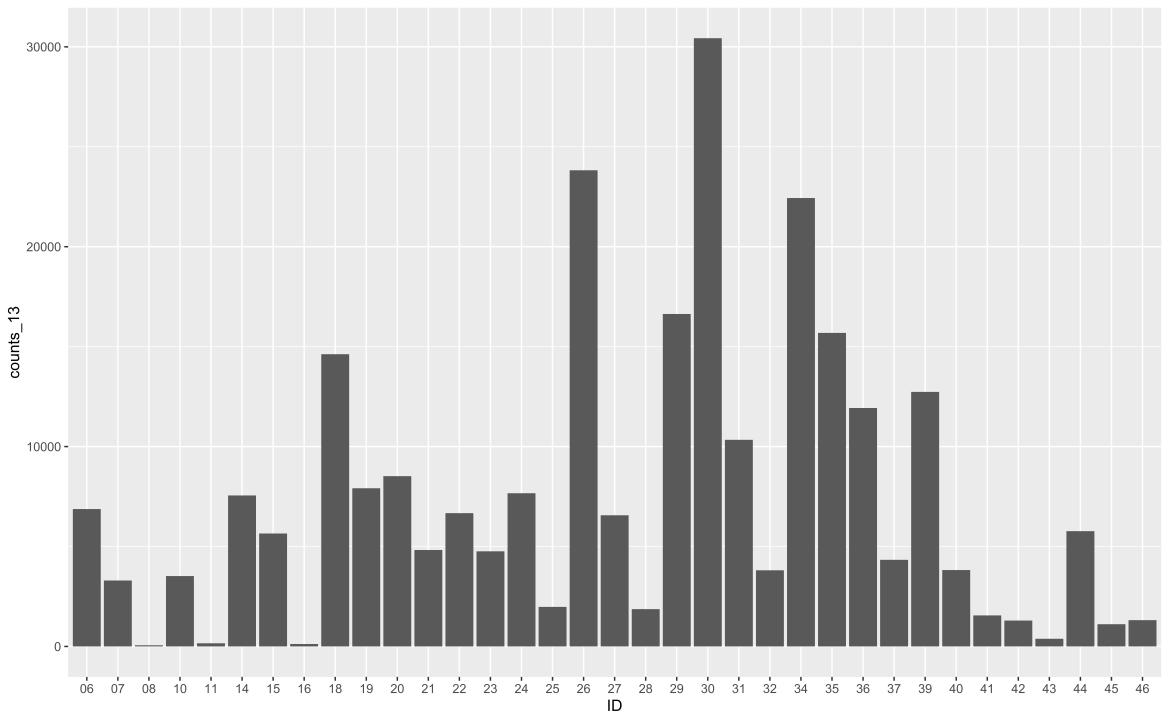


表 2 不同行业企业数目（2）



上图为我国1998年（上）与2013年（下）不同行业企业数目条形图。从上图可知，我国黑色金属冶炼和压延加工业的企业数目与其他数目飙升的行业不同，下降了约5000家企业，从1998年各行各业中的龙头跌落至末尾。排除国家政策的推广，原料的产品提升也可能是主要原因之一。

随着中国钢铁产业发展的需要，铁矿石进口量不断攀升，世界铁矿石供应商凭借自身的资源禀赋优势和在世界铁矿石贸易市场上的垄断地位，操纵市场价格，致使世界铁矿石价格不断上涨。2010年，世界三大铁矿石供应商——巴西淡水河谷公司、澳大利亚必和必拓公司以及力拓公司提出将铁矿石报价在 2009 年的价格基础上继续上涨 80%~90%，2011 年进口铁矿石年均价格高达 163.84 美元 /t，达到了史无前例的高价。催生全球新增产能加快释放，铁矿石供大于需的趋势逐渐显现，致使随后几年铁矿石生产企业急剧下降。

结合四图不难推测，沿海港口地区的铁矿石来源除了本身当地的储量，极大部分来源于国外进口，随着我国制造业的不断发展，对铁矿石的需求量不断增大，致使沿海黑色金属冶炼和压延加工业企业平均产值相较同时期内陆地区大幅下降。

废钢是现代钢铁工业不可缺少的铁素原料，也是唯一可以大量替代铁矿石的绿色原料，是节能载能的再生资源。与用铁矿石生产1t钢相比，用废钢生产1t钢，可节约铁矿石1.65t，降低能源消耗约合350 kg标准煤，减少约1.4t的CO2排放和约3t的固体废物排放。

受全面取缔“地条钢”影响，废钢资源流向发生明显变化，导致废钢价格下跌，废钢替代铁矿石的成本优势突显，钢铁企业开始加大废钢铁的使用量，在环保的重压下，电炉炼钢作为环保生产工艺，占比持续提升，进一步加大了对废钢的需求，京津冀等“2+26”地区的企业大量使用废钢，以增加钢材产量。由此河北省作为一个丝毫不占据上述优势的地区，在2013年有着高的黑色金属冶炼和压延加工业企业平均产值。

随着我国钢铁蓄积量和废钢资源产生量的增加及废钢价格优势，未来，我国的废钢资源总量将非常充足。未来充足的废钢资源将是我国钢铁工业强有力的支撑，同时也是实现黑色金属资源强国的重要支撑。我国对国际铁矿石资源的需求量将逐步下降，废钢资源产出量的大幅增加必将推动钢铁行业加快转型升级，并对钢铁工业生产流程结构的调整。故而现有模式必将被打破，拥有丰富矿产资源的地区将会限制冶炼，率先掌握废铁回收再利用的企业，势必成为未来行业龙头。

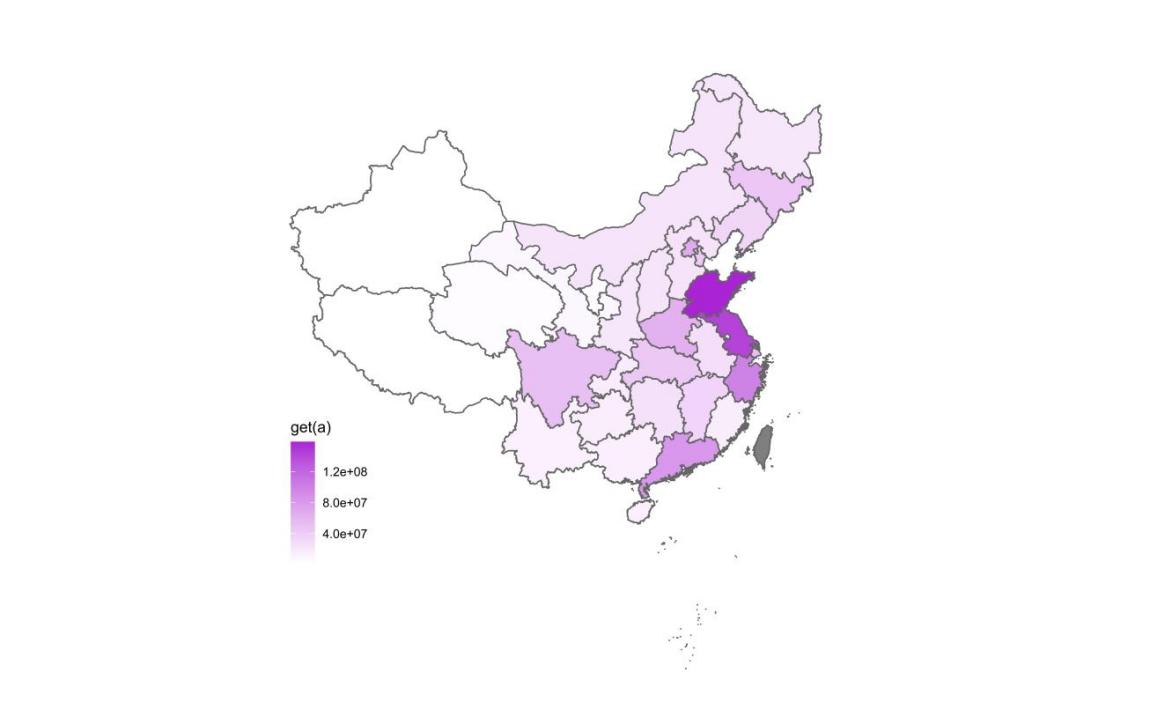
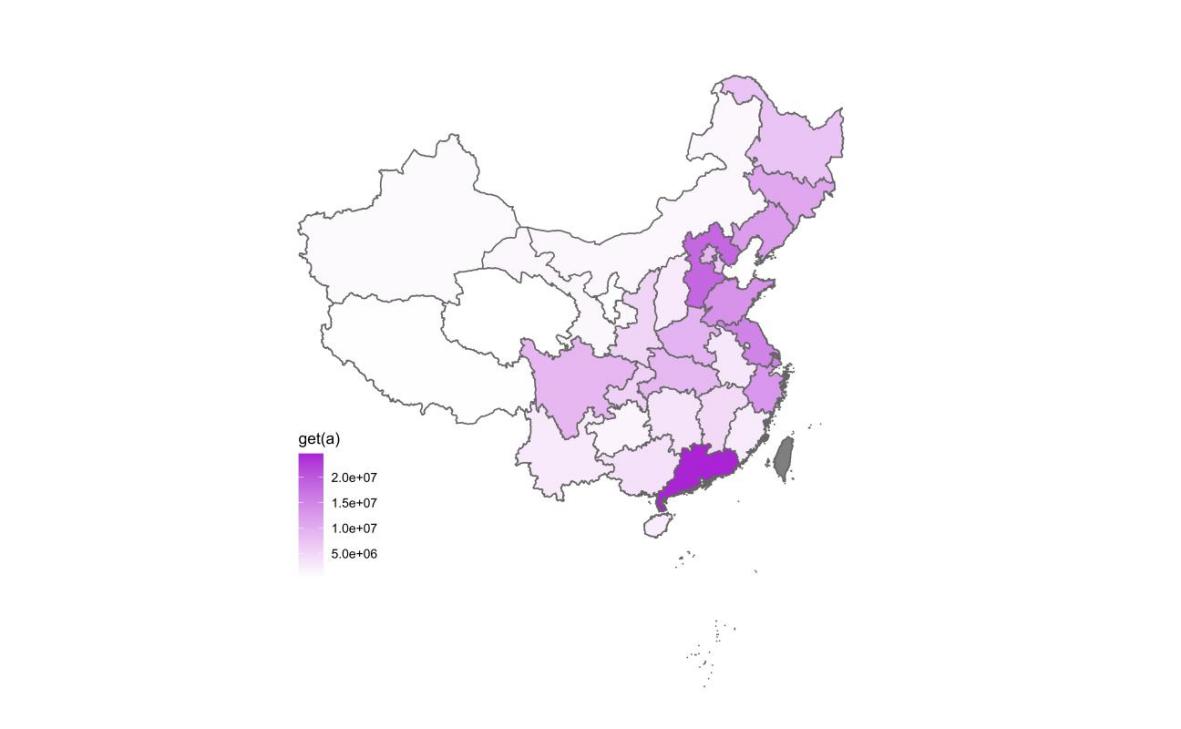


图 4 医药制造业

（4）医药制造业

上图为我国医药制造业（行业代码27）1998年（左）与2013年（右）产业规模对比图。人口自然增长及经济水平的提升带来医疗卫生费用的增长，我国是总人口数超过14亿的超级人口大国，且作为发展中国家，医疗卫生事业处于上升发展阶段。随着经济水平的增长，人们对生命健康的重视,医疗卫生总费用不断增长,医药企业得到快速发展，盈利能力和盈利水平不断增强。社会老龄化趋势明显，医疗保健需求增加，老龄人口对医疗、保健等消费支出较之其他年龄人口更为庞大，日益增多的老龄人口将提升医疗卫生行业的市场容量，促进医药消费量、保健品等行业的协同发展。

我国生物制药产业起步较晚，直到70年代初才开始将DNA重组技术应用到医学上，但在国家产业政策的大力支持下使这一领域发展迅速逐步缩短了与先进国家的差距。经过了数十年的发展，以基因工程药物为核心的研制、开发和产业化已经颇具规模，主要分布于环渤海、长三角、珠三角等经济发达地带。

由上图可知，我国的医疗行业发展迅猛，且以沿海地区及内地医学发达地区为最。但由于历史、人才、技术、资源等原因，我国生物医药行业不同区域的产业集群正在形成。比如广东的医疗器械行业，浙江的化学药、云南的中药等。推测企业规模巨大的地区应以基本的机械化工厂为主，而非研发企业。

与发达国家相比，我国生物医药产业还处于比较落后的状态。目前制约我国生物制药产业发展的主要因素有资金短缺、研发力量薄弱、缺乏产业化机制、科研成果转化率低等。其中，作为生物医药领域内重要部分的中成药产业在国际上发展步履维艰，而我国的生物制药公司对创新研发重要性认识不够，对高端人才及产品研发资金投入不足，大多集中在生物仿制药领域。

生物医药作为新兴产业，有着良好的发展前景和蓬勃的生命力，引起了国家的足够重视，近年来，中央和地方政府都在不断加大对生物医药的投入力度，从政策和资金等各方面扶持生物医药产业。

### 财务数据描述性分析

作为世界制造业中心,制造业是我国最重要的行业之一。研究制造业企业的财务状况、建立制造业企业的财务预测模型具有非常深远的现实意义。此外,制造业企业是我国企业的主体,其整体财务状况在一定程度上能代表我国企业总体的财务伏况。以制造业为研究对象的另一个好处是能够获得足够的样本。

财务困境定义识别

财务状况的技术性界定是财务预警研究领域的难点和重点。国内大多的研究者都从测度的简易性出发,将公司是否被特别处理作为判断其是否为财务困境公司的标准。关于财务困境的技术性界定,本文认为吕长江在2003年提出的观点最具操作性和可测性。根据我国大多数上市公司流动负债多,并长年积累转化为长期负债,从而导致偿债危机的实际情况,他提出用流动比率来定义财务困境。他认为财务困境是指企业陷入流动资产小于流动负债即流动比率小于1的境地,并在一年内一直持续该状态而不能逆转。本文认为该定义考虑到我国的实际情况,并且揭示了财务危机的成因和特点。

变量解释

企业财务危机的发生是由企业较小的财务风险积聚而成，而财务风险隐藏在企业的日常经营活动当中。基于以上对财务危机预测体系指标选取原则的讨论，结合实际情况和财务预测的可能性，本文所设计的财务预测指标主要涵盖财务指标和非财务指标以两个方面。所选取的指标主要由偿债能力、盈利能力、经营能力、发展能力、现金流、风险能力、非财务指标。

偿债能力直接衡量了企业的财务状况，通常是指企业到期债务能否有力偿还。偿债能力指标是衡量企业能否长期经营，稳定发展的重要指标。该部分共选择7个指标。

表 3 偿债能力指标

|  |  |
| --- | --- |
| 偿债能力指标 |  |
| X1流动比率 | 流动资产/流动负债 |
| X2速动比率 | （流动资产-存货）/流动负债 |
| X3利息保障倍数 | （净利润+财务费用）/财务费用 |
| X4资产负债率 | 负债合计/资产合计 |
| X5产权比率 | 负债合计/所有者权益 |
| X6长期资本负债率 | 非流动负债/（非流动负债+所有者权益） |
| X7长期债务营运比率 | 非流动负债/（流动资产-流动负债） |

盈利能力为企业在一定期限内获取收益的能力，受到企业管理层及投资者的强烈关注。盈利能力的大小不但能够支撑企业偿还债务，还能够支持企业主要业务的发展，并提供了企业扩展经营范围的能力。盈利能力是评价企业业绩、企业

价值的重要指标，该部分共选择9个指标。

表 4 盈利指标

|  |  |
| --- | --- |
| 盈利指标 |  |
| X8资产报酬率 | （利润总额＋财务费用）／平均资产总额 |
| X9流动资产净利润率 | 净利润／流动资产平均余额 |
| X10固定资产利润率 | 净利润／固定资产平均余额 |
| X11净资产收益率 | 净利润／股东权益平均余额 |
| X12营业利润率 | 营业利润／营业收入 |
| X13总营业成本率 | 营业总成本／营业总收入 |
| X14毛利率 | 毛利/营业收入 |
| X15销售期间费用率 | （管理费用+销售费用+财务费用）/主营业务收入（销售收入净额） |
| X16成本费用利润率 | 利润总额/成本费用总额×100%。 |

经营能力反映了企业对自身资产及资金的掌控能力，资产与资金的周转可以直接影响企业的盈利能力。当企业的经营能力较强时，则表现为资产的周转与流动性较强，正向带动企业运营效率，促进企业获利。同时，经营能力也体现了企业的管理水平，是一项反映企业健康发展的重要能力，该部分共选择4个指标。

表 5 运营指标

|  |  |
| --- | --- |
| 运营指标 |  |
| X17存货周转次数 | 营业成本／存货平均占用额 |
| X18流动资产周转率 | 营业收入／流动资产平均占用额 |
| X19固定资产周转率 | 营业收入／固定资产平均净额 |
| X20总资产周转率 | 营业收入／平均资产总额 |

发展能力，可以被认为是企业的成长能力。在企业的发展中，持续保持资产规模扩大、盈利水平提升是企业做大做强的根本。发展能力指标能够反映企业潜在的成长性，成长性高的企业能够长期获得资金的关注，该部分共选择4个指标。

表 6 发展能力指标

|  |  |
| --- | --- |
| 发展能力指标 |  |
| X21总利润增长率 | （总利润本年金额—上年金额）/总利润上年金额 |
| X22营收增长率 | （营业收入本年本期金额-营业收入上年同期金额）/营业收入上年同期金额 |
| X23所有者权益增长率 | （所有者权益本期期末值—所有者权益上年同期期末值）/所有者权益上年同期期末值 |
| X24资本累计率 | (所有者权益期末值-期初值)/所有者权益期初值 |

风险水平直接反映了企业潜在的风险，主要有财务风险和经营风险。通常来讲，杠杆加的越高，企业需要承担的风险越大，企业在这种杠杆下的盈利与发展的压力越大。

表 7 风险水平

|  |  |
| --- | --- |
| 风险水平 |  |
| X25财务杠杆 | （净利润+所得税费用+财务费用）/ （净利润+所得税费用） |
| X26经营杠杆 | （息税前收益+固定成本)/息税前收益 |
| X27综合杠杆 | 财务杠杆 \* 经营杠杆 |

此外加入非财务数据变量，根据在企业内部从业人员数量可以反应一个企业的生产规模以及人员成本等。

表 8 非财务指标

|  |  |
| --- | --- |
| 非财务指标 |  |
| X28从业人数 | 从业人数 |

本文的初选指标参考了多篇国内外文献和大量专业书籍。计算出了各个文献中常用于财务分析的最终指标共28个，取值范围及中位数如下:

表 9 财务数据特征



表 10 偿债能力指标数据特征

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 流动比率 | 速动比率 | 利息保障倍数 | 资产负债率 | 产权比率 | 长期资本负债率 | 长期债务营运比率 |
| Min.: | -1125 | -305.32 | -232344 | -371.1333 | -40454 | -140.954 | -63483.83 |
| 1st\_Qu: | 0.9 | 0.13 | 1.2 | 0.31 | 0 | 0 | 0 |
| Median | 1.2 | 0.35 | 4 | 0.5308 | 1 | 0 | 0 |
| Mean: | 15.2 | 3.93 | 99.1 | 0.5412 | 443 | 0.1371 | -1.81 |
| 3rd\_Qu | 2 | 0.71 | 18.3 | 0.74 | 3 | 0.1236 | 0.04 |
| Max. | 688102.5 | 140281 | 402120 | 536.9254 | 36756049 | 1237.4375 | 37889.88 |

表 11 盈利能力指标数据特征

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 销售期间费用率 | 成本费用利润率 | 存货周转次数 | 流动资产周转率 | 固定资产周转率 | 总资产周转率 | 总利润增长率 |
| Min.: | -1.63929 | -2.475046 | -293.4 | -897.208 | -9.3 | 0.001 | -13870.5 |
| 1st\_Qu: | 0.03825 | 0.008732 | 5.8 | 1.401 | 3.2 | 0.883 | -0.396 |
| Median: | 0.07977 | 0.03763 | 13.5 | 2.606 | 7 | 1.559 | 0.044 |
| Mean: | 0.10718 | 0.059149 | 155 | 9.917 | 42.2 | 3.166 | 0.195 |
| 3rd\_Qu: | 0.13582 | 0.09024 | 43.3 | 6.741 | 16.1 | 3.312 | 0.451 |
| Max.: | 117.42253 | 8.827498 | 519679 | 8674.2 | 904647 | 6675.333 | 6395 |

表 12 运营能力指标数据特征

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 资产报酬率 | 流动资产净利润率 | 固定资产利润率 | 净资产收益率 | 营业利润率 | 总营业成本率 | 毛利率 |
| Min.: | -41.0865 | -152.6487 | -1756 | -5978 | -7.037034 | 0 | -40.84007 |
| 1st\_Qu: | 0.019 | 0.0105 | 0.0286 | 0 | 0.007588 | 0.804 | 0.07293 |
| Median: | 0.0743 | 0.0731 | 0.1832 | 0 | 0.03514 | 0.8702 | 0.12471 |
| Mean: | 0.2322 | 0.5239 | 0.7002 | 267 | 0.045547 | 0.8512 | 0.13657 |
| 3rd\_Qu: | 0.2263 | 0.2987 | 0.6202 | 0 | 0.082597 | 0.9209 | 0.19 |
| Max.: | 1668.8 | 1291.5191 | 1488.25 | 23154806 | 20.200398 | 8.0991 | 1 |

表 13 成长能力指标数据特征

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 营收增长率 | 所有者权益增长率 | 资本累计率 | 财务杠杆 | 经营杠杆 | 综合杠杆 | 从业人数 |
| Min.: | -0.99974 | -9826 | -1905.82 | -2702.5 | -231030 | -217549 | 0 |
| 1st\_Qu: | -0.06055 | -0.04 | -0.01 | 1.001 | 7.5 | 7.6 | 263.2 |
| Median: | 0.0828 | 0.04 | 0.44 | 1.071 | 14.09 | 16.5 | 399 |
| Mean: | 0.12206 | 0.89 | 3.51 | 2.535 | 33.02 | 108.7 | 587 |
| 3rd\_Qu: | 0.24602 | 0.21 | 1.61 | 1.421 | 30.3 | 44.8 | 563 |
| Max.: | 24.75429 | 42423 | 44333.5 | 7330.5 | 194301 | 473511.7 | 147130 |

### 数据预处理：

对数据缺失值进行删除，对部分空值以均值中位数填补；

由于处理的28个变量存在离群值，观察变量分布情况：

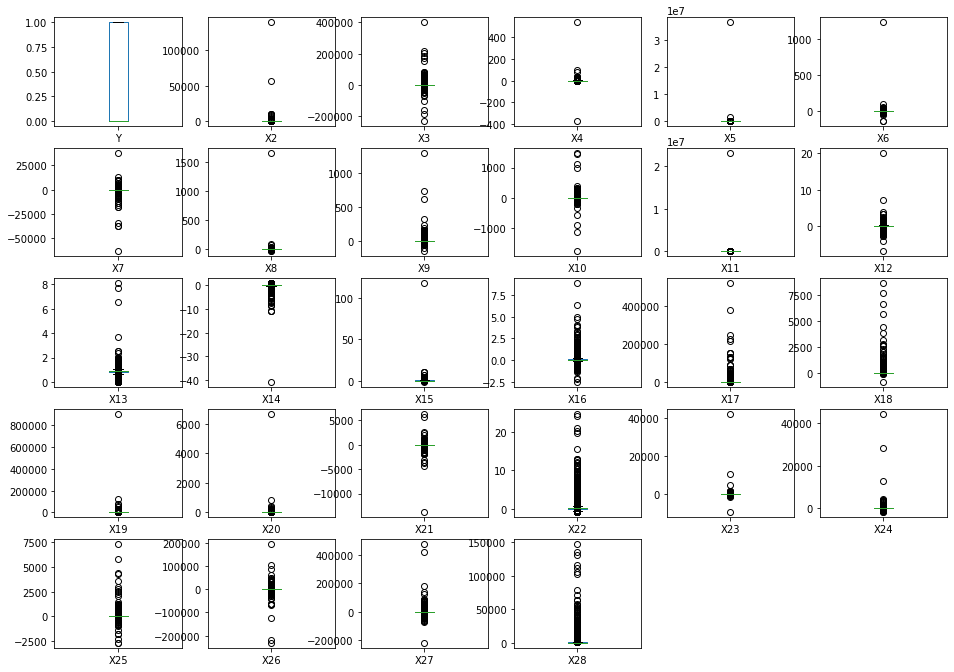


图 5 财务数据分布特征

将离群值以中位数替换：

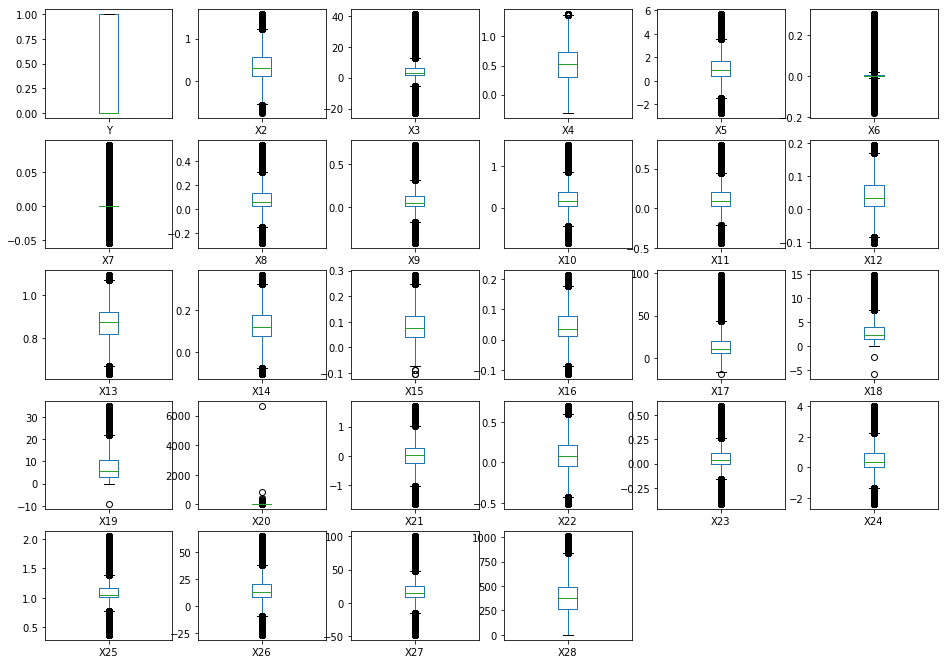


图 6 数据处理后的财务指标数据特恒

# 模型建立

建立的三个模型按照正确率，精准率，召回率，f1值以及ROC曲线，P-R曲线，AUC值进行评估。

Accuracy=TP+TF/(TP+TN+FP+FN) （1）

precision = TP/(TP + FP) （2）

recall = TP/(TP +FＮ) （3）

## 逻辑回归（logisticregression）预测

### 模型原理

二项逻辑回归模型是一种分类模型，主要运用于二分类问题，由条件概率概率P(Y|X)表示，形式为参数化的逻辑分布。这里，随机变量X取值为实数，随机变量Y取值为1或0，我们通过监督学习的方法来估计参数。因此，为了实现Logistic回归分类器，我们在每个特征上都乘以一个回归系数，然后把所有结果值相加，将这个总和带入到sigmoid函数中，进而得到一个范围在0~1之间的数值。任何大于0.5的数据被分入1类，小于0.5的被分为0类。

Sigmoid函数

 （4）

将Sigmoid函数的输入记为z，由以下公式得出

 （5）

当采用向量的写法时，即可转换为z=wTx，向量w也就是我们要找的最佳参数，为了得到更精确的分类效果，需要确定最佳参数，本文选择用极大似然参数法进行参数的估计。

### 结果分析

根据上市企业财务数据训练集与测试集进行结果分析：

确定最佳

得到利用logistic回归分类的正确率为0.7528232010972606，效果良好；

进一步计算其混淆矩阵：结果如下

表 14 混淆矩阵

|  |  |  |  |
| --- | --- | --- | --- |
|  | | 真实情况 | |
| 正常企业 | 财务困境企业 |
| 预测结果 | 正常企业 | TN 49293 | FN 7180 |
| 财务困境企业 | FP 16505, | TP 14932 |

计算其精确率，召回率，f1得分为：

表 15 精确率，召回率，f1得分

|  |  |
| --- | --- |
| 项目 | 得分 |
| precision\_score | 0.6752894356005789 |
| recall\_score | 0.4749817094506473 |
| f1\_score | 0.5576948215652953 |

但数据精确率，召回率，F1值较低，分类效果不理想；

计算P-R图：

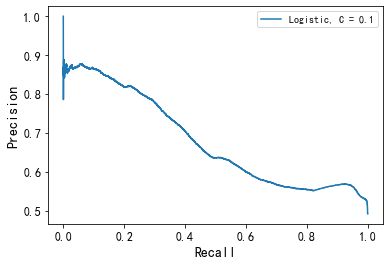


图 7 P-R图

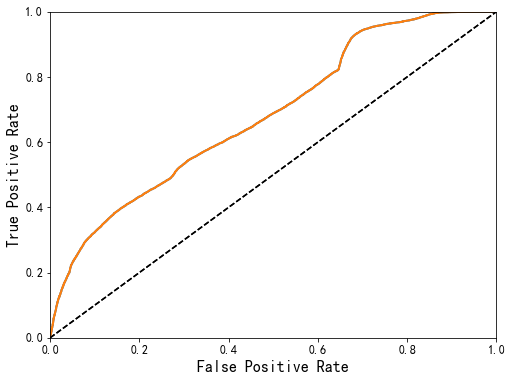
计算ROC图：

图 8 ROC图

可知ROC面积稍大于50%，预测效果不是很理想。

## 随机森林预测模型

### 模型原理

随机森林是一种比较新的数据挖掘模型。经典的数据挖掘模型是神经网络，有半个多世纪的历史了。神经网络预测精确，但是计算量很大。上世纪八十年代Breiman等人发明分类树的算法（Breiman et al. 1984），通过反复二分数据进行分类或回归，计算量大大降低。2001年Breiman把分类树组合成随机森林（Breiman 2001a），即在变量（列）的使用和数据（行）的使用上进行随机化，生成很多分类树，再汇总分类树的结果。随机森林在运算量没有显著提高的前提下提高了预测精度。随机森林对多元共线性不敏感，结果对缺失数据和非平衡的数据比较稳健，可以很好地预测多达几千个解释变量的作用（Breiman 2001b），被誉为当前最好的算法之一（Iverson et al. 2008）。

随机森林顾名思义，是用随机的方式建立一个森林，森林里面有很多的决策树组成，随机森林的每一棵决策树之间是没有关联的。在得到森林之后，当有一个新的输入样本进入的时候，就让森林中的每一棵决策树分别进行一下判断，看看这个样本应该属于哪一类（对于分类算法），然后看看哪一类被选择最多，就预测这个样本为那一类。

### 随机森林结果分析

训练数据集特征： (70328, 27) 测试数据集特征： (17582, 27)

训练数据集标签： (70328,) 测试数据集标签： (17582,)

将随机森林的决策树个数参数设置为100，并计算随机森林R2值

随机森林建模R2值

表 16 随机森林建模R2值

|  |  |
| --- | --- |
| 随机森林数据集 | R2 |
| 训练集得分 | 0.9499904565297775 |
| 测试集得分 | 0.6373759450995629 |

随机森林的预测效果很好，在训练集与测试集上都有着不错的表现。

将随机森林的预测结果可视化：

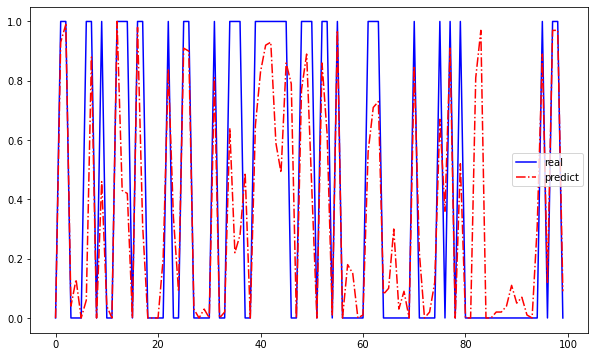


图 9 随机森林预测结果可视化

可见随机森林的预测效果很好，大部分数据都预测较为准确。

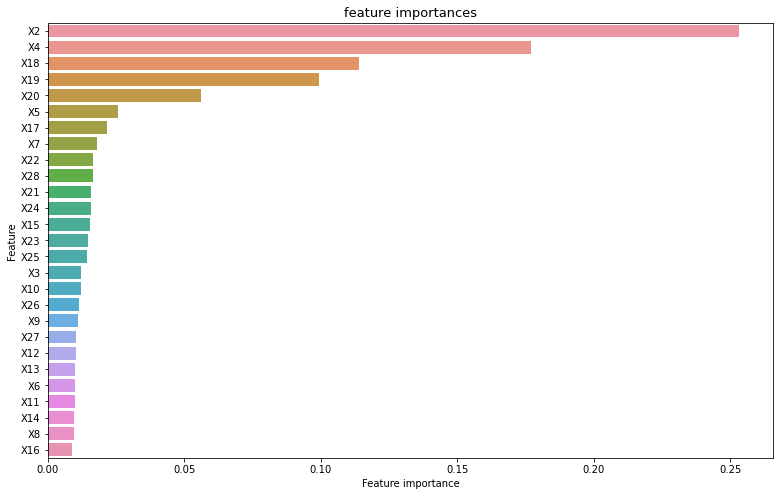
最后，得出自变量的权重分布为：

图 10 财务变量影响权重

可见X2速动比率,X4资产负债率,X18流动资产周转率,X19固定资产周转率,X20总资产周转率,X5产权比率对于企业财务困境有着决定性作用，权重分别为：

表 17 财务变量影响权重

|  |  |
| --- | --- |
| X2速动比率 | 0.2531355 |
| X4资产负债率 | 0.17704872 |
| X18流动资产周转率 | 0.11392285 |
| X19固定资产周转率 | 0.09922678 |
| X20总资产周转率 | 0.05623813 |
| X5产权比率 | 0.02564837 |
| X17存货周转次数 | 0.02179325 |
| X7长期债务营运比率 | 0.01814788 |
| X22营收增长率 | 0.01666904 |
| X28从业人数 | 0.01662664 |

## 多层神经网络预测模型

深度学习是数据挖掘研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据关系和变动规律。

神经网络由以输入层（inputlayer），隐藏层（Hiddenlayer），输出层（outputlayer）组成, 神经网络每层由单元(units)组成,根据不同的层和单元结构，神经网络可以用来做分类(classification）问题或回归(regression)问题。

多层神经网络模型被称为深度学习模型，因具有强大解释能力，在医药、安全、银行、金融、政府、农业和国防等领域有着广泛的应用。

### 密集连接网络模型与指标选取

目标变量是：速动比率，本案例借鉴了吕长江在2003年提出的观点最具操作性和可测性。根据我国大多数上市公司流动负债多,并长年积累转化为长期负债,从而导致偿债危机的实际情况,他提出用流动比率来定义财务困境。他认为财务困境是指企业陷入流动资产小于流动负债即流动比率小于1的境地。因此，本案例将速动比率小于1的企业标记为财务风险，将速动比率大于1的企业标记为财务正常。

自变量选取2013年财务数据参数，包括 从5个维度进行划分：所选取的指标主要由偿债能力、盈利能力、经营能力、发展能力、现金流、风险能力、非财务指标。

表 18 指标选取

|  |  |
| --- | --- |
| 维度 | 指标 |
| 偿债能力指标 | X1流动比率\_2013 |
| X2速动比率\_2013 |
| X3利息保障倍数-2013 |
| X4资产负债率\_2013 |
| X5产权比率\_2013 |
| X6长期资本负债率-2013 |
| X7长期债务营运比率-2013 |
| 盈利指标 | X8资产报酬率\_2013 |
| X9流动资产净利润率2013 |
| X10固定资产利润率2013 |
| X11净资产收益率\_2013 |
| X12营业利润率-2013 |
| X13总营业成本率2013 |
| X14毛利率2013 |
| X15销售期间费用率2013 |
| X16成本费用利润率\_2013 |
| 运营指标 | X17存货周转次数\_2013 |
| X18流动资产周转率（次）\_2013 |
| X29固定资产周转率\_2013 |
| X22总资产周转率（次）\_2013 |
| 发展能力指标 | X21总利润增长率2013 |
|  | X22营收增长率2013 |
|  | X23所有者权益增长率2013 |
|  | X24资本累计率 |
| 风险水平 | X25财务杠杆 |
| X26经营杠杆 |
| X27综合杠杆 |
| 非财务指标 | X28从业人数\_2013 |

正如在前面章节了解到的，一个神经网络包括以下几个部分：

* 层：构成神经网络的基本元素
* 输入数据和对应标签
* 损失函数：定义了作用于学习的反馈函数
* 优化函数：定义了学习的过程

密集连接网络是Dense层的堆叠，用于处理向量数据（向量批量）。这种网络之所以被叫作密集连接，是因为Dense层的每个单元都和其他所有单元相连接,确保了网络中层与层之间最大程度的信息传输。

密集连接网络有四个层组成，其中第一层有20个节点，使用relu激活函数；

设置网络类型为堆栈式神经网络，包含10个神经元的密集层，激活函数为 "relu"，修正线性单元(Rectified linear unit，ReLU）

 （6）

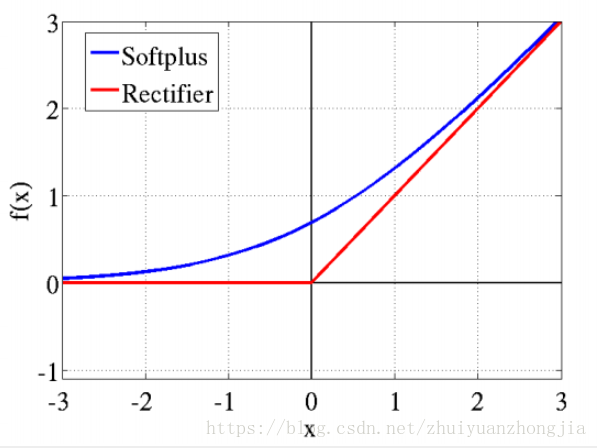


图 11 relu激活函数

第二层有15个节点，包含20个神经元的密集层，激活函数为 "relu"

第三层有7个节点，包含10个神经元的密集层，激活函数为 "tanh"

 （7）

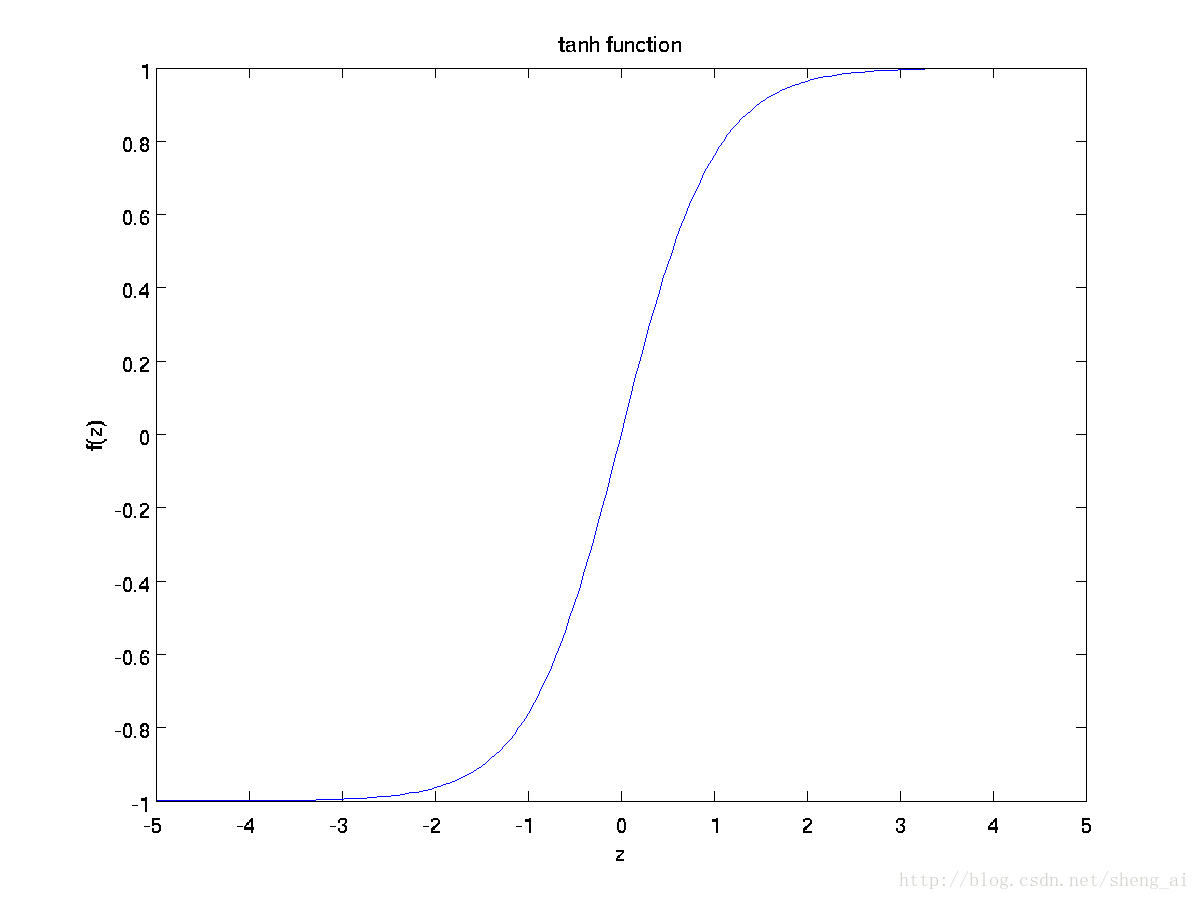


图 12 tanh激活函数

第四层有1个节点，包含1个神经元的密集层，激活函数为 "relu"，该层输出预测结果。

模型选用rmsprop作为优化函数，使用mse为损失函数，使用mae为监测函数。RMSProp算法的全称叫 Root Mean Square Prop，RMSProp算法对权重W和偏置b的梯度使用了微分平方加权平均数。其中，假设在第t轮迭代过程中，各个公式如下所示：

 （8）

 （9）

 （10）

 （11）

在上面的公式中sdw和sdb分别是损失函数在前t-1轮迭代过程中累积的梯度梯度动量，β是梯度累积的一个指数。所不同的是，RMSProp算法对梯度计算了微分平方加权平均数。

### 密集连接网络模型结果分析

在训练集上拟合四层神经网络的参数，其中80%用于计算参数，20%用于评估结果，batch\_size 取值为64，做50次计算。

第三层参数取值热点图

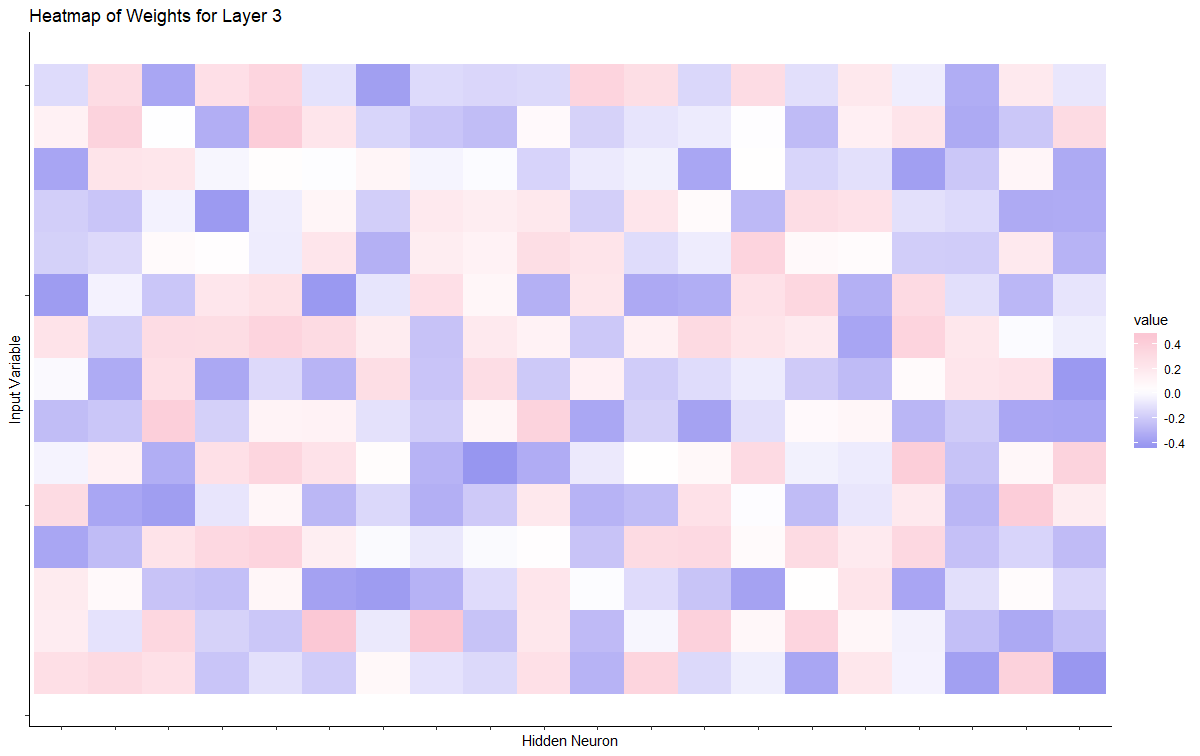


图 13 第三层参数取值热点图

下面展示了密集连接网络模型在财务识别领域的拟合效果（运行50个周期）。

在训练集上

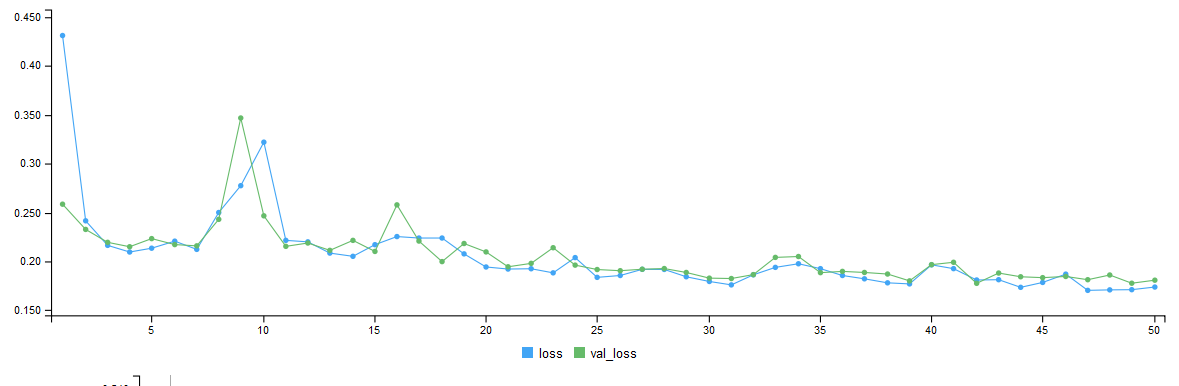


图 14 密集连接网络模型拟合效果

在评估数据集上在20 个周期后，准确度稳定在80%附近

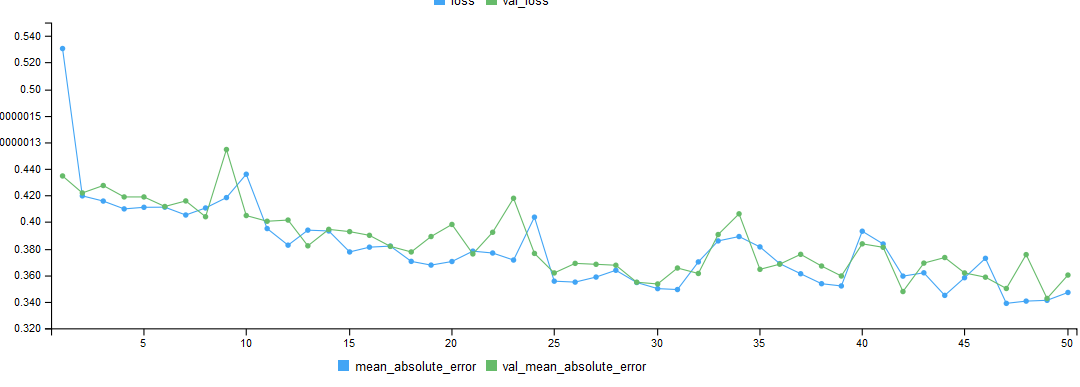


图 15 密集连接网络模型拟合效果

在测试集上在30个周期后，准确度稳定在70%附近。

测试集数据共有14965条数据。

其中5054条正向数据（速动比率小于1），9911条负向数据（速动比率大于1），模型较好的预测了存在财务风险企业和无财务风险企业，识别为0，事实也为0的企业占比为72.48%,识别为1，事实也为1的企业占比为75.32%。准确度稳定在70%附近。

表 19 密集连接网络模型混淆矩阵

|  |  |  |  |
| --- | --- | --- | --- |
|  | | 真实情况 | |
| 财务困境企业 | 正常企业 |
| 预测结果 | 财务困境企业 | TN 2062 | FN 2992 |
| 正常企业 | FP 783 | TP 9128 |

**Abstract**

Financial robustness is a necessary condition for the sustainable management of enterprises. This case uses 87 dimensions of financial data of 360 thousand industrial enterprises in China for statistical analysis to realize the goal of mining implied information from the data. The latent law and value of data are presented through data visualization. By processing missing values and outliers, the influence of interference information on modeling is eliminated, and the unified analysis of general and special cases is realized. According to the analysis objective, the index system is constructed, and the laws found in descriptive statistics are described quantitatively with the classical statistical analysis method. At the same time, due to the large amount of data and comprehensive collection of financial variables, the deep learning method is timely introduced into the analysis. In addition, the estimation effect of classical method and deep learning method is compared in this case. In general, the case data provides a good data basis for statistical methods and practical application analysis, that is, the standard analysis process can be used to explore general rules, but also meet the multi-angle, personalized data mining conditions, meet the needs of statistics teaching cases

**Key words: Industrial Enterprise; Financial Analysis Statistics ;Learning Inference Statistics ;Description Statistics；**

参考文献

1. 别秀梅,刘识非.财务综合分析模型:杜邦分析法[J].金融刊,1997(05):58-59.
2. 侯晋萍.哈佛框架下的财务报表分析:以A公司为例[J].经济师,2012(01):176-177.
3. 李先珍.A公司财务风险分析与预警研究[J].广西质量监督导报,2021(04):133-134.
4. 胡延杰,夏国平.BP神经网络财务危机预警实证研究[J].北京航空航天大学学报(社会科学版),2009,22(04):18-21.
5. 李扬,李竟翔,马双鸽.不平衡数据的企业财务预警模型研究[J].数理统计与管理,2016,35(05):893-906.
6. 徐伟,陈丹萍.财务风险预警建模原则及几种预警新模型[J].统计与决策,2016(08):150-153.
7. 张春梅,赵明清,官俊琪.基于Lasso+SVM的制造业上市公司财务风险组合预警模型[J].数学的实践与认识,2021,51(05):1-12.
8. 李星辰,王青龙,林国庆.基于机器学习方法的上市公司财务预警模型对比研究[J].商场现代化,2020(07):150-152.
9. 吴春雷,马林梅.上市公司财务预警模型构建与变量选取方法[J].中国高新技术企业,2007(06):19+21.
10. 姜玉婵,蔡岩.深度学习模型在医院财务预测中的应用[J].微型电脑应用,2021,37(04):49-52.
11. 吴世农,黄世忠.企业破产的分析指标和预测模型[J].中国经济问题,1987(06):8-15.
12. 程寅骁.基于因子分析法的城投公司财务预警体系研究[J].全国流通经济,2021(03):60-62.
13. 李嘉东. 基于过采样随机森林的工业上市企业财务危机预警模型理论与实证研究[D].山东大学,2020.
14. 陈静,余建波,李艳冰.基于随机森林的用户流失预警研究[J].精密制造与自动化,2021(02):21-24+51.
15. 解秀玉,管西三.企业财务风险预警模型研究——基于制造业数据[J].南京审计学院学报,2013,10(04):58-68.
16. 程寅骁.BP神经网络在城投公司财务预警的应用探讨[J].现代商业,2021(22):163-165.
17. 张春梅,赵明清,官俊琪.基于Lasso+SVM的制造业上市公司财务风险组合预警模型[J].数学的实践与认识,2021,51(05):1-12.
18. 杨保安,季海,徐晶,温金祥.BP神经网络在企业财务危机预警之应用[J].预测,2001(02):49-54+68.
19. 杨淑娥,黄礼.基于BP神经网络的上市公司财务预警模型[J].系统工程理论与实践,2005(01):12-18+26.
20. 梁琪,郝项超.最终控制人所有权和控制权对企业财务失败预警的影响——配对方法与嵌套模型的应用[J].金融研究,2009(01):107-121.
21. 秦小丽,田高良.基于灰色理论和神经网络的公司财务预警模型[J].统计与决策,2011(16):176-178.
22. 郭毅夫,权思勇.基于神经网络的创新型企业财务危机预警研究[J].统计与决策,2013(04):170-172.
23. 龚小凤.基于BP神经网络的企业财务危机预警变量改进探索[J].财经问题研究,2012(12):111-116.
24. 曹兴. 基于深度学习网络的企业财务风险预警研究[D].中国地质大学(北京),2016.
25. 马路静. 基于深度学习的房地产企业财务风险预警研究[D].山东科技大学,2018.
26. 李星辰,王青龙,林国庆.基于机器学习方法的上市公司财务预警模型对比研究[J].商场现代化,2020(07):150-152.
27. 姜玉婵,蔡岩.深度学习模型在医院财务预测中的应用[J].微型电脑应用,2021,37(04):49-52.
28. 圣文顺,赵翰驰,孙艳文.基于优化BP神经网络的销售预测模型研究[J].辽宁工业大学学报(自然科学版),2020,40(02):117-120.
29. 王艾娟.基于EM-BP神经网络的财务绩效评价模型的构建与检验[J].技术与市场,2021,28(03):13-16.
30. 曾驰. 基于财务分析的上市公司分类的改进K-Means与随机森林方法研究[D].湘潭大学,2020.