

● 牛高鸣, 姚长青, 高影繁, 梁娜 (中国科学技术信息研究所, 北京 100038)

# 一种上市企业风险披露量化方法<sup>\*</sup>

**摘要:** [目的/意义] 以一种更完善的测度方法度量上市企业年报文本中披露的风险信息。[方法/过程] 文章以百科词典为基础词库, 最大限度地抽取上市企业年报风险文本中的实体词, 使用文本向量化的方法将实体词转化为风险向量, 对企业年报风险文本中包含的风险信息进行度量, 以信息含量值量化企业年报中所包含的特有风险信息。[结果/结论] 通过对市场层面、行业层面的风险信息含量进行分析, 得出不同行业或企业风险信息披露的水平。

**关键词:** 风险信息; 风险披露; 文本向量化; 信息含量; 上市企业年报

## A Quantitative Method for Risk Disclosure of Listed Enterprises

**Abstract** [Purpose/significance] To measure the risk information disclosed in the annual reports of listed companies with a more perfect measurement method. [Method/process] Based on the Encyclopedia dictionary, this paper extracts the entity words from the risk text of listed companies' annual reports to the maximum extent, transforms the entity words into risk vectors by using text vectorization method, measures the risk information contained in the risk text of enterprise annual reports, and quantifies the unique risk information in enterprise annual reports by the value of information content. [Result/conclusion] Through the analysis of risk information content at market level and industry level, the level of risk information disclosure in different industries or enterprises is obtained.

**Keywords:** risk information; risk disclosure; text vectorization; information content; annual report of listed companies

企业年报描述了企业运营情况, 对投资者而言是了解企业发展状况的重要渠道。近年来年报中对于企业的文字性介绍在不断增加, 研究表明年报中的文字性介绍也包含着大量的重要信息, 同样影响着投资者的决策。年报中的财务数据反映的是历史性信息, 而包含的文本介绍则是在企业年度审计结束后编写的, 包含了更多企业未来发展状况的信息, 能够解释财务数据和弥补财务数据的不足, 结合财务数据和文本介绍能够更好地获取年报中包含的信息。对投资者而言, 企业未来的发展前景更能影响其决策, 从年报的文字性介绍获取企业的发展相关信息也越来越重要。

风险是企业发展过程中不可避免的重要因素, 风险的发现、披露以及应对都影响着企业未来的发展状况。企业年报中的风险信息是重要的非财务信息, 具有信息与风险的双重属性, 相较于财务数据, 年报中的风险披露包含着更多的展望信息, 更多地描述了企业未来发展中可能遇到的问题, 对企业来说, 算是“坏消息”。但对投资者和潜在投资者而言, 年报中的风险披露能够直接凸显出企业所

面临的状况, 揭示企业的相关管理策略, 降低和企业之间的信息不对称程度, 从而为其决策提供支持。

企业年报中风险信息的披露也是监管机构和投资者关注的问题之一, 并以此来研究上市企业的风险状况。当前对风险文本的研究多以内容分析法为主, 通过关键词和词频对文本进行量化。本文以上市企业年报的风险文本为研究对象, 通过风险文本向量化的方法, 以回归分析为媒介, 综合企业、行业、市场三个层面信息, 对企业年报风险文本进行量化, 从数量上对企业风险的披露情况进行分析, 继而全面测度证监会各门类、大类上市企业的风险信息披露情况。

## 1 国内外研究现状

风险信息披露是上市公司信息披露的重要元素, 国内外学者对其研究的热度也表明了其重要性。而风险的披露基本上都为文字性的定性信息, 所以对于风险信息的研究常使用处理描述性文本的分析方法, 比如使用关键词及词频来代表文本内容<sup>[1-2]</sup>, 或结合可视化技术挖掘文本深层次有价值的内容<sup>[3]</sup>。目前对年报中风险信息的研究主要在两个层面。一个是对企业风险信息披露行为的研究, 另一个是对风险信息处理方法的研究。

企业一方面希望通过风险信息的披露降低企业与外部

<sup>\*</sup> 本文为中国科学技术信息研究所重点工作项目“上市公司年报数据库建设及服务系统研发”的研究成果之一, 项目编号: ZD2019-09。

的信息不对称性,来增加投资者的信心;另一方面,又希望隐藏风险信息来降低风险披露的成本<sup>[4]</sup>。李慧雪等<sup>[5]</sup>利用风险披露等级和风险披露方式来衡量年报的风险信息披露程度。通过分析在风险披露方面存在的问题,从完善内部风险管理体制、加强政府监管和社会公众的监督、完善风险披露的内容和形式三个方面提出改善我国国有商业银行风险披露现状的方法。T. Kravet 和 V. Muslu<sup>[6]</sup>利用 OLS 变化模型来检验年报中新的风险披露的影响,研究表明年报中风险披露的增加与股票回报率增加以及申报前后的交易量有关,文本风险披露能够增加投资者的风险认知。P. Carmona 等<sup>[7]</sup>使用模糊集定性比较分析法建立上市公司财务与非财务数据库,研究表明董事会的独立性、规模、活动水平和性别多样性、首席执行官的兼任状况、审计委员会的独立性、四大审计公司的审计以及机构投资者的存在与高风险披露有关,说明了企业风险披露的重要性。

近年来,很多学者开始使用基于风险文本内容的分析方法来进行企业风险披露的研究。N. Moumen 等<sup>[8]</sup>使用内容分析法,将企业年报中的管理层讨论与分析(MD&A)与董事会报告(BR)中与风险相关句子的数量作为衡量风险披露水平的标准,证明了风险信息与市场未来两年的收益预测之间存在正向关系。LI 等<sup>[9]</sup>通过统计年报中与风险或不确定性相关词语的频率来衡量企业年报风险信息的披露水平。研究表明企业未来收益的降低与风险内容的增加有关,风险披露增加会对企业在下年的收益有负面的影响。Hanley 等<sup>[10]</sup>使用词内容分析法将 IPO 招股说明书中的信息分解为标准信息和信息含量来验证其与股票定价的关系,结果发现更加详细的风险披露能够更好地控制股票的发行价格。D. Zeghal 等<sup>[11]</sup>使用内容分析法验证企业风险披露范围和质量与企业盈利能力等相关特征的联系,说明了风险披露与危机、银行规模、董事会独立性等要素显著正相关,与盈利能力和董事会规模等要素呈显著负相关。孟庆斌等<sup>[12]</sup>采用文本向量化方法,对 A 股上市企业 MD&A 中披露的风险信息进行度量。研究表明 MD&A 中的信息含量越高,未来企业股价崩盘的风险越小。胡小荣等<sup>[13]</sup>采用基于多因素拟合的风险短语识别技术,将企业年报中“风险因素”描述字段进行处理,得到环保行业不同风险主题文本的主题短语。郝祥超等<sup>[14]</sup>利用文本分析法,以中文单个字为维度度量招股说明书中重大风险提示的标准风险提示信息和特有风险提示信息,研究表明主板上市公司的特有风险信息可以降低 IPO 抑价。

目前,越来越多的学者使用文本内容分析法来度量企业年报中的风险披露水平,以年报中的重点风险字、词或

者句代表其整体的表述,但存在以下问题:一是风险类句子和词语的判断主要依靠人工,比如统计与风险相关的句子或词语;二是采用文本向量方式进行风险信息量化时词典规模不足,无法识别未包含的风险词语;三是文本向量粒度太低,比如采用单字作为向量的分量,信息含量低,噪声大。所以本文在之前学者研究的基础上,采用千万级体量的百科实体词作为风险词遴选词库,极大地丰富了风险词汇的规模;同时采用风险短语而非单字作为分量来构建风险短语向量,以词为基础进行上市企业风险信息披露情况的量化,弥补了以往学者在相关研究中的不足。

## 2 基于文本向量化的风险信息提取方法

### 2.1 方法流程

在风险信息的抽取时,传统的方法是利用文本中有代表风险的关键词及词频来衡量风险信息,但这样做只能抽取最具有代表性的词,对于那些有意义但是不具有代表性的词则被舍弃。本文所使用的方法能最大限度地保留风险字段中的实体词,并依据其频次代表其重要性,能够挖掘出其中不同于其他企业的独特表述继而测度其风险信息。

本文将年报中上市企业风险的文本描述作为实验数据,利用百科实体词库从年报风险文本中抽取实体词及其频次,利用向量的方式将上市企业风险信息表述出来。之后计算出企业对应的行业标准化向量和市场标准化向量,以及计算出企业的信息含量。本文所使用的文本向量化的风险信息提取方法流程如图 1 所示。



图 1 风险信息提取流程

### 2.2 基于百科词典的风险短语识别方法

孟庆斌等<sup>[12]</sup>进行相似实验构建全集向量时,是以中国社会科学院语言研究所编纂的《现代汉语词典》作为基本的语料库,包括实体词 6.9 万个。为了最大限度抽取文本中的风险实体词,本文使用复旦大学提供的中文通用百科实体词库作为基本语料库,该词库包括 1200 多万个百科实体词,几乎覆盖所有领域,抽取精度高。为了提

高抽取效率,本文使用 Python 的中文分词组件 Jieba 进行分词,用百科实体词库替换 Jieba 自带的原词库,相较于原 Jieba 的 35 万词库规模提升了几十倍。此外, Jieba 中采用了基于汉字成词能力的隐马尔科夫模型 (HMM),用于识别词库中未包含的词,所以本文在分词时去除此模型,这样就能保证过滤掉结果中的单个字后所有的实体词都来自于百科实体词库。

### 2.3 企业文本向量标准化方法

要对文本进行量化,首先需要进行文本特征抽取并转化为文本向量,转化过程包括文本向量化和向量标准化两个步骤。首先利用百科实体库作为语料库构建风险信息全集向量,从全部风险文本中提取实体词,实体词的数量就等于全集向量的维度。然后,将每家企业的风险文本转化为向量,向量中每个维度代表一个实体词,数字表示该词出现的频次,为 0 表示该词没有在此企业的风险文本中出现。下面举例说明文本向量化的过程:设百科实体词库为 [产品、原材料、价格、上涨、供应不足……],两个简化的企业年报风险文本中,内容分别为“产品价格上涨,原材料价格上涨”和“原材料供应不足”,则全集向量为 [产品、原材料、价格、上涨、供应不足]。在第一个风险文本中,“产品”“原材料”“价格”“上涨”的出现次数分别为 1、1、2、2,而“供应不足”为 0,所以第一家企业的风险向量表示为 [1 1 2 2 0],第二家企业的风险向量表示为 [0 1 0 0 1]。

文本向量化之后,还可能存在文本不均衡的情况。即长文本中,某个词出现的概率高,在短文本中,出现的概率小。所以在文本向量化之后,需要将向量进行标准化处理,即将向量除以向量中各维度的和,也就是文本中实体词的总数。文本向量标准化过程如图 2 所示。

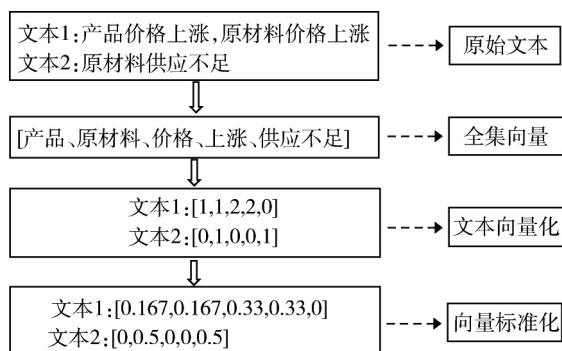


图2 文本向量标准化过程图

### 2.4 信息含量测度方法

本文借鉴参考文献 [7] 叙述的方法,从行业和市场两个层面来定义企业的信息含量。对于同行业的企业而言,在同一行业意味着面临着一些相似的风险,如市场竞

争风险、技术风险;而在同一市场中,这些企业也面临着相同的宏观经济环境以及政策背景,所以在风险描述方面每家企业不可避免地和其他企业存在着一定的相似性。在本文中,将这些与同行业或整个市场中的企业重复或相似的风险信息定义为标准信息,将那些企业自身独有的内容称为有价值的信息,定义为信息含量。

2.4.1 同行业标准化向量和市场标准化向量构建 各家企业风险描述不可避免的与其他企业存在着相似性,但是与其他行业中的企业相比,行业之间企业对风险描述的相似性会更高。首先利用 2.3 节所描述方法求出每个企业的标准化向量,然后计算每家企业对应的同行业标准化向量和市场标准化向量。同行业标准化向量指的是与企业  $i$  同行业的其他企业的标准化向量的平均值,定义为行业标准化向量  $Vector_{I_i}$ ,计算过程如公式 (1) 所示;市场标准化向量指的是企业  $i$  所处行业之外的所有企业的平均标准化向量,定义为市场标准化向量  $Vector_{M_i}$ ,计算过程如公式 (2) 所示。

$$Vector_{I_i} = \frac{1}{X-1} \sum_{j=1, j \neq i}^X Vector_{j_i} \quad (1)$$

$$Vector_{M_i} = \frac{1}{Y-X} \sum_{j=1, j \neq i}^{Y-X} Vector_{j_i} \quad (2)$$

2.4.2 信息含量测度方法 得到每家企业的标准化向量和行业标准化向量以及市场标准化向量之后,利用回归方程计算企业的信息含量  $Inf$  和信息标准  $Stdinf$ 。例如某年度企业  $i$  所处行业共有  $X$  家企业,整个市场共有  $Y$  家企业,则该企业的行业标准化向量  $Vector_{I_i}$  和市场标准化向量  $Vector_{M_i}$  计算公式如公式 (1) 和公式 (2) 所示,信息含量  $Inf$  和标准信息  $Stdinf$  的计算过程如公式 (3) 所示。

$$Vector_{i_i} = \alpha_0 + \alpha_1 \times Vector_{I_i} + \alpha_2 \times Vector_{M_i} + \mu_{i_i} \quad (3)$$

在公式 (3) 中,  $\alpha_1$  代表企业  $i$  风险文本中能够被同行业的其他企业所能解释的部分,  $\alpha_2$  代表企业  $i$  风险文本中能够被市场其他企业所能解释的部分。  $\alpha_1$  和  $\alpha_2$  越大越说明企业  $i$  的风险信息和其他企业的越相似,该企业风险文本中的信息越能被其他企业所解释。参考 Hanley 和 Hoberg (2010),将  $\alpha_1 + \alpha_2$  作为企业的标准信息,记作  $Stdinf_{i_i}$ ,则标准信息  $Stdinf_{i_i}$  的计算过程如公式 (4) 所示。

$$Stdinf_{i_i} = \alpha_1 + \alpha_2 \quad (4)$$

方程中的残差  $\mu_{i_i}$  表示用行业标准化向量和市场标准化向量估计企业标准向量的正面和负面的偏差,为了避免标准信息和信息含量之间的完全共线性导致的计算结果出现误差,将回归方程的残差  $\mu_{i_i}$  各维度绝对值之和作为企业的信息含量,记作  $Inf_{i_i}$ 。信息含量  $Inf$  的计算过程如公

式 (5) 所示。

$$Inf_{i,t} = \sum |\mu_{i,t}| \tag{5}$$

3 实验与结果分析

3.1 数据源及预处理

相较于之前研究者的实验数据，本文采用的实验数据来自于中国科学技术信息研究所自主加工的中国上市企业年报数据库，年报中的风险披露内容依据风险类型进行加工，因此数据质量更高，更有价值。本文从上市公司年报数据库中选取 A 股 1139 家上市公司 2014—2016 年半年报中的“风险字段”作为实验样本进行处理分析。在数据预处理方面，将 1139 家企业按照其门类行业进行行业划分，为了使实验结果更加准确，将只包含一家企业的行业去除。

3.2 信息含量及标准信息统计结果

实验结果各年份变量的描述性统计如表 1 所示，其中 Inf 表示企业的信息含量（见公式（5）），Stdinf 表示企业的标准信息（见公式（4））。为了避免某些异常值影响结果，在进行描述性统计的时候，进行了 1% 和 99% 水平上的缩尾处理。从表 1 中可以看到，无论是信息含量还是标准信息，各年份的最大值和最小值都有较大差异，信息含量值高于标准信息值，为本文后续的分析提供了支撑。

表 1 统计结果

变量	年份	样本量	均值	标准差	中位数	最大值	最小值
Inf (信息含量)	2014	1139	1.89	0.36	1.79	3.31	1.41
	2015	1139	1.86	0.33	1.78	3.11	1.42
	2016	1139	1.86	0.33	1.79	3.25	1.35
Stdinf (标准信息)	2014	1139	0.99	0.32	0.98	2.06	0.35
	2015	1139	0.98	0.30	0.97	1.86	0.36
	2016	1139	1.01	0.28	0.99	1.86	0.38

3.3 证监会门类和大类行业分析

3.3.1 证监会门类行业分析 表 2 为 2014—2016 年证监会的门类行业各年的信息含量均值（企业数小于 10 的行业被去除）。由表 2 可以看出，三年信息含量均值最高的行业为电力、热力、燃气及水生产和供应业，其次为制造业和房地产业，而信息含量均值最低的为科学研究和技术服务业。结果也与实际情况相符合，电力、热力、燃气及水生产和供应业因为原材料价格以及需求客户变动较大，制造业和房地产因为竞争强，且与政策及大经济环境相关性强，所以其半年报中风险信息含量较高。对于科学研究和技术服务业等行业，因其发展较为平缓且不确定性较弱，所以在半年报中，风险披露的内容较少。

此外，房地产业、建筑业、制造业、批发和零售业、电力热力燃气及水生产和供应业 5 个行业各年份的平均信息含量值都高于市场每年份平均信息含量值，说明这几个

行业整体的风险披露水平较高。而水利环境和公共设施管理业、租赁和商务服务业、文化体育和娱乐业、农林牧渔业、信息传输软件和信息技术服务业 5 个行业各年份的平均信息含量值都低于市场每年份平均信息含量值，说明其风险披露水平还有待提升。

表 2 各行业各年份平均信息含量值

证监会行业门类	上市企业数	信息含量			
		2014 年	2015 年	2016 年	三年平均值
市场	1139	1.89	1.86	1.86	1.87
电力、热力、燃气及水生产和供应业	25	2.02	1.94	1.97	1.98
房地产业	51	1.95	1.90	1.90	1.92
批发和零售业	67	1.93	1.88	1.91	1.91
制造业	702	1.92	1.88	1.88	1.89
建筑业	29	1.87	1.90	1.90	1.89
采矿业	43	1.83	1.84	1.93	1.87
金融业	23	1.75	2.00	1.84	1.86
交通运输、仓储和邮政业	32	1.91	1.83	1.84	1.86
综合	10	1.89	1.79	1.79	1.82
农、林、牧、渔业	15	1.79	1.80	1.81	1.80
文化、体育和娱乐业	12	1.77	1.82	1.80	1.80
水利、环境和公共设施管理业	12	1.74	1.73	1.83	1.77
租赁和商务服务业	13	1.78	1.71	1.74	1.74
信息传输、软件和信息技术服务业	88	1.72	1.69	1.65	1.69

通过统计各行业中信息含量值高于市场平均信息含量的企业数发现（结果如图 3 所示，其中企业数少于 5 的行业被去除），虽然农、林、牧、渔业每年的平均信息含量值都低于市场平均信息含量值，但是农、林、牧、渔业中有 26.7% 的企业三年的信息含量值都高于市场每年的平均信息含量值。这显著说明了这几家企业与同行业企业相比，风险信息披露水平更高，对于其他企业而言，具有参考的价值。其次是电力、热力、燃气及水生产和供应业、制造业和金融业，占比分别为 20%、17.5% 和 17.4%。这不仅说明这几个行业风险信息披露水平整体较高，还说明这些行业中其部分企业风险信息披露远高于市场平均信息含量值。

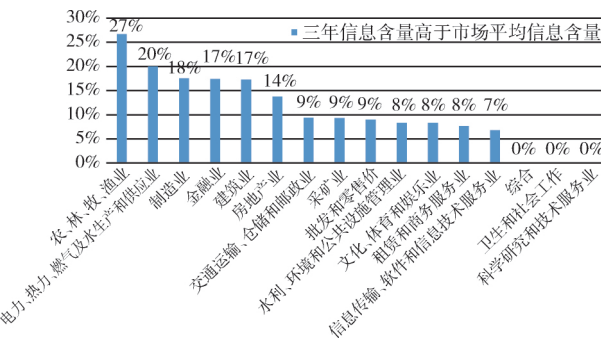


图 3 各行业高于市场平均信息含量企业数占比

3.3.2 制造业大类行业分析 由表 2 可知，每年份制造



业的平均信息含量明显高于市场的平均信息含量，而且制造业所包含的企业数量也最多，约占市场企业总数的61.63%。本节将以制造业为代表，分析制造业行业中大类行业信息含量分布情况。通过大类行业的划分，制造业可以划分为医药制造业等29个大类行业，制造业各大类行业各年的平均信息含量值如表3所示（企业数小于5的行业被去除）。由表3可知，在制造业的28个大类行业中，汽车制造业、纺织业、石油加工炼焦及核燃料加工业、黑色金属冶炼及压延加工、酒饮料和精制茶制造业、化学原料及化学制品制造业、化学纤维制造业各年的平均信息含量都高于制造业各年平均信息含量，这说明汽车制造业等7个大类行业的风险披露水平较高。

表3 制造业大类行业各年份平均信息含量值					
证监会制造业大类行业	上市企业数	信息含量			
		2014	2015	2016	三年平均值
制造业	702	1.92	1.88	1.88	1.89
石油加工、炼焦及核燃料加工业	9	2.09	2.23	1.91	2.08
汽车制造业	24	2.07	2.03	2.03	2.04
纺织业	14	1.92	2.13	2.00	2.02
黑色金属冶炼及压延加工	15	2.09	1.99	1.89	1.99
化学原料及化学制品制造业	70	2.08	1.97	1.91	1.99
酒、饮料和精制茶制造业	9	1.97	1.97	1.98	1.97
造纸及纸制品业	7	2.03	1.99	1.86	1.96
其他制造业	5	2.10	1.81	1.94	1.95
金属制品业	14	1.96	2.02	1.80	1.93
纺织服装、服饰业	11	1.97	1.86	1.95	1.93
化学纤维制造业	11	1.95	1.92	1.90	1.92
医药制造业	70	1.91	1.88	1.95	1.92
农副食品加工业	12	2.02	1.86	1.86	1.91
有色金属冶炼及压延加工	30	1.95	1.87	1.89	1.90
铁路、船舶、航空航天和其他运输设备制造业	14	1.92	1.99	1.80	1.90
通用设备制造业	47	1.90	1.87	1.91	1.89
橡胶和塑料制品业	25	1.94	1.81	1.92	1.89
食品制造业	10	1.79	1.77	2.01	1.86
非金属矿物制品业	37	1.95	1.81	1.80	1.85
计算机、通信和其他电子设备制造业	98	1.87	1.82	1.84	1.84
专用设备制造业	70	1.80	1.83	1.87	1.83
电气机械及器材制造业	71	1.79	1.78	1.82	1.80
仪器仪表制造业	18	1.77	1.81	1.66	1.75

图4为各大类行业中每年信息含量值都高于制造业平均信息含量值的企业数占比（企业数少于5的大类行业被去除），由图4可知石油加工、炼焦及核燃料加工业有44.4%的企业三年的信息含量值都高于制造业每年的平均风险信息含量值，其次是造纸和纸制品业和汽车制造业，占比分别为29%和25%。三个行业中汽车制造业企业数最多，竞争最为激烈，但其整体平均信息含量值高于制造业平均信息含量，四分之一企业每年的信息都高于制造业平均信息含量，更高于市场整体信息含量，说明了汽车制造业在风险信息披露方面做得较好。其中化学原料及化学制品制造业中有15家企业连续三年的信息含量值都高

于制造业每年的平均风险信息含量值，是所有大类行业中数量最多的，也说明其风险信息披露意识较强。

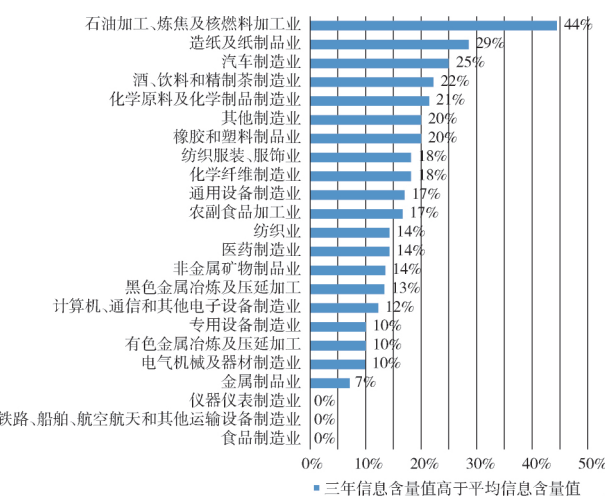


图4 各大类行业高于制造业平均信息含量值企业数占比

3.4 改善风险信息披露水平建议

根据以上分析，本文针对改善企业风险信息披露水平提出两点建议：一是证监会应根据行业来制定有差异的、针对性更强的信息披露规范，这样可以使整体的披露质量得到提升，可以使社会整体的福利水平得到提高。由以上的分析可知，上市企业的风险信息披露可能在不同的行业存在一定的差异。同行业公司特征比较接近，彼此的价值相关程度也很高，而不同行业门类间因产业、发展、政策的不同在进行披露时策略会有区别，国内外研究发现，像制造业、服务业等行业愿意披露更多的信息，所以其行业的信息含量值也正如本文所度量的值一样较高。

二是根据自愿性披露的行业差异，优化基于强制性披露的监管。强制性披露能够保护投资者的利益，而自愿性披露则是满足投资者的需求，那么对于不同行业来说其需求也不尽相同。比如高技术含量的行业企业自愿性披露程度较高，因为对于高技术企业而言，技术发展较为复杂，这也使得市场和投资者对于信息披露的需求较高。根据自愿性披露的行业差异，可以对不同的行业采取差异化的风险信息披露监管。通过相关手段，提高信息公开程度，降低投资者获得信息的成本，间接促进整体市场的自愿性披露水平。例如证监会发布的《公开发行证券公司信息披露编报规则第11号》，根据行业制定差异化的监管要求。

4 结论及不足

本文介绍了一种上市企业风险披露量化方法，以百科实体库为基础词库，用文本向量化的方法对2014—2016年A股上市企业年报中风险描述文本进行实体词的抽取，

构建文本向量并以此计算企业风险信息含量和标准信息。相较于之前研究者所使用的量化企业风险披露的方法,本文使用的方法能最大限度地抽取风险文本中有价值的实体词,但仍要进一步对噪声词进行处理。通过对市场层面、行业层面的风险信息含量进行分析,得出不同行业或企业风险信息披露的水平。分析结果能够向关心上市企业发展的利益相关者(如投资者、供应商)提供企业风险披露情况,另外监管机构也能够根据不同行业风险信息值评估其风险信息披露质量,根据披露情况制定相关政策以提高市场整体风险披露水平。□

# 参考文献

- [1] 刘琦岩,张泽玉,张运良,等.从主题词词频变化看我国科技政策演变[J].情报工程,2016,2(4):14-19.
- [2] 靳彤,张红伟,赵勇.政策文本计量视角下我国科研诚信治理的特征与启示[J].情报工程,2018,4(5):116-126.
- [3] 郭传斌,刘琦岩,赵婧,等.情报学视角下的文本可视化应用[J].情报工程,2017,3(4):48-61.
- [4] 陈艺丹.风险信息披露文献综述[J].财讯,2018(31):146-147.
- [5] 尚洪涛,李慧雪.我国国有商业银行流动性风险披露存在的问题与对策[J].经济论坛,2009(1):44-46.
- [6] KRAVET T, MUSLU V. Textual risk disclosures and investors' risk perceptions [J]. Review of Accounting Studies, 2013, 18(4): 1088-1122.
- [7] PEDRO C, DE F C, CARMEN R. Risk disclosure analysis in the corporate governance annual report using fuzzy-set qualitative comparative analysis [J]. Revista de Administração de Empresas, 2016, 56(3): 342-352.
- [8] MOUMEN N, OTHMAN H B, HUSSAINEY K. The value relevance of risk disclosure in annual reports: evidence from MENA emerging markets [J]. Research in International Business and Finance, 2015, 34: 177-204.
- [9] LI F, DECHOW P, DICHEV I, et al. Do stock market investors understand the risk sentiment of corporate annual reports? [R]. University of Michigan, 2006.
- [10] HOBERG G, HANLEY K W. The information content of IPO prospectuses [J]. The Review of Financial Studies, 2010, 23(7): 2821-2864.
- [11] ZEGHAL D, AOUN M E. The effect of the 2007/2008 financial crisis on enterprise risk management disclosure of top US banks [J]. Journal of Modern Accounting and Auditing, 2016, 12(1): 28-51.
- [12] 孟庆斌,杨俊华,鲁冰.管理层讨论与分析披露的信息含量与股价崩盘风险——基于文本向量化方法的研究[J].中国工业经济,2017(12):132-150.
- [13] 胡小荣,姚长青,高影繁.基于风险短语自动抽取的上市公司风险识别方法及可视化研究[J].情报学报,2017,36(7):663-668.
- [14] 郝项超,苏之翔.重大风险提示可以降低IPO抑价吗?——基于文本分析法的经验证据[J].财经研究,2014,40(5):42-53.

作者简介:牛高鸣(通讯作者),男,1996年生,硕士生。研究方向:科技金融数据挖掘。姚长青,男,1974年生,博士,研究员。研究方向:情报理论与方法。高影繁,女,1974年生,博士,副研究员。研究方向:文本挖掘,知识组织。梁娜,女,1995年生,硕士生。研究方向:科技大数据关键技术与应用服务研究。

作者贡献声明:牛高鸣,撰写论文,分析数据。姚长青,指导论文写作与修改。高影繁,构建论文整体框架。梁娜,分析方法指导。

录用日期:2019-10-25

(上接第136页)

- [27] BARABASI A, ALBERT R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512.
- [28] DYER J R G, JOHANSSON A, HELBING D, et al. Leadership, consensus decision making and collective behaviour in humans [J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2009, 364(1518): 781-789.
- [29] WU L, ZHANG J. Accelerating growth and size-dependent distribution of human online activities [J]. Physical Review E, 2011, 84(2): 026113.
- [30] LORENZ J, RAUHUT H, HELBING S D. How social influence can undermine the wisdom of crowd effect [J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(22): 9020-9025.

作者简介:项欣(ORCID:0000-0002-6623-6695),男,1982年生,博士生。研究方向:信息分析。祁彬斌(ORCID:0000-0002-4001-5712),男,1992年生,博士生。研究方向:虚拟现实与仿真,信息可视化技术及应用,信息分析。朱学芳(ORCID:0000-0002-8244-5999,通讯作者),男,1961年生,博士,教授,博士生导师。研究方向:多媒体信息处理及应用,数字信息资源管理,信息可视化技术及应用,模式识别与人工智能,信息安全,虚拟现实,数字人文等。

作者贡献声明:项欣,提出研究问题,设计研究思路,数据处理,撰写和修改论文。祁彬斌,修改论文。朱学芳,理论指导。

录用日期:2020-01-21