基于知识图谱的企业智能风险预警系统

7月8日汇报PPT讲稿

各位领导、老师、专家们上午好！

今天很荣幸为大家讲解和演示我们团队制作的基于知识图谱的企业智能风控系统。

首先说明一下软件的设计背景和意义。舆情监测服务在网络上定义为一种数据抓取，数据分析，数据实时监控分析的企业级服务。其中关键就在于数据以及实时两个方面，一方面要及时的获取到最新数据源，另一方面要对新闻数据做实时处理，实时分析，实时反馈。

这里我总结了大致三类传统的舆情预警系统。

第一种是基于关键词的舆情预警系统，这种系统会一直处于监控网络中出现的新闻的状态，一旦检测到的新闻信息中触发了相关的监测关键词，系统就会发出警告通知，从而实现预警的功能。

第二种是基于情感倾向类的舆情预警系统。这种系统通俗意义上来说就是对于获取到的网络上的新闻做情感分析，当分析得到的负面情绪超过一定比例后，系统会发出警告通知，从而实现预警的功能。

第三种是基于指标的预警系统。这种系统常用于事件发生一段时间之后的监控，主要是根据舆情的传播量，转载情况等指标进行预警，可用于追踪舆情话题的发展态势。

这三种舆情预警系统都还有一定的缺陷，有改进的空间，

如基于关键词的舆情预警就需要事先由专业人员给出相应的关键词并组成关键词词库，而词库的丰富度直接决定了预警系统是否可以准确的判断事件的风险类型。当出现某些未知类型的事件此系统就无法给出正确的判断。

第二种基于情感倾向类的舆情预警系统需要分析出事件中的情感倾向，但是在金融证券行业，大多数的新闻表述都是客观角度出发，这就导致很难获取到其中的情感特征。因此此类系统在企业新闻舆情分析方面效果并不明显。

第三种基于指标的舆情预警系统需要根据新闻的传播度、转发量来判断事件的风险特征，这显然有悖于风险监控的实时性的要求，对于企业的有关新闻应该第一时间对其作出判断并作出相应的处理。

基于此，我们团队结合了最新的对于文本事件的处理结束，提出了基于知识图谱的企业智能风险预警系统的构想，并实现了其功能。

我们的系统结合了

知识图谱，可实现多企业的多径推理扩展，知识图谱可以关联企业内部信息以及企业与企业间的关联信息，形象化展示企业间的关联信息。其次当获取到一条风险新闻后由关联的单一企业可在知识图谱中进行逻辑推理，快速定位多个相关联企业。提高业务人员的分析效率。

其次，在对于新闻的风险判断中，我们引入了机器学习，深度学习可通过预训练获得高效、精准的判别模型，在后期新闻事件分析判断中可实现判别精准度高、分析覆盖面广、实时计算度快的功能。

最后，为了应对互联网上数据量爆炸的新闻事件，引入了大数据技术，Hadoop集群可以实现大数据的存储和计算，利用spark内存级别的快速读写技术可实现海量数据实时存储及读取。对于获取到的新闻数据可进行多路并行处理来提高处理新闻事件的单一时间，实现实时化处理。

下面为大家介绍我们系统的功能和实现。

第一部分是有关知识图谱的搭建以及知识图谱的可视化功能演示。

构建企业知识图谱需要获得企业的相关信息，通过互联网爬虫或者是购买数据库内容的方式获取到基本的有关企业的文本信息及结构化信息。企业知识图谱主要是描述企业的内部结构信息以及由企业与企业间关联的关联信息。其中数据信息主要包括公司介绍、董监高、股东、债券、投资、行业、概念及主营业务等。

第二步是筛选完整数据信息，获取相互之间可利用可展示的关联、隶属关系，通过关联分析发掘存在大量数据集的对象之间的相关性及因果结构关系。

然后通过整理获取到的结构化信息组成三元组形式并利用编程语言将数据导入到图数据库中生成企业主体的知识图谱。

最后通过可视化插件将数据内容可视化到网页端并提供交互的功能。

普通用户在网页端就可以实现对于一个企业或者多个企业的图谱建立，用户只需要在网页端上传规定格式的excel文件，系统就会自动对文件进行解析并构建相应企业的知识图谱。右图为构建的企业图谱示例，中心主节点为企业主体，由企业主体向外发散延伸得到企业关联信息节点。

用户可在企业名称搜索框输入企业名，系统会根据模糊搜索得到相应的企业图谱展示在网页端。

除了可以以文件信息上传构建图谱外，系统在网页界面提供了小范围增删改查操作。

下面由我们团队的另外一名同学为大家演示知识图谱这块的操作。

好的，接下来我为大家介绍系统的另一部分内容，也是此系统的关键部分，对于新闻事件的处理及分析。

首先，实时新闻获取系统通过在互联网上的监控，其中主要集中在几大特定金融行业新闻事件发布网站。实时获取相关最新新闻信息并将其获取并存储到Hadoop数据库中进行存储。

第二阶段，新闻信息提取系统将存储在Hadoop中的新闻信息逐一提取到自然语言处理系统中进行事件分析和判断。NLP处理阶段通过Bert深度神经网络的预训练获得相应模型。模型可对新闻中出现的相关企业、证券、股票或行业进行关键词向量匹配进行提取。其次，对新闻整体内容进行切词分词处理，利用Bert技术做词向量向量间相似度计算，得出新闻事件的风险类型。

第三阶段，对于分析后的结果以信息流的形式存储到Hadoop数据系统中，等待下一步信息提取。最后由舆情信息发布系统将舆情信息发布到网页端，在网页端工作的业务人员就可以第一时间了解到风险舆情并作出相应的处理。

这里图中是存储在数据库中的新闻事件信息。由新闻事件获取管理系统负责控制。与后端事件分析系统分隔开可实现多路并行处理，提高处理效率。

图中展示的是已经分析处理好的舆情信息由舆情信息发布系统发布到网页端展示。可以看到，结果中包含了风险类型，相关企业，企业综合评估等信息。

这里显示的是网页上主界面实时滚动的新闻事件标题。实时滚动新闻从Hadoop数据库中获取，获取到的新闻均与所关注企业有关，无关新闻将会被删除。其中红色的代表有风险的事件，灰色的代表无风险事件。

页面左边是某一舆情信息分析处理的结果在网页端展示的界面，可以看到，结果信息中包含了，当前新闻的标题，与之相关的企业。企业的周期内的综合评分，事件的风险类型、事件严重等级以及事件可能持续影响的时间。

右边是新闻事件分线时系统给出的在不同风险类型中占据的权值比重，因为一条新闻不可能只反映一种风险类型，通过这个图可以直观得到新闻事件的不同类型的风险。业务人员可以根据此图作出更准确的判断和处理。

这里展示的是知识图谱与舆情分析系统的结合。上面是知识图谱的构建路线，下面是深度学习新闻舆情分析判别路线。两者在获取到分析结果后结合。根据分析所得结果判断相关企业或行业的风险类型。其次，在知识图谱中以所获得的企业或者行业进行扩展推理得到与之相关联的其他上下游企业。这样就可以通过一条新闻事件得到多个企业的市场风险值。

最后对项目做简单的总结。

1. 软件设计实现了基于知识图谱的企业智能预警功能。
2. 为企业搭建的知识图谱可以更好的帮助企业了解自身以及与其他企业之间的关联关系
3. 为企业搭建的知识图谱可以更好的帮助企业了解自身以及与其他企业之间的关联关系

功能展示

企业图谱展示

新闻舆情展示

数据应用

数据处理

数据获取

企业信息整理

NLP及深度学习

企业图谱构建

新闻数据信息存储分发

网络新闻数据

网络企业信息

数据库企业信息