

基于文本挖掘技术的煤炭行业环境责任披露研究

费 琪, 龚方圆, 程相素, 阎虎勤

厦门国家会计学院, 福建 厦门

Email: 156208301@qq.com, 351003774@qq.com, Sophia_cxs@163.com, yanhuqin@xnai.edu.cn

收稿日期: 2021年5月17日; 录用日期: 2021年6月4日; 发布日期: 2021年6月11日

摘 要

在可持续发展政策的要求下, 环境保护是煤炭企业履行社会责任的重要部分。本文利用Python语言所提供的BeautifulSoup网页爬虫工具、Wordcloud词云分析工具、以及Jieba汉字处理工具, 选择A股市场十家具有代表性的煤炭企业披露的环境保护社会责任报告, 进行文本挖掘, 得出煤炭行业信息披露的现状及存在的缺陷, 并结合煤炭行业自身特点提出了实际可行的建议。

关键词

文本挖掘, 环境责任, 煤炭行业, 披露质量

Analysis on Environmental Responsibility Disclosure of Coal Industry Based on Text Mining Technology

Qi Fei, Fangyuan Gong, Xiangsu Cheng, Huqin Yan

Xiamen National Accounting Institute, Xiamen Fujian

Email: 156208301@qq.com, 351003774@qq.com, Sophia_cxs@163.com, yanhuqin@xnai.edu.cn

Received: May 17th, 2021; accepted: Jun. 4th, 2021; published: Jun. 11th, 2021

Abstract

According to the requirement of sustainable development policy, environmental protection is an important part of coal enterprises' social responsibility. In this paper, we select the environmental protection social responsibility reports of ten representative coal enterprises in the A-share mar-

ket, using the language of Python provided by the BeautifulSoup web crawler tool, word cloud analysis tool, and Jieba Chinese character processing tool to carry on the text mining, obtain the status and defects of the coal industry information disclosure, and finally put forward some practical suggestions combined with the characteristics of the coal industry.

Keywords

Text Mining, Environmental Responsibility, Coal Industry, Disclosure Quality

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自我国进入工业时代以来,经济蓬勃发展,作为国家能源重要占比的煤炭更是推动国家经济发展的重要动力。然而,煤炭行业在开采过程中却存在破坏地表植被、水土流失和污染环境的隐患[1],为了与可持续发展的理念相符合,国家也在新《环境保护法》中出台了相应政策,要求企业每年按照规定进行环境、社会和治理报告披露。企业按照国家相关政策规定定期披露在环境保护方面的社会责任不仅有利于企业自身建立良好的公众形象,而且有利于推行党的十九大提出的“建设美丽中国”方针[2]。如今煤炭行业中多数公司虽然会每年定时披露环境责任履行报告,但是目前企业披露的环境保护信息质量如何却没有分析结论。上市公司披露的信息质量如何?是否相关可靠全面?信息披露标准是否统一?各企业披露的信息之间又是否具有可比性?为了进一步规范企业的环境信息披露质量,这些问题都是必须分析回答的。

在以往学者对于企业环境责任报告的研究中,由于受到了分析工具的限制,复杂的分析环境为分析工作的进行带来了巨大难题[3]。方面海量的责任报告信息难以收集整理,不同企业的社会责任报告零星分散在企业官网中,难以收集提取关键信息进行企业间的横向对比。另一方面,受到信息识别技术的约束,在进行环境责任报告重要信息提取时,报告中图表混杂,关于煤炭行业的专有词汇更是晦涩难懂,为信息质量分析工作增添了难度。而 Python 语言所提供的 BeautifulSoup 网页数据爬虫、和 Jieba 汉字词库技术完美得解决了以上难题。BeautifulSoup 网页数据爬虫能够快速爬取网页信息,并将其转换为文本格式,以 Jieba 汉字词库为依据,进行文本拆分,再利用 Wordcloud 词云统计文本中出现的高频词,最终依据词频生成客观形象的词云[4]。

本文借助 Python 的 BeautifulSoup + Wordcloud + Jieba 文本挖掘方法,对煤炭行业十家上市公司的社会环境责任报告进行了文本挖掘。首先提取责任报告中高频词,分析煤炭行业在进行社会责任报告时的侧重点。再利用 ROST CM6 中的加载项工具 Net Draw,分析所提取出的文本中各节点之间的逻辑关系及联系的紧密程度,形成语义网络图,并从宏观和微观方面分析整个行业在进行环境保护信息披露时所包含的层次[5]。最后,我们单独对煤炭行业的龙头企业中国神华进行了词频统计和网络语义图的绘制,进行行业内信息披露质量的横向对比,从而能够对煤炭行业环保信息披露质量形成充分而深刻的剖析。本文的研究方法运用了目前最新的计算机语言技术,为煤炭行业环保信息披露质量的研究提供了新思路。

2. 数据获取及数据预处理

在充分考虑市值大小及数据可靠性等因素后,本文选取了 A 股中十家具有代表性的企业,分别是中国神华、陕西煤业、中煤能源、兖州煤业、淮北矿业、西山煤电、潞安环能、冀中能源、平煤股份和开

根据环境报告分词后出现的词语和频数, 本文发现报告中提及较多的词语是“环保”、“排放”、“节能”、“治理”、“环境”、“改造”等, 如“环保”一词在文中出现了 130 次, 而“排放”一词出现了 93 次。同时我们发现报告中多次提及的污染物主要有“废弃物”、“污水”、“瓦斯”、“废气”等。依据词频的统计结果, 本文选取了前 100 个高频词汇, 绘制的词频统计表如表 1。

Table 1. Coal industry environmental protection word frequency statistical table

表 1. 煤炭行业环境保护词频统计表

序号	关键词	频数	序号	关键词	频数	序号	关键词	频数	序号	关键词	频数
1	环保	130	26	技术	26	51	积极	15	76	质量	12
2	排放	93	27	燃煤	25	52	烟气	15	77	员工	12
3	节能	64	28	应急	25	53	方案	14	78	气候变化	11
4	治理	61	29	废弃物	25	54	年度	14	79	控制	11
5	环境	53	30	污水	24	55	指标	14	80	监督	11
6	改造	52	31	标准	23	56	设备	14	81	完善	11
7	能源	49	32	清洁	22	57	系统	14	82	突发	11
8	煤矸石	42	33	生态	22	58	大气污染	14	83	法律法规	11
9	污染	40	34	矿区	21	59	发电	14	84	生态环境	11
10	矿井	40	35	社会	21	60	安全	13	85	达标排放	11
11	煤炭	39	36	生活	19	61	气体	13	86	资源	11
12	绿色	39	37	锅炉	19	62	目标	13	87	有害	11
13	生产	39	38	瓦斯	19	63	报告	13	88	改善	11
14	发展	37	39	持续	18	64	案例	13	89	标煤	11
15	监测	37	40	落实	18	65	能力	13	90	余热	11
16	煤矿	37	41	责任	18	66	土地	12	91	照明	11
17	开展	36	42	能耗	17	67	制定	12	92	总量	10
18	设施	36	43	矿山	17	68	综合治理	12	93	温室	10
19	国家	34	44	强化	17	69	复垦	12	94	管理制度	10
20	污染物	33	45	废水	17	70	回收	12	95	政府	10
21	减排	30	46	高效	16	71	危废	12	96	风险	10
22	处置	30	47	管控	16	72	理念	12	97	隐患	10
23	防治	29	48	排放量	16	73	开采	12	98	备案	10
24	工程	27	49	全面	15	74	监控	12	99	煤场	10
25	电厂	26	50	废气	15	75	水平	12	100	井下	10

3. 基于文本挖掘的煤炭行业对环境的社会责任披露状况分析

从词频统计表中不难看出, 煤炭企业对于环境方面社会责任的履行主要集中在“环保”“节能减排”、“治理污染”、“绿色生产”等关键词上, 这一方面是基于全球气候变暖、能源危机、生态环境恶化等宏观环境的要求, 另一方面也是煤炭行业实现可持续发展的必要条件。词频统计表仅仅统计了关键词的

频数，而关键词语义网络图不仅展示出了关键词之间的语义联系，还反映了煤炭行业履行对环境的社会责任的过程与具体措施，从这些关键词以及其间的联系，可以大致看出煤炭行业对环境方面的社会责任披露的重点内容及其影响因素。

3.1. 披露的影响因素

位于关键词语义网络图中心位置的就是“中国”和“政府”，这是大部分国内企业履行社会责任的重大推动力。尤其是对于煤炭行业这样高能耗、高排放、高污染的行业来说，要履行环境方面的社会责任是一项成本极高，且经济效益低的行为，是不符合企业追逐经济利益的目标的。但走出微观，放眼宏观，为了人类社会的可持续发展，企业必须承担起其应有的社会责任[7]。我国对于企业在环境方面需要履行的义务主要是以污染防治的各项具体指标来规定的，在关键词语义网络图中可以很清晰的看到，与“中国”和“政府”相关联的有“管理”、“排放”、“报告”等，这类关键词的关联主要是源于中国政府要求企业对其污染排放进行管理并对外报告其管理过程与成效。此外，在词频统计表中也有一些关键词表明了包括政府在内的外部环境对企业在环境方面的社会责任的要求，例如“达标排放”、“法律法规”、“监督”、“气候变化”、“指标”、“积极”、“管控”、“责任”、“减排”等，这其中不仅有国家法律法规的要求，也有社会公众、企业的利益相关者等对于身处煤炭行业的企业履行其社会责任的期盼。

3.2. 披露的内容

外部环境的要求催生了企业内部的变革。首当其冲的就是企业内部环境管理体系的建立与完善，在词频统计表中，“生产”、“监测”、“防治”、“工程”、“综合治理”、“理念”、“管理制度”等关键词就表明了作为样本选取的这些煤炭企业在环境保护措施与内部管理制度方面的探索过程。一套有效运行的管理制度不仅可以实现从管理层到“员工”的每一个关键节点的全程“监控”与“责任”“落实”，还能及时发现“风险”“隐患”并迅速反应，积极“备案”。而对于煤炭行业来说，其整个生产过程中都会伴随着各种风险，这里的风险不仅仅是破坏环境的风险，还包括员工的安全风险、意外事故风险等，环境管理体系的建立不仅对环境保护起作用，还会对整个生产环节起作用，不只保护环境，还能保障安全生产。因此，企业对环境管理体系的建设、改进与披露是很有必要的。

在完成了企业内部环境管理体系的建设之后，就应当通过具体的措施来实现对环境的治理。在关键词语义网络图中，与“企业”相关联的关键词就很明显的展现出了企业履行社会责任的具体措施。

首先是对污染进行防治，结合图表，“减排”、“管控”、“检测”、“废气”、“排放”、“废水”、“防治”等关键词都表明了现有的环境管理体系之下，煤炭行业最首要的任务仍是控制污染物的排放，从源头上减少对环境的破坏，通过制定标准和指标，全面落实责任管控，从而实现污染物的减排。

其次是降低能耗，提高资源的利用率，具体的关键词有“高效”、“利用”、“能源”、“技术”、“能耗”、“改造”、“电厂”、“优化”等。能耗的降低离不开技术的进步，而产业链的延伸又能够提高资源的利用效率，通过“矿山”、“电厂”、“矿区”、“煤场”等区块的结合，加以资源再利用的技术，可以最大限度的提高整个生产区域的资源使用效率。

最后是对环境已经被破坏的区域进行修复，建设生态矿区。图表中的“综合治理”、“复垦”、“绿色”、“生态”、“生活”、“发电”、“改造”、“供热”、“照明”等关键词就反映了这一点。煤炭行业的特殊之处在于其与自然环境的紧密关联性，矿区内既要进行资源开采的生产活动，又要保障员工的正常生活。由于早年间企业的责任意识薄弱，在进行生产活动的同时对矿山的生态环境造成了严重的破坏，尤其是煤炭产业大多分布在我国北方地区，资源开采的过程中破坏过的土地大多变成了风沙

区，并且逐年扩大，对人们的生产生活造成了恶劣的影响。当企业实现了自身的节能减排之后，环顾自身周围的环境，保持或恢复其原有的生态就成为了首要任务。从词频统计表的相关关键词中不难看出，煤炭企业主要依靠能源的高效利用和对矿山的综合治理来实现生态建设。

4. 基于行业内龙头企业的纵向对比

我们在整理样本社会责任报告的环境责任披露信息中发现，作为稳居煤炭行业市值第一且遥遥领先的中国神华能源股份有限公司，在 A 股上市之后一直保持发布社会责任报告，且披露内容最为详细和充分，于是我们对其环境责任披露内容做了同样的文本化处理，并用文本挖掘技术对词频最高的前 100 个词进行提取，然后利用 excel 的 vlookup 函数与前文中 A 股 10 家上市公司样本数据中频数最高的前 100 个词进行匹配，进行比例的计算，然后筛选出前者词频/后者词频>10%的词语，并按比例大小进行降序排列后，得到以下数据：

Table 2. Shenhua environmental responsibility disclosure word frequency list
表 2. 中国神华环境责任披露词频表

关键词	神华词频	所有样本词频	占比	关键词	神华词频	所有样本词频	占比
隐患	9	10	90.00%	目标	4	13	30.77%
监督	8	11	72.73%	管理制度	3	10	30.00%
风险	7	10	70.00%	发展	11	37	29.73%
管控	11	16	68.75%	强化	5	17	29.41%
温室	6	10	60.00%	技术	7	26	26.92%
气候变化	6	11	54.55%	处置	8	30	26.67%
矿山	9	17	52.94%	全面	4	15	26.67%
煤炭	20	39	51.28%	生活	5	19	26.32%
年度	7	14	50.00%	治理	16	61	26.23%
气体	6	13	46.15%	应急	6	25	24.00%
能源	22	49	44.90%	排放	22	93	23.66%
燃煤	11	25	44.00%	煤矸石	9	42	21.43%
高效	7	16	43.75%	方案	3	14	21.43%
环境	23	53	43.40%	防治	6	29	20.69%
监测	16	37	43.24%	改造	10	52	19.23%
生态	9	22	40.91%	矿区	4	21	19.05%
井下	4	10	40.00%	排放量	3	16	18.75%
电厂	10	26	38.46%	清洁	4	22	18.18%
安全	5	13	38.46%	减排	5	30	16.67%
废弃物	9	25	36.00%	污水	4	24	16.67%
落实	6	18	33.33%	污染	6	40	15.00%
土地	4	12	33.33%	环保	16	130	12.31%
绿色	12	39	30.77%	污染物	4	33	12.12%

紧接着，我们用 ROST CM6 进行语义网络分析，生成了中国神华环境责任披露的语义网络图：

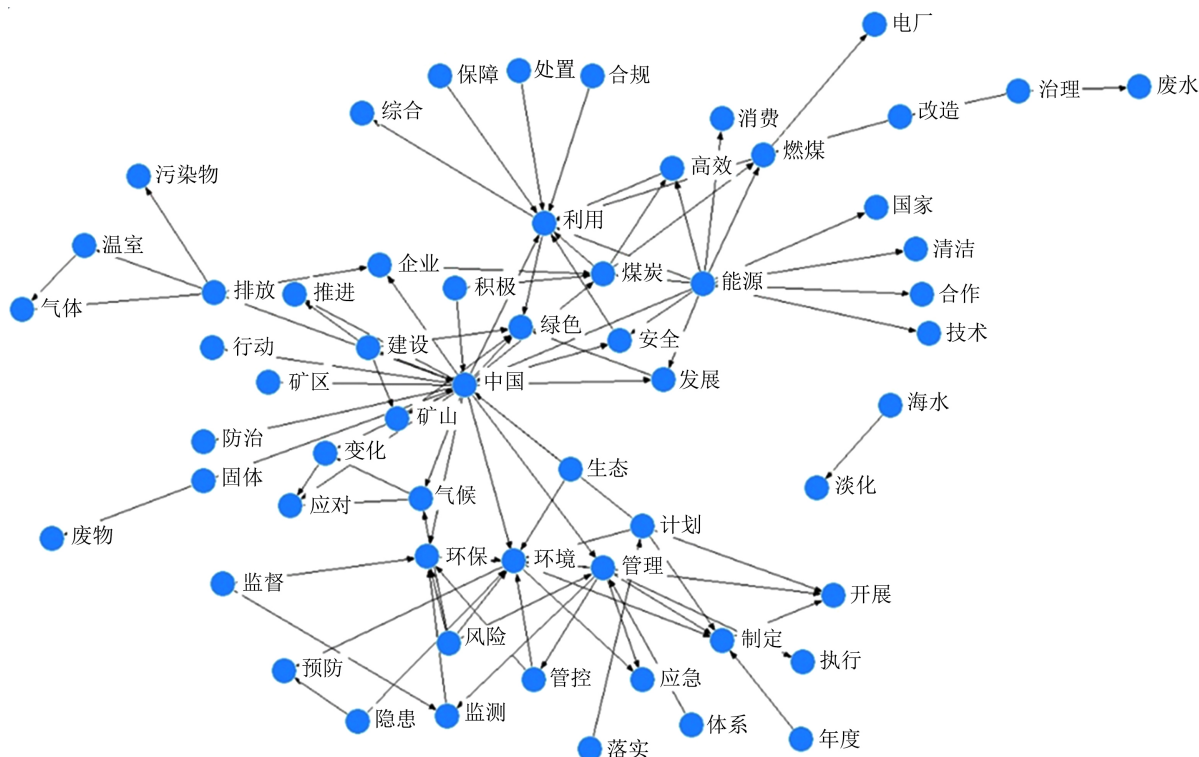


Figure 3. Shenhua environmental responsibility disclosure semantic network graph

图 3. 中国神华环境责任披露语义网络图

结合图 3 和表 2，我们进行更加深入的分析了中国神华环境责任履行和披露的特点。

4.1. 积极应对气候变化且披露充分

从图 3 左列图中可以看出“风险”、“气候变化”、“温室”、“气体”占比均超过了 40%，究其原因在于神华披露了大量有关于应对气候变化的责任履行情况。中国神华支持和参与温室气体减控方案的倡议，依据相关法律法规的要求，合理建立碳排放管理制度，同时积极响应国家气候相关政策的号召，依据企业自身污染物排放情况，进行了气候风险评估，将风险分类为一类风险、二类风险和三类风险，并依据国家政策方针和行业领先技术水平制定缓解，适应和积极管理的政策和战略，以应对气候变化。

4.2. 环境管理体系有效且披露完善

“隐患”、“监督”、“管控”、“监测”、“治理”、“管理制度”等词占比高，其中，“隐患”占比高达 90%，原因在于神华充分披露了整个环境管理体系架构、具体职责和工作内容，其为强化对环境风险的管控能力，尽可能避免出现违反法律法规的情况，建立了严格有效的全过程环境风险管控机制，涵盖了环保监督检查，环境隐患预防，风险应急管理这三个方面。加强对污染物的监测管理力度和环保隐患排查治理力度，严格遵循突发环境事件应急预案备案程序[8]。

4.3. 节能降耗表现突出且披露具体

“能源”“高效”“煤炭”“绿色”“技术”“清洁”等词占比也不逊色，原因在于神华相比于其

他煤炭企业对节能降耗方面的高度重视。其积极开展燃煤绿色升级和燃煤工业绿色改造，系统提升能源使用效率，进一步深化节能降耗。在发挥煤炭价值的过程中，为避免因排放带来的污染，同时加大清洁煤炭产品研发力度，并采用先进技术，通过设备改造等手段，不断优化煤炭开采、洗选加工、运输流程，让煤炭的清洁利用和绿色转化成为现实。

4.4. 注重生态环境维护且披露全面

“生态”、“矿山”、“土地”、“绿色”等词反映了神华对生态环境问题的重视。煤炭企业生产运营原本就会给生态环境带来很多影响，环境污染和生态破坏的问题十分突出，而近年来对这些生态环境问题的全面的排查和整改，通过水土保持、防风固沙、塌陷区治理、复垦绿化、生态园区建设等工作，有效地恢复了井工矿区、露天矿区、铁路等作业区域生态环境，保护区域周边生态多样性。其以货币形式披露了相关的财务信息，也定量地披露了治理的成果，如植被覆盖率、复垦率、水土流失控制率等。

由于我们的分析指标是占比，也就是说神华越是重视、做得充分的方面恰是其他煤炭企业还有努力和改善空间的地方，我们通过对其他企业社会责任报告的具体阅读也佐证了这个结论。不论是从披露篇幅还是披露质量来说行业龙头的几家公司，都胜于其他公司。

5. 结论及建议

综上，基于文本挖掘技术和分析，我们总结了煤炭行业在环境责任披露方面主要存在以下两个问题：

5.1. 披露不全面、不具体

主要体现在两方面，一是重正面与轻负面，从词频表来看大部分都是体现企业积极举措的正面词汇，而鲜有负面词汇的出现。可见大部分企业的大部分披露内容均是一些正面信息，而少有甚至几乎没有负面信息，报喜不报忧，也就是人们常说的，太完美的东西反而显得不那么真实，事实是否如此我们不得而知。二是重定性而轻定量。大部分高频词在文本中检索时少以指标的形式来出现，大部分都是定性的文本描述，而定性和定量内容的结合能够增强报告信息的可靠性、可比性和相关性，便于使用者进行不同企业和不同期间的对比分析[9]。

5.2. 披露不均衡、不主动

近年来，我国煤炭行业上市公司环境信息披露数量渐渐增多，但其披露质量却良莠不齐，行业龙头的模范作用显著，而其他公司相比之下有所差距。而且，大部分企业是由于政策压力以及为了提升企业声誉才进行披露，而非主动披露。在对行业内企业横向对比分析时我们发现，有的企业只用了个位数的篇幅披露环境责任履行的信息，而对同一家企业做纵向的对比分析时发现，有的企业连续多年套用同样的披露模板，仅对部分定量数据做改动，而文字性的描述继续沿用。披露缺乏主动性，企业就会抱着应付相关部门的心态选择性披露而非全面披露，最终导致披露在完整性和可比性上有所欠缺。

针对以上披露现状，我们提出以下几点建议：

首先，煤炭企业的环境责任披露意识有待提高。披露情况不佳其实也从侧面反映出了企业环境责任意识不够强。而自觉按照相应社会契约的要求承担环境责任以及对其进行充分的披露不仅可以增加企业知名度，也能主动配合监管部门，以降低企业被干涉的风险和被干预后经济利益的流出。

其次，企业环境责任披露意识的提高离不开国家制度法规的指导和约束作用。环境信息披露制度的建立是依托于相关法律法规的建立，因此想要完善环境信息披露制度就应该先完善我国的环境法律法规，使之系统化、体系化。比如完善和统一煤炭行业污染排放标准，为企业提供环境披露的依据和标准，从而提高所披露环境信息的可比性和准确性[10]。

最后,相关部门的监管以及社会责任的监督也是不可忽略的因素。多层次、多主体的环境监管体系有助于督促企业做好环境责任披露工作。不论是相关部门的处罚还是社会舆论的监督 and 曝光也有助于倒逼企业更真实、完整和充分地披露环境责任履行情况,履行好企业环境责任。

致 谢

感谢厦门国家会计学院阎虎勤老师在《Python 财务数据分析》课程上对 BeautifulSoup + Wordcloud + Jieba 文本挖掘方法的讲解,让我们对于文本挖掘这一技术及其应用有了初步的认识与了解,帮助我们能够顺利进行信息搜集和整理,这是我们完成本篇论文不可或缺的前提。

基金项目

本论文得到了厦门国家会计学院 2019 年“云顶课题: Python 财务数据分析”项目和大米(厦门)科技股份有限公司的支持。

参考文献

- [1] 刘福民,高纪创,范开阳. 煤炭业上市公司环境会计信息披露研究[J]. 会计之友, 2019(7): 133-138.
- [2] 高彰. 浅析环境会计信息披露有效性问题——以煤炭业上市公司为例[J]. 知识经济, 2019(31): 85-87.
<https://doi.org/10.15880/j.cnki.zsji.2019.31.046>
- [3] 王军会,宿旭. 能源行业上市公司环境信息披露的数量和质量分析——基于年报和社会责任报告比较的视角[J]. 中国管理信息化, 2020, 23(23): 111-114.
- [4] 杨承启,庄心怡. 基于文本挖掘的能源与环境保护政策大数据研究[J]. 财会学习, 2021(7): 146-149.
- [5] 宋英慧,黄麒. 基于文本挖掘技术的财务报表附注披露研究[J]. 会计之友, 2019(1): 142-147.
- [6] 阎虎勤. Python 财务数据分析(讲义) [Z]. 厦门: 厦门国家会计学院, 2021.
- [7] 赵美,黄麒. 文本挖掘在企业社会责任报告研究中的应用探析[J]. 中国市场, 2017(19): 14-16+27.
- [8] 陈思言. 煤炭企业环境会计信息披露评价体系构建研究[D]: [硕士学位论文]. 昆明: 云南财经大学, 2020.
- [9] 崔思凡. 探析我国煤炭行业上市公司环境会计信息披露[J]. 河北农机, 2020(7): 115-116.
- [10] 刘银玲. 煤炭上市公司环境信息披露研究[J]. 会计之友, 2019(3): 117-122.