

Catching the Signal of False Rating—The Text Analysis of Credit Rating Reports

by

Tao Zhang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Business Administration

Approved March 2021 by the
Graduate Supervisory Committee:

Xiaochuan Huang, Co-Chair
Feng Li, Co-Chair
Huibing Zhang

ARIZONA STATE UNIVERSITY

May 2021

捕捉信用评级虚高的信号--债券信用评级报告的文本分析

张韬

全球金融工商管理博士
学位论文

研究生管理委员会
于二零二一年三月批准：

黄晓川，联席主席
李峰，联席主席
张慧冰

亚利桑那州立大学

二零二一年五月

ABSTRACT

Since the year of 2014 when the first bond default happened, the number and amount of bond default have increased a lot year by year, in which there are many bond issuers with the credit rating of A, AA or even AAA. Investors have gotten the real feeling of the bond market risk, so they have to take the bond credit rating much more seriously. The credit rating is closely related with the bond issuance, the market price and the investment decision, but it has been for the long time that the bond rating level has generally higher by domestic rating institutions than in the overseas market. In terms of that, it is worthy of investors to pay more attention on identifying the individual bond with extra high rating so called the false rating. The credit rating report, the key document supporting the rating result, is the important information resource and investment reference for investors. Given that the current reports have been templated extremely, it is not possible for investors to determine whether the rating is false by only relying on the reports.

This article focuses on the text analysis of credit rating reports with the use of tools of natural language and machine learning. The PKU-SEG is used for Chinese word segmentation and tagging, TextRCNN for text categorization, both the emotion dictionary and a machine learning & predicting model for text sentiment analysis. The research sample consists of 400+ bond issuers and almost 800 rating reports. Among reports with

different rating, the text difference is founded by tools and testified with the issuer samples of the bond default and the rating decrease. The research result shows that the text analysis of credit reports is helpful to catch the signal of false rating.

摘要

自 2014 年首家债券违约以来，违约金额和主体数逐年大量增加，其中不乏拥有 A、AA 甚至 AAA 评级的发债主体。投资者切实感受到债券市场的风险，因而不得不谨慎看待债券的信用评级。信用评级影响着债券的发行、市场价格以及投资者决策，然而一直以来，国内评级机构给予企业债券的评级普遍高于海外市场的相应评级；在此背景下，投资者更加关注如何识别出评级愈发虚高的债券个体。信用评级报告是支持评级结果的核心文件，也是资者了解债券的重要信息来源和进行投资决策的参考。然而，信评报告已然非常模板化，投资者很难通过报告本身来判断评级结果是否虚高。

本文采用自然语言和机器学习工具分析信评报告的文本，其中应用 PKUSeg 进行中文分词和标注、TextRCNN 进行文本分类分析、基于情感词典和机器学习预测模型共同进行文本的情感分析。研究样本涉及 400 多家债券发行主体和近 800 份信评报告，本文研究发现不同评级机构和不同档次的评级报告之间存在文本和情感因素的差异，通过抽取发生主体违约和评级下调的信评报告对差异加以验证。研究结果表明信评报告的文本分析有助于捕捉评级虚高的信号。

目录

	页码
表格列表.....	vi
章节	
一、导论及研究方向介绍	1
1.1 导论.....	1
1.2 研究方向.....	5
1.3 研究对象、方法和实践意义.....	9
二、研究样本—信评报告.....	13
2.1 研究样本选择.....	13
2.2 研究样本的整理.....	14
2.3 信评报告的中文分词和词性标注.....	18
三、信评报告的分类差异	22
3.1 信评报告的文本统计差异	22
3.2 信评报告的文本分类差异.....	25
四、信评报告的文本情感分析.....	28
4.1 情感分析应用方法说明	28
4.2 构建信评报告的情感指标.....	28
五、发现评级虚高信号的实验与结论.....	35

章节	页码
5.1 归纳实验样本数据.....	33
5.2 构建信评报告的情感指标计算模型	33
5.3 实验结果说明	35
5.4 结论	36
参考文献.....	37

表格列表

表格	页码
1 债券存量（截至 2020 年 12 月 31 日）	1
2 企业债券历年发行规模	3
3 企业债券历年违约规模	4
4 信评报告的基本内容	10
5 民企和国企债券的违约率	14
6 发行主体个数	15
7 发行主体列表-大公国际	15
8 违约主体列表-大公国际	17
9 评级下调主体列表-大公国际	17
10 样本数据库结构	18
11 词性标注对照表	20
12 词频统计列表-富宇化工	22
13 情感分数表-示例	31
14 情感分析分类测试结果	34
15 样本数据归纳结果	35
16 情感指标映射表	36
17 情感值区间与权重	36

一、导论及研究方向介绍

1.1 导论

中国债券市场的发展始于改革开放后的二十世纪九十年代，以 1990 年上海和深圳证券交易所的设立和 1997 年全国银行间债券市场的成立为重要标志，债券由银行柜台销售转向通过证券交易所或银行间债券市场公开发行业和交易。如表 1 所示，截至 2020 年 12 月 31 日，全市场债券存量 4 万余只，存量余额超过 100 万亿，包括国债、地方政府债、金融债、企业债、公司债、资产支持证券（ABS）等十余个大类品种，其中未包括央行票据和同业存单。

表 1 债券存量（截至 2020 年 12 月 31 日）

债券类别	债券数量 (只)	债券数量比重 (%)	债券余额(亿 元)	余额比重 (%)
国债	271	0.65%	206,859	20.04%
地方政府债	6,230	14.99%	254,864	24.70%
金融债	2,205	5.30%	270,698	26.23%
政策银行债	288	0.69%	183,110	17.74%
商业银行债	325	0.78%	18,144	1.76%
商业银行次级债券	539	1.30%	40,582	3.93%
保险公司债	64	0.15%	3,036	0.29%
证券公司债	738	1.78%	17,548	1.70%
证券公司短期融资券	55	0.13%	1,429	0.14%
其它金融机构债	196	0.47%	6,851	0.66%
企业债	2,602	6.26%	22,642	2.19%
一般企业债	2,597	6.25%	22,606	2.19%
集合企业债	5	0.01%	36	0.00%

债券类别	债券数量 (只)	债券数量比重 (%)	债券余额(亿 元)	余额比重 (%)
公司债	8,975	21.59%	89,490	8.67%
私募债	5,217	12.55%	42,902	4.16%
一般公司债	3,758	9.04%	46,589	4.51%
中期票据	6,262	15.06%	74,784	7.25%
一般中期票据	6,262	15.06%	74,784	7.25%
短期融资券	2,292	5.51%	20,929	2.03%
一般短期融资券	503	1.21%	4,800	0.47%
超短期融资债券	1,789	4.30%	16,130	1.56%
定向工具	2,986	7.18%	21,521	2.09%
国际机构债	16	0.04%	320	0.03%
政府支持机构债	166	0.40%	17,225	1.67%
资产支持证券	9,050	21.77%	45,566	4.42%
银保监会主管 ABS	1,155	2.78%	15,921	1.54%
证监会主管 ABS	6,329	15.22%	22,583	2.19%
交易商协会 ABN	1,566	3.77%	7,062	0.68%
可转债	396	0.95%	5,326	0.52%
标准化票据	26	0.06%	29	0.00%
可交换债	96	0.23%	1,730	0.17%
合计	41,573	100%	1,031,983	100%

数据来源：WIND 资讯

存量债券中，超过 70%的债券余额属于政府信用背书和金融机构发行的债券，企业债（2602 只）与公司债（8975 只）合计数量虽占存量的近 30%，但二者的存量余额略高于 11 万亿，不足总余额的 11%。（注：以下本文所称“企业债”是对企业债和公司债的统称，对应数据是二者的合计。）

企业债券市场始于 1998 年，如表 2 数据所示，当年发行量不足百亿，2007 年年度发行量首次超过 1 千亿元，2015 年年度发行超过 1 万亿，2020 年接近 4 万亿。

表 2 企业债券历年发行规模

统计年份	发行量（亿元）	年度增量（亿元）	增长率
1998	98	n/a	n/a
1999	128	30	31%
2000	85	-43	-33%
2001	164	79	92%
2002	360	196	120%
2003	336	-24	-7%
2004	272	-64	-19%
2005	604	332	122%
2006	615	11	2%
2007	1,234	619	101%
2008	1,855	621	50%
2009	3,992	2,137	115%
2010	3,344	-648	-16%
2011	3,791	447	13%
2012	9,137	5,347	141%
2013	6,455	-2,682	-29%
2014	8,443	1,988	31%
2015	13,720	5,277	62%
2016	33,835	20,116	147%
2017	14,756	-19,080	-56%
2018	18,994	4,238	29%
2019	29,063	10,069	53%
2020	37,624	8,561	29%

数据来源：WIND 资讯

随着企业债券市场的发展，债券违约已是不期而至，2014年“11超日债”违约是企业债的首次违约。如表3数据所示，企业债历年违约金额从2014年的13亿快速增长至2020年的1334亿，占当年存量金额的比例从万分之四发展到2019年的1.25%和2020年的1.19%。

表3 企业债券历年违约规模

年份	企业债违约金额 (亿元)	企业债当年存量金 额(亿元)	违约金额占比
2014	13	36,917	0.04%
2015	73	47,369	0.15%
2016	24	75,880	0.03%
2017	114	81,409	0.14%
2018	735	83,943	0.88%
2019	1,164	92,837	1.25%
2020	1,334	112,132	1.19%

数据来源：WIND 资讯

面对企业债高于1%的违约率，和不足5%的市场平均到期收益率，即使充分分散的企业债投资组合也要面对可能高达20%的收益减值。另外，考虑到企业债流动性差，组合很难充分分散，收益减值风险会更高。当然，市场会要求更高的发行利率作为风险补偿，在违约率与收益率之间取得平衡；但对于投资者而言，如何加强对拟投债券的风险识别与跟踪，尽可能规避违约风险，才是当务之急。

1.2 研究方向

1.2.1 债券信用评级

债券信用评级（bond credit rating）或简称为“债券信评”是以具有独立法人资格的机构主体所发行某一特定有价债券为对象进行的信用评级，对其按期还本付息的可靠程度进行评估，并标示其信用等级的等级，债券信用评级行业与债券市场的发展相生相伴。鉴于有政府直接或间接的兑付担保，在国内发行的国债和国有银行的金融债券，不需要进行信用评级；因此，债券信评通常是指企业债券的信用评级，本文即以企业债券的信用评级为研究方向。

债券信评是发行利率的重要参考，可以降低高信誉企业的融资成本。一般来说，信评等级越高的债券，越容易得到投资者的信任，能够以较低的利率出售；而信评等级低的债券，风险较大，只能以较高的利率发行。投资者购买债券要承担信用风险，风险度因发债主体的偿付能力不同而有所差异。对于投资者而言，投前了解债券的信用风险非常重要，但受知识、信息、效率等多方面因素的限制，无法对众多债券进行深入分析。因此，需要专业机构对准备发行的债券进行债券信评，即对其还本付息的可靠程度进行客观、公正和权威的评定，从而方便投资者进行债券投资决策。

信用评级的行业管理主要根据 2015 年 1 月 15 日正式公布实施《公司债券发行与交易管理办法》（证监会令第 113 号，以下简称“《管理办法》”）中关于评级的规定，如“公开发行公司债券，应当委托具有从事证券服务业务资格的资信评级机构进行信用评级，评级

机构按照《管理办法》第四十六条规定履行定期跟踪评级和重大事项评级调整的职责。”；
“非公开发行公司债券是否进行信用评级由发行人确定，并在债券募集说明书中披露。”

《管理办法》中“具有从事证券服务业务资格的资信评级机构”的“业务资格”包括四项，分别是：

- 1) 中国人民银行认定的全国性企业债券及银行间债券信用评级机构；
- 2) 中国证监会认定的从事证券市场资信评级业务信用评级机构；
- 3) 国家发改委认定的企业债券、中小企业、担保公司信用评级机构；
- 4) 中国保监会认定的对保险公司投资债券进行信用评级的机构。

中国第一家从事信用评级的专业机构是中国诚信证券评估有限公司，现改称为中诚信国际信用评级有限责任公司（简称“中诚信国际”），始创于 1992 年 10 月，经中国人民银行总行批准成立。目前符合《管理办法》中“具有从事证券服务业务资格的资信评级机构”的“资信评级机构”包括以下九家，分别是：

- 1) 中诚信国际信用评级有限责任公司（简称“中诚信国际”）；
- 2) 中诚信证券评估有限公司（简称“中诚信证评”）；
- 3) 联合资信评估有限公司（简称“联合资信”）；
- 4) 联合信用评级有限公司（简称“联合评级”）；
- 5) 大公国际资信评估有限公司（简称“大公国际”）；
- 6) 上海新世纪资信评估投资服务有限公司（简称“上海新世纪”）；
- 7) 东方金诚国际信用评估有限公司（简称“东方金诚”）；

8) 中证鹏元资信评估股份有限公司（简称“中证鹏元”）；

9) 远东资信评估有限公司（简称“远东资信”）

行业管理的相关的规定和明确的资质要求奠定了债券评级行业的基本格局，即由获得业务资格的九家评级公司对企业债券的进行发行评级和后续的跟踪评级。

1.2.2 关于债券信评虚高的猜想

A 级评级是债券信评的最高等级，通常再从高到低进行内部分档，如 AAA、AA+、AA、A+、A 等。获得 A 级评级的债券发行主体偿债能力强，企业经营受经济形势影响的程度较小；债券本金和收益的安全性最大，因而它们的收益水平较低，发债企业的筹资成本也低。中国债券市场的主要投资资金来自于各类机构，这些机构大都接受外部约束或称监管约束，即监管部门设定的投资规则，比如中国证券监督管理委员会要求货币基金不得投资“信用等级在 AA+ 以下的债券与非金融企业债务融资工具”，原保险监督管理委员会要求保险资金投资“无担保非金融企业（公司）债券，具有国内信用评级机构评定的 AA 级或者相当于 AA 级以上的长期信用级别。”（《保险资金投资债券暂行办法》（保监发〔2012〕58 号））。

国内发行的企业债多数获得了 A 级评级，这与机构投资者所受的监管约束有关。评级公司出具的评级结果是机构投资的基本前提，进而影响债券的定价和发行，唯有评级为 A 级及以上的企业才有可能成功进行市场化发行。国内企业在海外的发债则鲜有 A 级，多是 B 级。B 级债券的安全性、稳定性以及利息收益会受到经济中不稳定因素的影响，经济形

势的变化对这类债券的价值影响很大。B级债券的收益水平较A级高，发行主体的筹资成本也较高，而投资者面对的投资风险也高。

与海外市场的评级相比，国内市场给予企业的债券评级整体偏高。尽管如此，债券市场对于投资者也算是公平的，毕竟大家共同面对一个整体评级偏高的市场进行决策。在评级整体偏高的前提下，投资者更加关注如何识别出评级虚高的个体。以下是导致个别债券评级虚高的两个猜想。

猜想之一是评级机构或评级者的主观判断造成的个体虚高。评级公司所应用的信用评级方法是以发行主体的财务和经营数据为基础，从历史数据中提炼信息；在分析方法上，除定量分析外，还使用定性分析，例如：专家分析法、OOC评级法、信用得分法等。评级方法中的定性部分和对未来的预测都受到评级相关人员主观偏好和判断的影响，从而造成个体评级虚高。

猜想之二是评级业务竞争压力下的刻意虚高。某家企业拟发行债券并委托评级机构进行发行评级，评级机构要秉承“公平、公证、公开”的三公原则进行评级，同时又要向发债主体收费，与发债主体之间存在一定的利益关系。在坚持原则与经济利益之间难免存在矛盾，当市场上有几家评级机构竞争同一评级业务时，可能会满足发债主体的高评级要求，从而导致个体评级虚高，即所谓的“竞争评级”。

对于个别债券评级虚高的原因，很难有一个量化或是确定的解释；对投资者而言，大多数情况下也无需探究导致个体评级虚高的具体原因。投资者真正关心的是是否能够识别

出评级虚高的个体或者发现可能虚高的信号，从而提升投资决策过程中的风险控制，避免不合理的价格乃至可能的违约风险。

本文的研究方向不是债券信用评级的方法，也不是验证关于评级虚高的猜想，而是确认债券信用评估报告是否释放出评级虚高的信号以及如何捕捉到这一信号。

1.3 研究对象、方法和实践意义

1.3.1 研究对象

本文的研究对象是债券信用评级报告（本文简称“信评报告”），信评报告是债券信评的核心，用以揭示债券发行主体和所发债券的基本信息，并给予分档评级结果。信评结论最直接的体现是评级结果，可被简化成几个字母的表达，如 AA、BB 等，评级报告通过分析债券发行主体的偿债能力、评价其能否按期偿还本息、说明投资该债券所需承担的风险程度等来解释评级结论。信评报告全面体现评级公司所做的工作及其成果，包括信息收集、实地尽调、行业研究、评级模型的运用等。针对一只债券，评级公司给予的评级结果是满足很多机构进行合规投资的基本条件，评级机构出具的信评报告则是投资者了解该债券的重要信息来源和重要的决策参考。

《公司债券发行与交易管理办法》中明确了债券信用评级报告的基本要求：

1) 按照规定或约定将评级信息告知发行人，并及时向市场公布首次评级报告、定期和不定期跟踪评级报告；

2) 在债券有效存续期间，应当每年至少向市场公布一次定期跟踪评级报告；

3) 应充分关注可能影响评级对象信用等级的所有重大因素，及时向市场公布信用等级调整及其他与评级相关的信息变动情况，并向证券交易所或其他证券交易场所报告。

各家评级机构在满足发布频率等基本要求之外，已逐步将信评报告模板化，大致结构和内容如表 4 所示：

表 4 信评报告的基本内容

序号	栏目	主要内容
第一部分	发债主体概况	对发债主体进行简单介绍
第二部分	本期债券概况	介绍本期债券的发行规模、利率结构等基本要素
第三部分	宏观经济和政策环境	分析可能影响发债主体经营的宏观经济和相关政策
第四部分	行业及区域经济环境	介绍并分析发债主体所属行业和地域的经济
第五部分	基础素质分析	股权状况、企业规模和竞争力、人员素质、信用记录
第六部分	法人治理结构	法人治理结构、管理水平
第七部分	经营分析	基于收入、费用等数据对主要业务进行分析
第八部分	财务分析	基于财务报表的基本信用分析，包括杠杆率、盈利能力、现金流等比率指标
第九部分	本期债券偿还能力分析	对现有债务的影响、本期债务的本息偿还预测
第十部分	结论	评级结果

信评机构撰写信评报告的信息来源来自公开信息和通过对发债主体进行实地调研所获取的资料。在准备发债的过程中，发债主体会配合评级机构对其进行实地调研，并且提供撰写评级报告所需的信息和材料，这是评级机构全面并深入了解发债主体的难得机会，而对于一般投资者而言，基本上很难获得类似的机会。通过实地调研，评级机构能否获得

“非公开信息”，并在评级报告中能否充分应用此类信息以支持评级逻辑，也是衡量评级质量的关键因素。

在评级模型中，发债主体的财务数据、经营数据和行业数据等量化信息可进行横向和纵向的分析，从而奠定评级模型的基础；而评级者的主观态度是否真的与评级结果一致呢？本文的研究对象是信评报告的文本，即从文本挖掘的角度，进行评级报告信息含量和情感因素的区分，比较评级结论与报告文本之间的差异，进而发行凭借虚高的信号。

1.3.2 研究方法

债券信用评级的相关研究大致可分为两类：一类是构建或改进信用评级制度层面的研究，如信评行业制度建设与发展、国内外评级制度比较、评级付费模式的影响、再评级制度设计等论题；另一类是关于信用评级方法层面的研究，如评级模型分析、如何在对某行业（如房地产行业）或某类企业（如中小企业）进行信评时调整参数等。在定性分析之外，两类研究主要应用基于数据的比较、归因、统计等量化方法。

本文不属于以上两类研究，也未应用常规的定性和定量研究方法，而是采用自然语言和机器学习技术对信评报告的文本分类和情感因素分析。随着机器学习领域的发展，已有越来越多的较为成熟的技术工具支持研究者可直接进行对文本信息的研究，本文应用 PKUSeg 进行分词和标注、TextRCNN 进行文本分类、基于情感词典的 BosonNLP 和 LSTM 机器学习预测模型共同进行文本的情感因素分析。

本文是对债券全样本的研究，并未对发债主体按行业或其它标准对发债主体进行区分，而是对文本进行了分组。一是按不同评级机构出具的信评报告分组，已发现在评级机

构之间是否存在文本差异。二是将债券中发生违约和评级下调的债券发行主体归为一组，定义其发行信评报告为不可靠报告组，将该组报告与其它报告的情感因素进行比较，已验证虚高信号的存在。。

1.3.3 研究的实践意义

本文研究是来自投资者的视角，研究方法和成果适用于债券投资的实务操作。在整体评级偏高的市场中，提供捕捉到债券个体评级虚高信号的可能性和方法论。捕捉信号只能是提醒作用或是对违约或下调评级可能性的预测，并不能替代投资者的决策环节。因此，本文研究可作为债券投资的辅助评价工具，为投资决策提供支持。同时为阅读信评报告提供了直接线索。协助投资人员将信评报告中重要信息提取出来，或是提炼出报告的要点，以提高研究效率。

本文研究为辨析各类商业报告和公告进行技术应用的探索。在商业实践中，各种量化和统计工具已广范应用于各种数据的分析，而针对文本分析的机器学习等技术工具还待被广泛市场使用。当然，一些新闻和简报，已经实现人工智能的文本写作乃至创作，我们未来的探索或是该如何区分人与机器间文本的差异，毕竟在当前的认知中，人拥有更加丰富的情感因素。

二、研究样本—信评报告

2.1 研究样本选择

2.1.1 选择“委托评级”样本

评级公司作为金融中介机构受托于某家发债主体对其进行评级（称为“委托评级”），同时向作为评级委托方的发债主体收费；评级行业另有承做“主动评级”业务的评级机构，其收费对象是投资者，也就是说其是为付费的会员提供债券评级信息服务。

本文选择“委托评级”的评级报告作为样本，主要原因是：1）“委托评级”向发债主体收费，产生个体评级虚高的可能性大。2）“主动评级”业务并没有发债主体的配合，因此其评级的信息来源均是来自公开信息，无法进入企业实地调研和获取一手资料，而“委托评级”则可获得发债主体的信息支持。3）“委托评级”的评级结果和评级报告是作发债主体发行债券的重要材料，是信息披露的一环，可比较容易的收集；而“主动评级”的债券评级资料只对会员有限开放，较难获得。

2.1.2 选择评级 AA 及以上的发债主体

本文选择评级 AA 及以上的企业债为研究样本，主要是考虑到这部分债券中可能存在个体评级虚高的现象。AA 评级的债券才能满足多数机构投资者的进行基本要求，才有可能进行成功的债券发行。因此，根据前述“关于债券信评虚高的猜想”，将评级拔高至 AA 级对于发债主体有利益驱动，评级机构也很有可能采用“竞争评级”。

2.1.3 选择民营企业的发债主体

在评级 AA 级以上的企业债券中，本文将研究样本进一步聚焦于民营企业发行的企业债。如图表 5 中数据所示，民营企业的发行主体违约率是 7.38%，而国有企业的发行主体违约率是 0.85%，前者显著高于后者，并构成一定数量的违约样本。违约债券中更有可能存在评级虚高的个体，违约样本则可用以验证虚高信号的存在和有效性。

表 5 民企和国企债券的违约率

	民营企业	AAA	AA+	AA	民企合计	国有企业
企业债券	存量余额（亿元）	8,327	4,688	2,456	15,471	277,756
	债券数量（只）	684	831	605	2,120	28,646
	发行主体（个）	64	149	234	447	4,226
违约企业债券	存量余额（亿元）	9	97	181	287	1,180
	债券数量（只）	1	27	39	67	169
	发行主体（个）	1	10	22	33	36
违约率	存量余额（亿元）	0.10%	2.07%	7.39%	1.86%	0.42%
	债券数量（只）	0.15%	3.25%	6.45%	3.16%	0.59%
	发行主体（个）	1.56%	6.71%	9.40%	7.38%	0.85%

数据来源：WIND 资讯

2.2 研究样本的整理

2.2.1 发行主体

表 6 发行主体个数

评级机构	未违约主体个数	违约主体个数	合计数
大公国际	43	9	52
东方金诚	40	3	43
联合信用	93	6	99
联合资信	59	4	63
上海新世纪	60	3	63
中诚信国际	82	5	87
中证鹏远	33	2	35
合计	410	32	442

市场共有 9 家评级机构，本文选取其中的 7 家机构，因为“远东征信”和“中诚信证评”仅参与了个位数的主体信评，未予选择。市场存量数据，基于 7 家信评机构的 442 家发债主体（其中有 32 家违约主体）的发行信用评级报告。下表是以“大公国际”一家评级公司的评级主体和报告样本。

表 7 发行主体列表-大公国际

序号	债券发行主体	信评报告
1	保集控股集团有限公司	保集控股-大公国际-20200629
2	北京捷成世纪科技股份有限公司	捷成世纪-大公国际-20200629
3	传化智联股份有限公司	传化智联-大公国际-20200629
4	大连万达商业管理集团股份有限公司	万达商业-大公国际-20200628
5	东营市亚通石化有限公司	亚通石化-大公国际-20200630
6	福建福晟集团有限公司	福晟集团-大公国际-20190626
7	福建阳光集团有限公司	阳光集团-大公国际-20200627
8	刚泰集团有限公司	刚泰集团-大公国际-20180925
9	顾家集团有限公司	顾家集团-大公国际-20200714
10	海亮集团有限公司	海亮集团-大公国际-20200619
11	杭州正才控股集团有限公司	正才控股-大公国际-20200629

序号	债券发行主体	信评报告
12	河北海伟交通设施集团有限公司	海伟交通-大公国际-20190729
13	红星美凯龙家居集团股份有限公司	红星美凯龙家居-大公国际- 20200715
14	红星美凯龙控股集团有限公司	红星美凯龙控股-大公国际- 20200630
15	湖南友谊阿波罗控股股份有限公司	友谊阿波罗-大公国际-20200703
16	江苏永钢集团有限公司	永钢集团-大公国际-20200702
17	江西正邦科技股份有限公司	正邦科技-大公国际-20200616
18	金世旗国际控股股份有限公司	金世旗-大公国际-20200629
19	精功集团有限公司	精功集团-大公国际-20190717
20	开元旅业集团有限公司	开元旅业-大公国际-2020611
21	龙光交通集团有限公司	龙光交通-大公国际-20200615
22	能兴控股集团有限公司	能兴控股-大公国际-20200618
23	宁波奥克斯置业有限公司	奥克斯-大公国际-20200623
24	宁夏泰瑞制药股份有限公司	泰瑞制药-大公国际-20190617
25	荣盛房地产发展股份有限公司	荣盛房地产-大公国际-20200427
26	山东富宇化工有限公司	富宇化工-大公国际-20190626
27	山东海科化工有限公司	海科化工-大公国际-20190626
28	山东齐悦科技有限公司	齐悦科技-大公国际-20200629
29	山东三星集团有限公司	三星集团-大公国际-20200629
30	山东胜通集团股份有限公司	胜通集团-大公国际-20190322
31	山东鑫海科技股份有限公司	鑫海科技-大公国际-20191224
32	深圳市海王生物工程股份有限公司	海王生物-大公国际-20200624
33	同益实业集团有限公司	同益实业-大公国际-20181026
34	五洋建设集团股份有限公司	五洋建设-大公国际-20170816
35	西王集团有限公司	西王集团-大公国际-20190725
36	星星集团有限公司	星星集团-大公国际-20200629
37	雪松实业集团有限公司	雪松实业-大公国际-20200702

序号	债券发行主体	信评报告
38	阳光城集团股份有限公司	阳光城集团-大公国际-20200708
39	阳光凯迪新能源集团有限公司	阳光凯迪-中证鹏元-20190701
40	亿利洁能股份有限公司	亿利洁能-大公国际-20200703
41	亿利资源集团有限公司	亿利资源-大公国际-20200629
42	浙江恒逸集团有限公司	恒逸集团-大公国际-20200629
43	中弘控股股份有限公司	中弘控股-大公国际-20181029

表 8 违约主体列表-大公国际

序号	违约主体	信评报告
1	宁夏远高实业集团有限公司	宁夏远高-大公国际-20160408
2	新华联控股有限公司	新华联控-大公国际-20140218
3	北京信威通信技术股份有限公司	北京信威-大公国际-20151218
4	华泰汽车集团有限公司	华泰汽车-大公国际-20160728
5	精功集团有限公司	精功集团-大公国际-20130930
6	东辰控股集团集团有限公司	东辰控股-大公国际-20140128
7	刚泰集团有限公司	刚泰集团-大公国际-20160922
8	新光控股集团集团有限公司	新光控股-大公国际-20170922
9	中融双创(北京)科技集团有限公司（企业曾用名：邹平长城集团有限公司）	邹平长城-大公国际-20160608

表 9 评级下调主体列表-大公国际

序号	评级下调主体
1	亿利资源集团有限公司
2	雪松实业集团有限公司
3	佳源创盛控股集团集团有限公司
4	杭州正才控股集团集团有限公司
5	北京信威通信技术股份有限公司

2.2.2 样本的结构化处理

将各主体的评级报告文件（PDF 版）767 份（部分主体多次发债，对应多份评级报告）读取为文本数据；过滤其中部分信息，如评级公司的评级声明、跟踪评级安排等；进行数据结构化处理，形成如下数据表并建立数据库（图表 10）。进一步结构化，标题识别：信评报告中的各级标题识别，将文本对应到各个标题；文件中表格数据读取及结构化：信评报告中以表格形式呈现的数据，写入数据库

表 10 样本数据库结构

数据库分表	数据项
评级公司表	编号、公司名称
信评报告表	信评报告表：评级企业、评级公司、评级日期、报告正文原文
企业表	编号、名称、发行评级、当前评级、发行信评报告编号、当前信评报告编号、是否违约

2.3 信评报告的中文分词和词性标注

2.3.1 中文分词和标注的方法说明

中文分词是进行文本分类和情感分析基础，根据实现原理和特点，分词方法主要分为两类。第一类是基于词典的分词方法：该类方法按照一定的策略将待匹配的字符串和一个已建立好的“充分大的”词典中的词进行匹配，若找到某个词条，则说明匹配成功，识别该词；常见的基于词典的分词方法包括正向最大匹配、逆向最大匹配、N-最短路径法和双向匹配分词法等。第二类是基于统计的分词方法：该类方法在给定大量已经分词的文本的前提下，利用统计机器学习模型对汉字进行标注训练，从而实现对未知文本的切分，目前常用的算法包括隐马尔科夫模型（HMM）、条件随机场模型（CRF）、最大熵模型

(ME)、支持向量机模型 (SVM) 以及包括 textCNN 和 BERT 在内的深度学习模型。所谓模型，就是指能对输入数据进行处理，并且给出最优的输出。对于字标注的分词方法来说，输入就是 n 个字，输出就是 n 个标签，从而实现对文本的标注。

本文采用的是北京大学语言计算与机器学习研究组研制推出的一套全新的中文分词工具包，简称 PKUSeg，是利用隐马尔科夫加假设和模型开发的分词工具。其中 pkuseg 相比于其他的分词工具具有如下特点：1) 高分词准确率。相比于其他的分词工具包，该工具包在不同领域的数据上都大幅提高了分词的准确度。根据测试结果，pkuseg 分别在示例数据集 (MSRA 和 CTB8) 上降低了 79.33% 和 63.67% 的分词错误率。2) 多领域分词。该分词包训练了多种不同领域的分词模型，根据待分词的领域特点，用户可以自由地选择不同的模型。3) 支持用户自训练模型，即用户可以使用全新的标注数据进行训练。

2.3.2 分词和词性标注示例说明

以信评报告《富宇化工-大公国际-20170331》为例说明对信评报告文本进行中文分词和词性标注说明。

信评报告原始文本摘录：

“公司地处东营市，周边石化企业较多，原材料采购及产品销售渠道较畅通，区位优势较为明显；公司是山东省规模较大的地炼企业，丰富的产品结构可为整体抗风险能力提供保障；芳烃精制项目的完工提升公司炼油产品加工能力；公司期间费用在营业收入中占比较低，期间费用控制能力较强。2016 年以来国内燃料油市场价格持续增长，受此影响，芳烃等石油化工类产品成本有所增长公司总有息债务在总负债中占比很高，以短息有

息债务为主，短期面临一定偿债压力公司担保比率较高，被担保企业区域集中度较高，存在一定或有风险。”

表 11 词性标注对照表

N	名词	C	连词	p	介词
T	时间词	U	助词	w	标点符号
S	处所词	Y	语气词	nr	人名
F	方位词	E	叹词	ns	地名
M	数词	O	拟声词	nt	机构名称
Q	量词	I	成语	nx	外文字符
B	区别词	I	习惯用语	nz	其它专名
R	代词	J	简称	vd	副动词
V	动词	H	前接成分	vn	名动词
A	形容词	K	后接成分	vx	形式动词
Z	状态词	G	语素	ad	副形词
D	副词	X	非语素字	an	名形词

基于《词性标注对照表》，对上述原始文本的摘录实现如下分词与词性标注结果：

[('公司', 'n'), ('地处', 'v'), ('东营市', 'ns'), (' ', 'w'), ('周边', 'n'), ('石化', 'j'), ('企业', 'n'), ('较', 'd'), ('多', 'a'), (' ', 'w'), ('原材料', 'n'), ('采购', 'v'), ('及', 'c'), ('产品', 'n'), ('销售', 'vn'), ('渠道', 'n'), ('较', 'd'), ('畅通', 'a'), (' ', 'w'), ('区位', 'n'), ('优势', 'n'), ('较为', 'd'), ('明显', 'a'), (';', 'w'), ('公司', 'n'), ('是', 'v'), ('山东省', 'ns'), ('规模', 'n'), ('较', 'd'), ('大', 'a'), ('的', 'u'), ('地炼', 'vn'), ('企业', 'n'), (' ', 'w'), ('丰富', 'a'), ('的', 'u'), ('产品', 'n'), ('结构', 'n'), ('可', 'v'), ('为', 'v'), ('整体', 'n'), ('抗', 'v'), ('风险', 'n'), ('能力', 'n'), ('提供', 'v'), ('保障', 'vn'), (';', 'w'), ('芳烃', 'n'), ('精制', 'n'), ('项目', 'n'), ('的', 'u'), ('完工', 'vn'), ('提升', 'vn'), ('公司', 'n'), ('

炼油', 'vn'), ('产品', 'n'), ('加工', 'vn'), ('能力', 'n'), (':', 'w'), ('公司', 'n'), ('期间', 'f'), ('费用', 'n'), ('在', 'p'), ('营业', 'vn'), ('收入', 'n'), ('中', 'f'), ('占', 'v'), ('比较', 'd'), ('低', 'a'), (';', 'w'), ('期间', 'f'), ('费用', 'n'), ('控制', 'vn'), ('能力', 'n'), ('较', 'd'), ('强', 'a'), ('.', 'w'), ('2016 年', 't'), ('以来', 'f'), ('国内', 's'), ('燃料油', 'n'), ('市场', 'n'), ('价格', 'n'), ('持续', 'vd'), ('增长', 'v'), (';', 'w'), ('受', 'v'), ('此', 'r'), ('影响', 'vn'), (';', 'w'), ('芳烃', 'n'), ('等', 'u'), ('石油', 'n'), ('化工', 'n'), ('类', 'n'), ('产品', 'n'), ('成本', 'n'), ('有所', 'v'), ('增长', 'v'), ('公司', 'n'), ('总', 'd'), ('有息', 'v'), ('债务', 'n'), ('在', 'p'), ('总负债', 'n'), ('中', 'f'), ('占比', 'v'), ('很', 'd'), ('高', 'a'), (';', 'w'), ('以', 'p'), ('短息', 'n'), ('有息', 'v'), ('债务', 'n'), ('为主', 'v'), (';', 'w'), ('短期', 'd'), ('面临', 'v'), ('一定', 'b'), ('偿债', 'vn'), ('压力', 'n'), ('公司', 'n'), ('担保', 'vn'), ('比率', 'n'), ('较', 'd'), ('高', 'a'), (';', 'w'), ('被', 'p'), ('担保', 'v'), ('企业', 'n'), ('区域', 'n'), ('集中度', 'n'), ('较', 'd'), ('高', 'a'), (';', 'w'), ('存在', 'v'), ('一定', 'b'), ('或', 'c'), ('有', 'v'), ('风险', 'n'), ('.', 'w)']

本文使用 PKUSeg 对所有的信评报告进行分词和词性标注，准确率达到 85%左右，当然，在此基础上，即完成基本词性标注后，还可加入专有名词和完善停用词表，加入新词发现机制，进而进行细分领域模型训练，实现在信评报告细分领域分词和标注更高的准确率。本文研究是发现文本分类和情感因素在不同信评报告之间的差异，85%分词准确率已是可以作为实现文本分类研究和情感分析的基础。

三、信评报告的分类差异

3.1 信评报告的文本统计差异

本文尝试从文本统计方面对信评报告的差异加以确认，如报告字数、包含数字的句子占比、句子数量、句子长度、词频等。以信评报告《富宇化工-大公国际-20170331.》为例，共有 17708 个字，1045 个句子（按逗号划分），单个句子平均长度约为 17 个字；共有逗号（，或，）1045 个，以逗号为分隔将文本切分为句子，其中包含有数字的句子有 348 个，数字信息占比为 $348/1045 = 33.3\%$ ；词频统计结果（出现超过 20 次的词，去除标点符号及停用词之后）如下：反映出该报告在数字方面的论据支撑强度。

表 12 词频统计列表-富宇化工

序号	词	出现次数	出现频率
1	2016	123	0.6811
2	增长	116	0.6423
3	烃	85	0.4707
4	2017	79	0.4374
5	产品	79	0.4374
6	化工	79	0.4374
7	资产	77	0.4264
8	芳	74	0.4097
9	负债	62	0.3433
10	债务	57	0.3156
11	流动	55	0.3045
12	收入	53	0.2935
13	有限	52	0.2879
14	企业	51	0.2824
15	经济	50	0.2769

序号	词	出现次数	出现频率
16	业务	46	0.2547
17	价格	43	0.2381
18	辰	42	0.2326
19	分别	42	0.2326
20	炼油	40	0.2215
21	原油	40	0.2215
22	利润	37	0.2049
23	流	37	0.2049
24	营业	37	0.2049
25	规模	36	0.1993
26	费用	35	0.1938
27	能力	35	0.1938
28	石油	35	0.1938
29	2014	34	0.1883
30	设备	34	0.1883
31	我国	34	0.1883
32	发展	33	0.1827
33	高	33	0.1827
34	影响	32	0.1772
35	生产	31	0.1717
36	项目	31	0.1717
37	油	31	0.1717
38	致	31	0.1717
39	利息	30	0.1661
40	2015	29	0.1606
41	提供	29	0.1606
42	下降	29	0.1606
43	评级	28	0.155

序号	词	出现次数	出现频率
44	其中	28	0.155
45	燃料油	28	0.155
46	比率	26	0.144
47	电力	26	0.144
48	产业	25	0.1384
49	持股	25	0.1384
50	持续	25	0.1384
51	结构	25	0.1384
52	情况	25	0.1384
53	石化	25	0.1384
54	一定	25	0.1384
55	债券	25	0.1384
56	担保	24	0.1329
57	根据	24	0.1329
58	投资	24	0.1329
59	现金	24	0.1329
60	增加	24	0.1329
61	东营	23	0.1274
62	风险	23	0.1274
63	进口	23	0.1274
64	贸易	23	0.1274
65	销售	23	0.1274
66	构成	22	0.1218
67	保障	21	0.1163
68	短期	21	0.1163
69	集团	21	0.1163
70	资金	21	0.1163
71	国际	20	0.1107

序号	词	出现次数	出现频率
72	借款	20	0.1107
73	来源	20	0.1107
74	票据	20	0.1107
75	其他	20	0.1107
76	市场	20	0.1107
77	受	20	0.1107
78	为主	20	0.1107

通过对 400 份报告的文本统计，可以收集出句子、数字等部分文本特征，但并未体现出债券评级与文本统计之间的相关性。

3.2 信评报告的文本分类差异

3.2.1 文本分类方法说明

文本分类是对文本按照一定的分类体系或标准进行自动分类标记。通俗的讲，文本分类不是按统计指标将文本加以区分，而是将文本的每个词做成词向量，放到工具黑盒中加以识别；文本中的每个词都可以在计算机里表示成数字结构的向量，这个向量一般是经过大量语料库训练出来的。

随着时间的推移，文本分类的主流方法从浅层学习的文本分类模型（例如，朴素贝叶斯模型，K 紧邻模型，支持向量机模型）占主导地位，到现在变成了以深度学习模型为主导地位。目前主流的研究方法通过改进 CNN、RNN 和添加注意力机制或进行模型融合和多任务方法，提高了针对不同任务的文本分类性能，目前常用的方法包括 TextRCNN、FastText、DPCNN、Transformer 等。本文使用 TextRCNN 作为文本分类方法。若能验证信评报告的文本间存在差异，就可能通过比较差异，发现评级虚高的个体。

3.2.2 不同评级机构的文本分类差异

单一评级机构的文本分类测试。以中诚信为例，输入中诚信的评级报告，判断是否为中诚信的信评报告。通过 150 份报告的训练，以 31 份样本进行测试，判断是否为中诚信报告的二分类准确率达到 90.3%，即在 31 份测试信评报告中，正确识别出 28 份。

多家评级机构的文本分类测试。按照不同的信评公司进行分类，选取信评报告较多的几家评级公司：中诚信（184 份），大公国际（135 份），联合信用（114 份），东方金诚（85 份）。输入未知评级机构的信评报告，输出是这份报告是否属于中诚信、大公国际、联合信用或东方金诚。使用 518 份信评报告中的 120 份作为测试数据，其中正确识别出 111 份，准确率为 92.5%，即一份未知机构的评级报告有 92.5% 可能性被判断出是哪家评级公司出具的。

由此可见，通过 TextRCNN 的方法测试得出结论一：不同评级机构的信评报告之间存在文本差异。不同评级公司出具的信评报告具体自身公司的文本特征，或许是因为各家机构的信评报告已是非常模板化；如果出现一份信评报告的文本偏离公司的整体特征，就释放出文本差异的信号，需要投资者加以关注。

3.2.3 不同评级的信评报告分类差异

使用 TextRCNN 进行不同评级的信评报告的二分类（不低于 AA 评级，低于 AA 评级）模型训练。输入未知评级的信评报告，输出是这个报告低于 AA 还是不低于 AA 的，与实际评级相比，它输出的分类的正确比例就是准确率

同一评级公司的二分类模型构建。使用大公国际 135 个信评报告进行训练和测试，其中包括不低于 AA 评级的信评报告 101 个，低于 AA 评级的信评报告 34 个。按 5:5 进行训练集和测试集划分，二分类准确率为 89.6%。

所有评级公司的二分类模型构建。使用 518 个信评报告进行训练和测试，其中包括不低于 AA 评级的信评报告 454 个，低于 AA 评级的信评报告 64 个。按 5:5 进行训练集和测试集划分，二分类准确率为 84.9%。

由此可见，通过 Text RCNN 的方法测试得出结论二：不同评级的信评报告之间存在文本差异。信评报告的文本存在评级分档和评级机构两方面的差异。举例说明，评级 AAA、AA 和 A 的评级报告在文本方面存在差异，用 TextRCNN 方法可以识别出偏离的个体，如被给予 AA 评级但确未被确认为符合 AA 群体的文本。

四、信评报告的情感分析

4.1 情感分析应用方法说明

情感分析是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。中文情感分析按照处理文本的粒度不同，可分为词语级、句子级、文本级三个研究层次，即从分析某一个词的正负极性到分析某一句的情感，到分析某一段话的情感。随着深度学习的发展，使用深度学习来进行情感分析也逐渐流行起来。近年来该领域的研究方法主要有基于字典和基于语料库两种。基于语料库的词语级的情感判别主要是根据它们的语法特性，对大规模语料库进行信息挖掘，从而得到统计数据并对其极性做出判断。基于词典的词语级的情感判别，主要是根据词典 WordNet 或 HowNet 中词语间的关联来判别词语的极性。

BosonNLP 是基于情感词典的方法，基于半监督机器学习引擎的中文语义开放平台，拥有情感分析，信息分类，实体识别，典型意见，文本聚类等多个工具。BosonNLP 情感分析引擎具有预测准确率高的优点，正负面情感分析准确度达到 80%~85%，经过行业数据标注学习后可达到 85%~90%。BosonNLP 实体识别引擎基于自主研发的结构化信息抽取算法，F1 分数达到 81%。通过对行业语料的进一步学习，可以达到更高的准确率。可定制数据分析模型和解决方案，针对需求提供分类、消歧、典型意见提取等定制机器学习模型的建立和 API 服务。

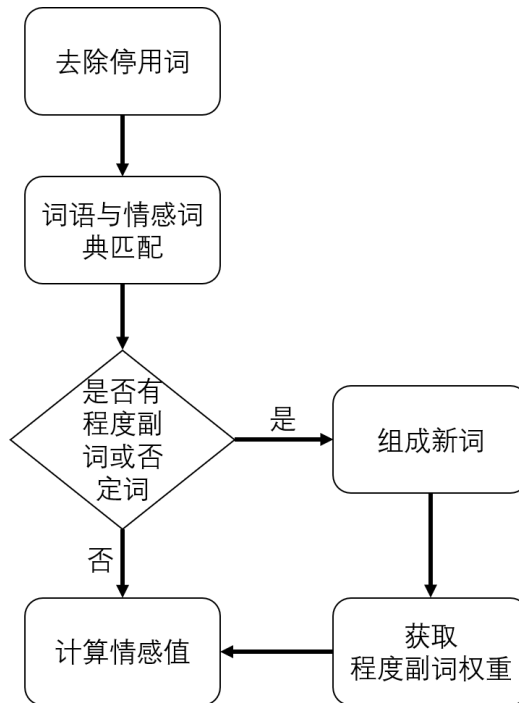
常用的深度学习分类器 LSTM 是 Hochreiter 和 Schmidhuber 在 1997 年提出的用于解决梯度消失和梯度爆炸的循环神经网络。相比于之前的 RNN，LSTM 能够在长序列上得到更好的表现。本文使用 LSTM 来进行句子情感值计算模型的训练。

4.2 信评报告的情感分析

本文应用BosonNLP和LSTM从词语级别、句子/文档级别两个方面来对文本进行情感分析，在分词的基础上，对信评报告的每一个词，每一个句子都进行情感计算，为后续的情感指标体系构建提供基础。

4.2.1 词语级别情感分析

本文基于BosonNLP情感词典、停用词典、否定词典、程度副词词典进行基于词典的词语情感分析，在情感词典中，词语标注的情感值大于0，则词语偏积极，情感值小于0，则词语偏消极，并且情感值的绝对值越大，其积极或消极的程度越大。进行词语级别情感分析的具体流程如下图所示。



词语情感值计算流程

首先根据停用词表去除文本；然后将文本中的每个词与情感词典中的词语进行匹配，

得到其情感值；之后判断词语前是否存在否定词或者程度副词，如果有，；如果有否定词，将情感值取反，如果有程度副词，从程度副词词典中查找该程度副词的权重，将情感值乘以程度权重，得到词语的最终情感值。

4.2.2 句子级别情感分析

从信评报告中的“优势”部分取出 3000 条语句作为积极语句，情感值标注为 1，从“关注和风险”部分取出 3000 条语句作为消极语句，情感值标注为-1，按 8:2 的比例（4800 条作为训练数据，1200 条作为测试数据）进行训练集和测试集划分，使用这些已经标注好情感极性的句子进行句子级别情感计算模型的训练或评估。

积极语句示例：

- 1) 公司位列最具规模中国饭店集团前列，在国内星级酒店行业仍具有较强竞争优势；
- 2) 公司向托管酒店输出成熟的管理模式和管理系统并收取管理费，盈利能力依然较强；
- 3) 乡村度假酒店作为公司开发的新型特色业务模式，有助于公司进一步提升品牌竞争力；
- 4) 公司酒店板块拥有网络化的销售渠道和成熟的中央预定体系，应用领先的 CRM 客户管理系统，客户稳定性较高；

消极语句示例：

- 1) 2020 年一季度，受疫情影响，公司营业收入大幅下滑，净利润同比转亏，公司面临一定的经营风险；

2) 2019 年，公司期间费用对利润形成一定侵蚀，且非经常性损益对利润的影响程度较高，影响盈利的稳定性；

3) 2019 年以来，公司其他应收款规模较大且部分为关联方往来款，对资金形成较大的占用；

4) 截至 2020 年 3 月末，公司受限资产规模仍较大，占净资产的比重仍较高，影响公司资产流动性；

5) 公司部分被担保企业经营亏损，仍存在一定代偿风险；

为了便于后续情感指标体系的构建，本文将句子的情感值归一化到[-1,1]之间，大于 0 表示句子情感值偏积极，1 表示最积极，小于 0 表示句子情感值偏消极，-1 表示最消极。

情感分析评分示例如下：

表 13 情感分数表-示例

语句	情感评分（-1 为最消极，1 为最积极）
----	----------------------

公司位列最具规模中国饭店集团前列，在国内星级酒店行业仍具有较强竞争优势；	0.86（积极）
公司向托管酒店输出成熟的管理模式和管理系统并收取管理费，盈利能力依然较强；	0.54（积极）
乡村度假酒店作为公司开发的新型特色业务模式，有助于公司进一步提升品牌竞争力；	0.73（积极）
公司酒店板块拥有网络化的销售渠道和成熟的中央预定体系，应用领先的 CRM 客户管理系统，客户稳定性较高；	0.42（积极）
2020 年一季度，受疫情影响，公司营业收入大幅下滑，净利润同比转亏，公司面临一定的经营风险；	-0.67（消极）
2019 年，公司期间费用对利润形成一定侵蚀，且非经常性损益对利润的影响程度较高，影响盈利的稳定性；	-0.34（消极）
截至 2020 年 3 月末，公司受限资产规模仍较大，占净资产的比重仍较高，影响公司资产流动性；	-0.12（消极）
公司部分被担保企业经营亏损，仍存在一定代偿风险	-0.58（消极）

本文使用两种方法分别进行句子级别情感分析的研究。一是基于情感词典，即在词语级别情感分析的基础上进行句子情感值计算，对每一个句子，其情感值为句子中所有词语的情感值之和，加和后对句子的情感值进行归一化处理，得到[-1,1]区间范围内的句子情感值。此种方法不需要训练，只需要测试。二是使用机器学习预测模型对句子进行情感值计算，本文使用 LSTM 来进行句子情感值计算模型的训练。将情感分析作为一个有监督的分类问题，对训练文本进行人工标注，然后进行有监督的机器学习，从而得出模型来进行预测。此种方法需要用 6000 条中 4800 条做模型训练。另外的 1200 条用来评估这两种方法的效果。

基于以上两种方法分别构建情感二分类模型，如果情感值大于 0，将其归类为积极句子，否则，将其归类为消极句子。使用测试数据 1200 条对两种模型进行评估，以积极句子为正例，消极句子为负例，使用准确率、精确率、召回率以及 F1 值作为二分类模型的评估指标，首先对几个变量进行定义：

TP(True Positive): 把积极句子分类为积极句子的数量，是正确分类；

FN(False Negative): 把积极句子分类为消极句子的数量，是错误分类；

TN(True Negative): 把消极句子分类为消极句子的数量，是正确分类；

FP(False Positive): 把消极句子分类为积极句子的数量，是错误分类；

准确率是指有在所有的判断中有多少判断正确的，其计算方法如下：

$$\text{Acc} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FN} + \text{FP})$$

精确率是指在预测为正例的样本中有多少是正确的，其计算方法如下：

$$\text{Pre} = \text{TP} / (\text{TP} + \text{FP})$$

召回率是指预测正确的正例占有所有正例数据的比例，其计算方法如下：

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

精确率和召回率通常是相互矛盾的，F1 是精确率和召回率的加权调和平均，其计算方法如下：

$$\text{F1} = \frac{2 * \text{Pre} * \text{Recall}}{\text{Pre} + \text{Recall}}$$

本文使用准确率、召回率以及 F1 值作为二分类模型的评估指标，使用测试集进行情感极性分类测试，结果如表所示。

表 14 情感分析分类测试结果

	准确率	精确率	召回率	F1
情感词典计算模型分类	90.2%	89.2%	90.6%	89.89%
机器学习预测模型分类	93.1%	92.6%	93.8%	93.20%

由以上结果可知，两种方法都能够较好地对句子情感极性进行分析。为了提高情感分析模型的容错能力，本文融合两种方法构建句子情感值的计算模型，句子级别情感值 SentiScore 计算方法如下：

$$\text{SentiScore} = \begin{cases} S1, & \text{if } |S1| \geq |S2| \\ S2, & \text{if } |S1| < |S2| \end{cases}$$

其中，S1 为使用基于词典的情感分析计算模型计算出的句子情感值，S2 为使用机器学习模型计算出的句子情感值，取二者中极性程度较大的作为最终的句子情感值。

五、发现评级虚高信号的实验及结论

5.1 归纳实验样本数据

本文以 518 份债券发行信评报告为样本，采用自然语言和机器学习工具进行样本数据的整理，就其结果归纳如下：

表 15 样本数据归纳结果

序号	数据项	数值	应用方法	方法和结果说明
1	报告数（份）	518	n/a	
2	中文分词准确率	85%	PKUseg	85%是应用 PKUseg 的大致分词准确率，不是特指信评报告的分词准确率；本文不是研究提高分词准确率，因而此数据不是本文的重点，支持本文实验即可。
3	信评公司正确分类（份）	479	TextRCNN	数据说明不同信评公司之间的信评报告存在特征差异
4	信评公司分类准确率	92.5%		
5	评级正确分类（份）	440		数据说明不同评级的信评报告之间存在特征差异
6	评级分类准确率	84.9%		
7	基于情感词典的文本分析准确率	90.2%	BosonNLP	此准确率是对报告文本每一个句子进行情感分析的准确率，训练数据是从所有报告中选出的“风险和关注”语句和“优势”语句各 3000 条；准确率数据说明句子级别情感分析模型的准确性高，基于此准确率进行情感分析是可行的。
8	基于机器学习的文本分析准确率	93.1%	LSTM	

5.2 构建信评报告的情感指标计算模型

在第四部分情感分析的研究基础上，本文针对信评报告进行情感指标计算模型的构建，以发现评级虚高的信号。以下是构建情感指标计算模型的三个基本假设：

(1) 对于信用评级越高的主体，其信用评级报告的情感属性越偏向积极；

(2) 信用评级机构会刻意提高信用评级，但不会过多粉饰信用评级报告；

(3) 对发行评级以后未下调过评级或有过违约行为的主体，其发行评级报告相对客观公正或者说虚高情况较少较轻，称这一部分评级报告为“可靠报告组”。

以“可靠报告组”作为训练数据进行情感指标计算模型的构建。具体方法如下：

(1) 按照“可靠报告组”的评级对其进行情感指标标注，将其中 80%的数据作为训练数据进行模型训练，20%的数据作为测试数据，具体情感指标映射表如下：

表 16 情感指标映射表

评级	情感指标
AAA	1.0
AA+	0.9
AA	0.8
AA-	0.7
A	0.6
BBB	0.5
BB	0.4
B	0.3
<B	0.2

(2) 将“可靠报告组”中所有句子按情感值区间进行分类，按照步长 0.4，统计每个情感值区间的句子数量占比，如下表所示：

表 17 情感值区间与权重

句子情感值区间	计算权重
[-1,-0.6)	W_0
[-0.6,0.2)	W_1
[-0.2,0.2)	W_2
[0.2,0.6)	W_3
[0.6,1.0]	W_4

其中 W_i 为对应区间的句子数量占比，并且有

$$W_0+W_1+W_2+W_3+W_4 = 1.0$$

(3) 使用神经网络进行情感指标计算模型训练，输入参数为第 2 步中的每个情感值区间的句子数量占比 (W_i)；

使用测试集评估该模型的效果，使用平均绝对误差作为评估指标，其计算方法如下所示，

$$avgd = \frac{\sum |P_i - T_i|}{N}$$

其中， P_i 为预测情感值， T_i 为实际情感值， N 为信评报告总数。

针对“可信报告组”的测试组数据，模型进行预测的平均绝对误差为 0.083。

5.3 实验结果说明

将主体发生评级下调或者违约的发行评级报告作为不可靠报告并列入“不可靠报告组”，同时将主体发生违约的作为违约报告并列入“违约报告组”，比较这两组与“可靠报告组”的实验数据，以验证评级虚高信号。具体方法如下：

(1) 根据不可靠报告和违约报告的评级赋予各报告一个情感指标。

(2) 用已构建的信评报告情感指标计算模型对不可靠报告和违约报告的情感指标进行计算。

(3) 将计算结果与第 1 步的情感指标比对，计算平均绝对误差。

在全部 518 份报告中，可靠报告组的报告 422 份，不可靠报告组的报告 96 份（其中 60 份构成违约报告组），基于信评报告情感指标计算模型得出的各组平均绝对误差如下：

	违约报告组	不可靠报告组	可靠报告组
报告数（份）	60	96	422
预测平均绝对误差	0.157	0.124	0.083

数据说明预测模型在违约报告、不可靠报告和可靠报告之间是有差别的；就平均绝对误差而言，违约报告组最高，不可靠报告组次之，可靠报告组最低。平均绝对误差值高说明在报告组中偏离正常情况的异常报告更多，并且大部分是计算所得的文本情感值低，即评级虚高的报告。三个报告组之间平均绝对误差值的差异说明情感指标差异可以作为评级虚高的一个信号。

5.4 结论

本文通过文本分类方法来对不同评级公司的、不同评级的报告进行分类模型构建，得到效果良好的分类模型，说明不同评级公司的信评报告之间、不同评级级别的信评报告之间存在文本差异。这些差异是能够通过机器学习模型训练得到的，尤其是不同评级报告之间的差异可以用于发现评级虚高的信号。

运用文本情感分析方法，本文以可靠信评报告构建了信评报告的情感指标计算体模型，通过与不可靠信评报告和违约信评报告的对比实验，发现不同组别的报告存在情感偏差程度的差异，这一通过文本和情感分析所获得的差异有助于捕捉评级虚高的信号。

基于本文情感分析的实验结果，可参考如下步骤通过情感分析进行评级虚高的识别：

- (1) 根据报告的评级赋予该报告一个情感指标；
- (2) 使用已构建的信评报告情感指标计算模型对报告的情感指标进行计算；
- (3) 将第 2 步的计算结果与第 1 步比对，如果差距大于某阈值，则可判定为该报告为评级虚高报告。

参考文献

- [1] Lai, Siwei, et al. "Recurrent convolutional neural networks for text classification." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 29. No. 1. 2015.
- [2] Luo, Ruixuan, et al. "Pkuseg: A toolkit for multi-domain chinese word segmentation." *arXiv preprint arXiv:1906.11455* (2019).
- [3] Min, Kerui, et al. "BosonNLP: An ensemble approach for word segmentation and POS tagging." *Natural Language Processing and Chinese Computing*. Springer, Cham, 2015. 520-526.
- [4] Hochreiter, Sepp, and Jürgen Schmidhuber. "LSTM can solve hard long time lag problems." *Advances in neural information processing systems* (1997): 473-479.
- [5] Feng Li. "Annual report readability, current earnings, and earnings persistence" *Journal of Accounting and Economics* 45(2008)221-247
- [6] Wenxing Lu, Yanfei Wang. "Review of Chinese text sentiment analysis" *10.3969/j.issn.1001-3695.2012.06.003*