

面向产学研服务的政策匹配方法与系统^①



罗 扬¹, 颜瑞新², 李伟刚², 朱登明^{3,4}

¹(江苏省生产力促进中心, 南京 210042)

²(西北工业大学, 西安 710072)

³(太仓中科信息技术研究院, 太仓 215400)

⁴(中国科学院 计算技术研究所, 北京 100190)

通信作者: 颜瑞新, E-mail: 13203992171@163.com

摘 要: 近年来, 产学研合作成为推动产业升级和经济发展的重要因素, 产学研服务可以快速整合多方资源, 提高创新效率, 增强企业的综合竞争力, 因此各级政府也围绕产学研合作推出了各种支持政策. 然而各类企业, 作为产学研合作的重要一环, 在申报过程中难以快速收集、整理与其相关的支持政策. 因此, 我们面向产学研服务, 利用人工智能文本分析的技术, 设计和实现了政策匹配系统, 可以对各类政策进行分析和预处理, 使得企业在政策池中能快速地获取与自身条件相匹配的政策, 从而节省企业的人力, 提高企业申报政策的效率.

关键词: 产学研; 政策匹配; 人工智能

引用格式: 罗扬, 颜瑞新, 李伟刚, 朱登明. 面向产学研服务的政策匹配方法与系统. 计算机系统应用, 2022, 31(11): 139-147. <http://www.c-s-a.org.cn/1003-3254/8794.html>

Policy Matching Method and System for Industry-university-research Service

LUO Yang¹, YAN Rui-Xin², LI Wei-Gang², ZHU Deng-Ming^{3,4}

¹(Productivity Centre of Jiangsu Province, Nanjing 210042, China)

²(Northwestern Polytechnical University, Xi'an 710072, China)

³(Taicang-CAS Institute of Information and Technology, Taicang 215400, China)

⁴(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: In recent years, industry-university-research cooperation has become an important factor promoting industrial upgrading and economic development. Industry-university-research services are conducive to quickly integrating multiple resources, improving innovation efficiency, and enhancing the comprehensive competitiveness of enterprises. Governments at all levels have also launched various policies to support industry-university-research cooperation. However, a wide variety of enterprises, as an important part of such cooperation, find it difficult to collect and organize befitting supporting policies efficiently during the application process. Therefore, this study employs the technology of artificial intelligence-based text analysis to design and implement a policy matching system for industry-university-research services. This system can analyze and preprocess various policies from different sources, thereby enabling enterprises to quickly obtain supporting policies that match their specific conditions and ultimately saving manpower for the enterprises and improving their efficiency in applying for a project.

Key words: industry-university-research cooperation; policy matching; artificial intelligence

随着我国互联网技术的飞速发展, 信息数字化的交流逐步成为新型发展趋势, 因此许多地方政府部门

与组织都开始尝试通过打造线上产学研合作平台来推动当地产学研工作的发展, 这已成为社会经济快速发

^① 基金项目: 江苏省创新能力建设项目 (BM2018018)

收稿时间: 2022-03-06; 修改时间: 2022-04-02; 采用时间: 2022-04-12; csa 在线出版时间: 2022-07-14

展的重要推动力之一。产学研合作整合了政府、企业以及高校院所等各方优势资源,各个单位主体可以优势互补,使得利益最大化,风险最小化^[1]。相关研究表明,提升科技成果的转化率,需要进行多方面的综合管理投入,如通过政策激励优化科技成果的转化机制、加大各方对科技研发的投入力度等^[2,3]。

各级政府为推动产学研合作推出了各种支持政策,希望借此推动资源整合,提高创新效率,增强各方的综合竞争力。然而,由于各种支持政策的类型不同、发布时间不同和所属管理部门不同,使得这些政策分布非常零散。对于企业来说,需要花费大量的时间和精力来搜索政策来源、验证政策时效性、评估申报可行性等,这不利于各类创新主体及时全面了解政策信息,也不便于其从海量政策中筛选出符合其自身申报的政策,不能充分有效地做好项目规划和申报准备,使得政策效用难以最大化。

结合产学研工作开展的实际需要,当前很多地方政府搭建了不同形式的产学研合作平台,在这些平台上的数据来源和数据类型各种各样,其中就包括各种支持政策,但当前大部分平台所采用的传统展示型的组织和呈现方式难以实现政策数据的充分利用^[4,5]。因此,应用文本分析技术研究政策文本的分析和政策匹配,在这一背景下具有十分重要的意义。

为了能够解决当前产学研合作系统存在的局限,帮助企业快速匹配适用的产业科技政策,本文研究了一种基于 BERT 的政策属性抽取方法,设计并实现了支持政策匹配的系统,能根据政策条件和录入的企业信息,自动计算政策匹配度,企业通过简单的操作,就可以获得与自身条件相匹配的科技政策。通过智能政策匹配技术和系统,进一步优化了产学研平台的业务功能,提高了政策信息的流通效率,优化了资源配置,提升了产学研合作的管理服务水平。

1 相关研究

所谓的政策文本是指因政策活动而产生的一些文献文本记录,包括官方文献(如法律、法规、部门规章)、公文档案(如研究、咨询、听证或决议)、政策舆情文本^[6]。随着数据处理技术的快速发展,政策文本分析呈现出大样本、细粒度的趋势,这主要是依赖于大数据分析、自然语言处理、数据挖掘、知识图谱等技术的日渐成熟。因此,一系列模型和方法的应用在政

策计算分析领域产生了一系列新的理念,如政策文本计算、政策语义分析、政策情感分析等新的研究视角^[7,8]。

近几年,在政府公文文本分析领域,国内外的研究主要在使用数据挖掘技术探索政策深层的内容,从而不断拓展政策研究的广度与深度。一些学者通过人工智能技术研究信息、特征提取,有力的支撑起政策的主题分析和热点识别等任务^[9]。赵洪等人^[10]提出了面向政府公文的知识发现与分析系统及相关技术,可实现政府公文的内容结构解析、主题自动标引、产生式自动摘要、重点内容提取等功能。邓文红针对政府办公信息系统智能化的迫切需要,研究了在知识管理下的公文自动分发、文件检索、数据抽取及报告自动生成问题^[11]。

对政策匹配系统的研究,目前陆川等人^[12]公开了一种政策大数据智能匹配系统及匹配方法,该专利的技术方案主要通过各类创新主体的浏览习惯与感兴趣勾选等给出相应的文章精准匹配度。李发明^[13]公开了一种用于企业服务的政策智能匹配系统的构建方法,其所述的政策匹配模型是政府工作人员根据政策条件去自定义配置匹配模型,每条政策通过配置的模型,和收集的企业信息进行匹配。上述的方法在收集政策方面相比纯手工方法有了一定的进展,但仍然存在智能化程度不高等问题,在政策匹配的效率方面还需进一步的提升。

2 政策匹配方法

本文将政策匹配问题分解为数据采集、政策的属性抽取和政策匹配度计算 3 方面的问题,采用了自然语言处理文本分析的方法,着重研究了政策文本的采集和预处理方法,并基于 BERT 模型对政策文本进行属性抽取,将该模型应用在政策匹配系统中,设计政策匹配度的计算方法,在系统中支持企业用户和政策条件的匹配度计算。

本文的研究内容框架如图 1 所示,主要分为以下几个部分:1) 数据预处理阶段,从网络上不同来源采集相关数据并进行清洗,处理成我们所需要的规范化的数据;2) 文本的属性条件抽取阶段,首先对文本句子进行词法和句法分析,利用依存句法分析方法,判断词的结构和类别,标记目标词的特征,确定句子的依存关系。然后再采用属性抽取方法处理政策文本,抽取政策中

的属性条件,从而得到按权重分类的条件池;3)政策文本匹配分值计算阶段,根据属性抽取的结果,基于条件池形成的条件模板,将用户填入信息与政策条件进行匹配,并根据每个条件的匹配情况打分;4)将模型用于政策匹配系统的设计与实现。

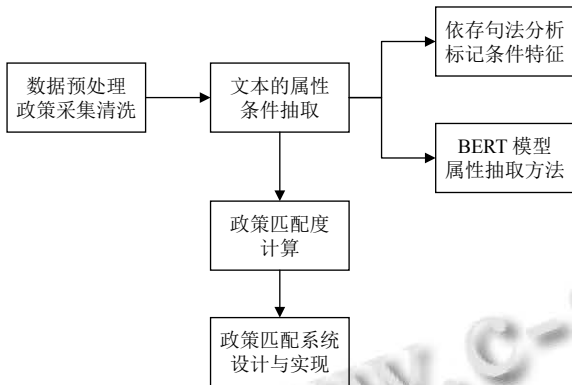


图1 本文研究内容框架

2.1 数据采集及预处理

政策发布来源于国家级、省级、市级、区级等不同的政府相关部门,这些数据信息的获取需要用到比较成熟的网络爬虫技术,每个Spider都可以通过定义特定网站的抓取和解析规则,自动地从网络上提取政策信息。本文系统使用Python语言,Requests网页请求工具及特定的正则表达式进行操作。网络爬虫动态跟踪政府发布的源网页,预设置为每日定时、自动爬取新政策。

政策数据具体的爬取流程如图2所示,首先在Requests工具中设置需要爬取的域名,Requests给目标网站发送请求。内部的搜索引擎会打开进行域名解析,并且可以得到一个URL链接。引擎在已知需要爬取的地址后,发送一个请求去调度相应的内容。然后将调度的内容和需要爬取的地址URL作为结果返回,含有这个网站页面所包含的数据,获得了爬虫所爬取的网络数据,包括下一个URL和相应的文本。接下来使用政策表达式解析Requests结果值中的文本数据和下一个URL。最后保存数据,然后给引擎发送新的请求开始新一轮的爬取操作,直到调度的请求全部响应结束断开连接。

通过网络爬虫技术获取了数据以后,再利用正则表达式简单高效的特点,使用政策表达式来匹配文本中的特殊字段,比如 `re.findall(<title>(.*)</title>)` 可以

查到政策标题, `re.findall(time="(.*?)"/>)` 找到政策的发布时间等。

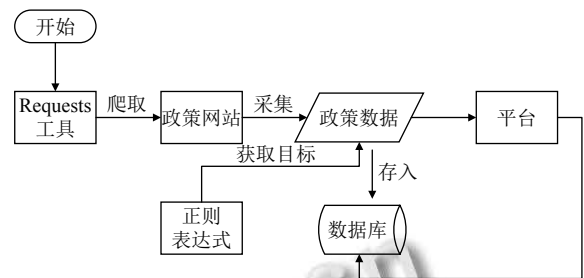


图2 政策采集流程图

2.2 基于BERT的政策属性抽取

在本文中,将一个政策表示为一组政策条件的集合 $z=\{c_1, c_2, \dots, c_n\}$, 一个政策条件进一步表示为三元组 $c=(s, p, o)$ 的形式, 其中 s 代表主实体 subject, 即是句子中的执行者; o 代表客实体 object, 即是所抽取属性的属性值; 而 p 表示主客体间的关系 predicate, 即是我们想要抽取的属性。即可以理解为 (s, p, o) 的意思是“主体 s 和客体 o 的关系是 p ”。如表1所示, 我们从典型的政策条件中抽取了对应的三元组信息, 比如, 企业 (s) 的注册资金 (p) 要大于 500 万元 (o)。

表1 政策条件提取三元组示例

条件ID	实体s	属性p	属性值o
①	企业	企业类型	科技、新能源、机械...
②	企业	注册地	苏州
③	单位	在职人员	大于100人
④	项目	投入资金	大于500万元
⑤	单位	经营时间	不少于2年

在政策属性抽取模型中, 借鉴基于文本分析的属性抽取方法^[14], 将政策中有效的、可以量化的关键属性提取出来。文本属性抽取的主要任务是从源文本文档中抽取既定目标的主体、客体、及两者之间关系。在此次任务中我们的目标是抽取政策文档中包含限制条件的句子, 进一步抽取句子中包含的主体实体(如企业、申报单位), 客体实体(如苏州市、200万元), 两者的关系(如: 注册地、研发投入)。我们将抽取信息三元组采用式(1)的方式建模, 先预测 s , 然后根据上一步传入的 s 去预测这个 s 所对应 p 和 o 。

$$P(s, p, o) = P(s)P(o|s)P(p|s, o) \quad (1)$$

我们通过构建并训练政策属性抽取的神经网络模型, 使得输入一个政策文本, 可以抽取该句子包含

的政策条件的三元组内容. 该方法主要分为以下 3 步: 首先, 对政策文本进行句法分析, 得到其句法结构^[15], 提取每个词的词性、实体类型、词位置、词在 WordNet 中的上位词标签等特征, 再采用字词混合的 embedding, 存储数据的语义信息; 其次, 使用 BERT 预训练模型对输入的数据进行编码, 采用“半指针-半标注”的结构方法预测 s 所对应的首、尾的位置; 最后, 将上一步预测的 s 对应的 embedding 与 BERT 输出的 hidden-states 进行拼接继续预测 p 和 o , 这里也对 p 构建相应的“半

指针-半标注”结构预测对应的 o 的首尾位置, 这样就可以把 p 和 o 同时预测出.

本文基于 BERT 的模型结构^[16], 其输入是文本的 embedding 求和, 分别是当前词的 token embedding, 当前词所在句子的 segment embedding, 当前词所在位置的 position embedding. 传入 BERT 模型的编码器得到编码序列, 然后将这个编码序列接入两个二分类器, 分类器的主要作用就是利用给定的类别对输入数据预测 s 的内容, 预测过程如图 3 所示.

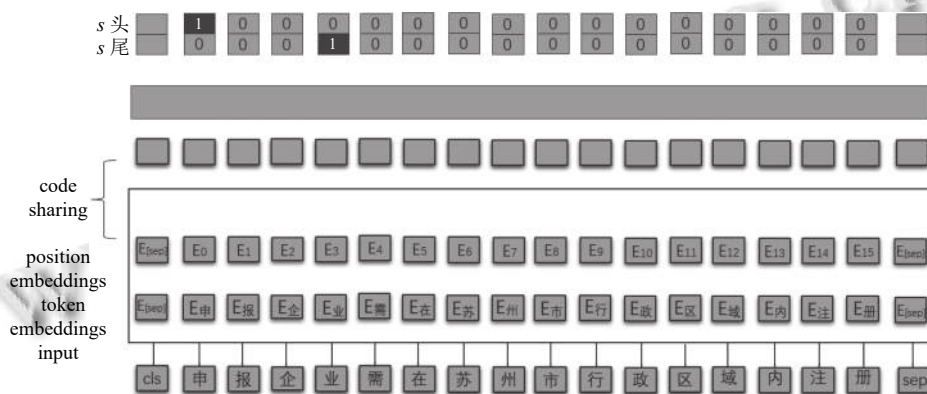


图 3 BERT 模型抽取 s 的结构表示

预测出了 s 以后, 找出 s 的首尾位置信息的编码, 对这个编码序列做 conditional layer normalization; 然后将上一步得到的条件序列去预测 s 所对应的 p , 对每一

种 p , 分别用同样的方式抽取其对应的 o . 这样经过模型的训练, s 、 p 、 o 就都可以完成预测, 预测过程如图 4 所示.

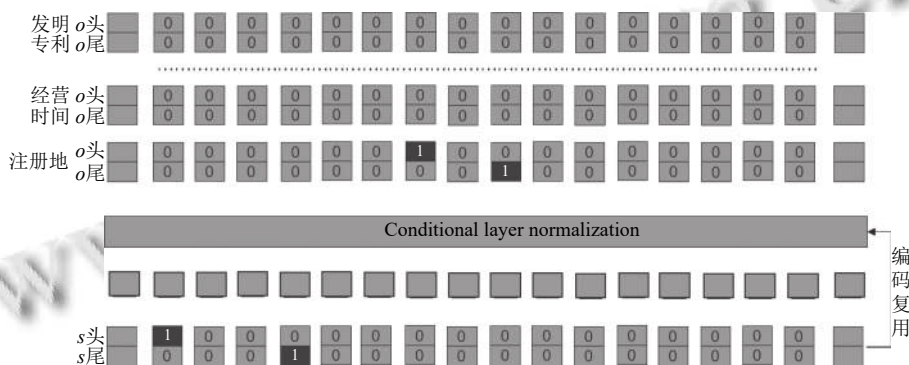


图 4 BERT 模型抽取 o 的结构表示

通过使用 BERT 模型, 可以很好地利用 BERT 的双向功能, 因为 Transformer 编码有注意力机制 self-attention, 还可以通过 BERT 获取比词汇更高级别的句子级别的语义表征, 因此能获得更好的匹配效果. 另外, BERT 的输入层和输出层更通用, 它的微调成本也比较小.

通过属性抽取模型, 即可在政策文本中自动获取

政策的属性条件及条件的指标设置, 只需要后续再进一步获取用户填入的企业信息就可以进行政策的匹配度计算.

2.3 政策匹配度计算

在产学研相关的政策池中, 适用于快速匹配的政策属性一般包括: 适用地区、申报起始日期、经营范

围、企业类型要求、企业注册地区、注册时间、注册资金、场地面积、经营范围、研发投入、营业收入、在职员工。

通过上面的政策属性抽取模型,我们可以获得每个条件匹配的信息三元组,在这一步,我们就可以将政策条件信息和企业自身信息进行匹配度 I 的计算。由于当前涉及企业用户的相关信息并没有一个固定的来源,暂时不能自动获取企业相关信息,故当前系统中企业方的信息需要用户逐项录入。首先,我们需要获取企业的条件信息,企业用户可以在系统中进行完善,填入的信息越完整越有利于做到精准的匹配;然后,通过将政策的条件与企业的条件进行逐一的数据匹配,判断企业的条件信息是否满足政策的条件,给每个条件 c_i 的逻辑结果置为 $\text{true}(l_i=1)$ 或者 $\text{false}(l_i=0)$, 然后为每个条件设置一个权重值 ω_i 来衡量每个条件对整篇政策得分的贡献值,权重的比例设置为用户可以自定义,匹配度的计算表达如式(2)所示:

$$I = \sum (\omega_i \times l_i) \quad (2)$$

对一些硬性的条件指标,比如注册地,如果政策条件要求是在苏州,那么注册地在南京的企业则不符合此篇政策,此时可以判定该政策与企业条件不匹配。针对这种情况,通过设置一个布尔类型的字段 `isImportant` 作为硬性指标的标识,优先遍历查找 `isImportant` 为 `true` 的条件。如果有一个这样的条件不满足,则可以给整篇文档匹配 0 分,此篇不符合要求的政策也不会推荐给该企业用户。如果硬性指标都满足,再进行常规判断。最后,在每个条件遍历匹配后,综合整篇政策的条件匹配情况计算总得分,即为整篇政策的匹配度。

在实际系统中,政策匹配计算过程采取自动识别与人工校准相结合,对于无法自动识别条件范围的政策,由管理员进行校准,设置政策中各条件指标的区间值。

企业填报企业信息后与政策库内政策进行匹配计算,匹配后按照申报起始日期及匹配度得分情况进行排序展示。

3 政策匹配系统实现

本文设计政策匹配系统架构如图5所示,共分为4层,分别是数据库层,业务层,平台层,用户交互层。不同功能构件被组织到分层的体系结构中,可以使每个构件处于只依赖于所处位置的下层而非上层的构件的

层次。这种分层体系结构便于在将来可以有更好的适应性变化。

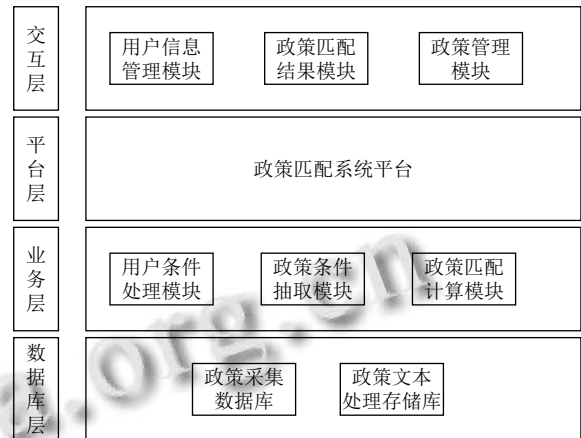


图5 系统结构框架图

政策匹配系统中的数据库层负责政策的采集存储,以及对政策文本预处理过后的存储库。业务层系统主要有对用户条件的处理功能,对政策属性条件抽取功能,对政策匹配计算功能。平台层是政策匹配系统的管理平台,用于连接业务层功能和用户的一个交互平台,提供系统业务层的功能接口。用户交互层包括企业用户和管理员的交互操作。用户操作主要是信息管理模块和政策匹配结果模块。管理员操作主要是政策业务的多方面管理功能。

系统功能结构如图6所示,主要是分为用户管理模块、政策匹配模块和政策管理模块。

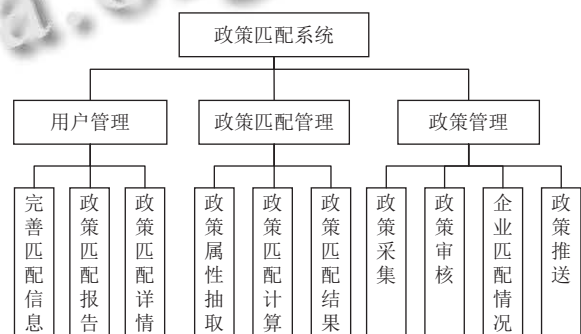


图6 系统功能结构图

在用户管理模块,主要是用户完善政策匹配所需企业信息,填入个体信息后获取匹配政策的报告,并查看政策匹配详情。系统通过业务逻辑层的政策匹配度计算,给用户显示相应的政策匹配报告。进入每篇政策的详情页面,可以查看每一项从政策中抽取出来的属

性条件与当前用户匹配的结果,显示符合或不符合,以及政策的原文及附件内容。

政策匹配模块是本系统中的核心功能,在系统业务层中实现,包括政策条件的属性抽取和政策匹配计算。政策条件的属性抽取模块,调用前述的基于BERT的政策属性抽取模型,实现政策中条件属性的自动抽取。属性抽取以后,在政策匹配计算模块,结合用户信息和抽取信息,进行政策匹配度计算得分,将结果值的返回给平台层显示。

在后台的政策管理模块,主要包括政策采集处理和存储、政策审核、查看企业匹配数据和政策推荐。首先是对政策数据的获取及存储,政策文本的获取主要来源于网络爬虫,也可以手工添加;随后存储在MySQL数据库中,对于特定的需要操作的文本,也使用了Python中的Pandas库用于分析结构化数据,还可以提供数据清洗功能。其次是政策数据的预处理,通过爬虫获取的政策数据可能存在不完整、不一致、异常值等低质量的数据,所以需要对一手原始数据进行一定的预处理操作,比如数据清洗、数据归约、数据变换等^[17]。这里使用的方法有去符号,去停用词,中文分词,去标记信息等操作。进而是利用标点作为句子分隔符对于文章语句进行划分,形成待处理的政策文本数据更新存放在数据库中。

在后台管理平台,管理员还可以对政策的录入进行审核,对企业政策匹配后的结果进行查看,并可以将与相关企业匹配度高的政策进行推送。政策推送分为即时消息推送、短信推送。系统将匹配度较高的政策推送给企业账户联系人,企业登录系统后能查询匹配的政策信息,企业选定关联政策后,在政策申报截止等关键时间节点系统自动通过短信通知或系统通知的方式告知用户,便于用户及时知晓当前政策,避免错过申报。经过本系统了解到的政策,不像传统的方式那样是需要完全人工分析解读的,而是经过筛选后符合条件的,是相对较精准的。

4 实验与应用成效

通过训练政策属性抽取模型,进行实验验证后,部分政策实例的条件分析结果展示如表2-表4所示。

政策实例 1: 在江苏境内注册,具有独立法人资格,有必要的研发投入,研发投入不低于 200 万元,需实际注册并运营满 2 年。

表 2 实验抽取结果示例 1

<i>s</i>	<i>p</i>	<i>o</i>
申报主体	注册地	江苏
申报主体	研发投入	不低于200万元
申报主体	经营时间	满2年

政策实例 2: 申报单位为在盐城市注册的软件企业,软件业务收入不低于 1 000 万元。

表 3 实验抽取结果示例 2

<i>s</i>	<i>p</i>	<i>o</i>
申报单位	注册地	盐城市
申报单位	营业收入	不低于1 000万元

政策实例 3: 机构用于电子商务相关产业的总体办公建筑面积应大于 3 000 平方米,产业规模较大、从业人员超过 100 名,且配套设施完善。

表 4 实验抽取结果示例 3

<i>s</i>	<i>p</i>	<i>o</i>
机构	企业类型	电子商务
机构	场地面积	大于3 000平方米
机构	在职人员	超过100名

从上述实验结果中可以看出,通过我们的政策属性抽取模型,可以在整篇政策文档中有效的抽取出政策文本中的结构化条件属性,从而对每篇政策文档做出快速的条件解读,通过政策匹配加快企业获取信息的效率,解决了企业用户准确获取政策难,匹配信息繁琐的问题。

本文的政策匹配系统已经在江苏省产学研合作智能服务平台中进行了集成和示范应用。下面是一些政策匹配系统的使用操作页面,首先是用户登录系统,完善企业的相关信息,如图7所示。

企业登录系统完善信息后,可以获取根据自身条件的匹配政策的列表,政策的展示是按匹配度得分情况和发布时间排序展示的,如图8所示。

点击任意一条政策,可以在政策详情页面看到每篇政策的条件指标匹配详情,同时还有政策的原文供企业用户详细查看,政策详情页面如图9所示。

目前我们的政策匹配系统,在政策管理采集模块,采集到的有效政策有 536 条,已经成功注册并且有效使用过政策匹配功能的用户有 386 个,未登录注册只作为游客访问该功能的用户达 800 多个。图 10 为系统运营情况,展示了部分注册登录的用户政策匹配的使用情况(为了隐私保护,企业名称暂且隐藏)。

注册资本: 1008.00 万元 成立日期: 2012-01-01

注册地址: 江苏省 淮安市 淮安国家高新区

企业类型: 民营

企业性质: 科技型企业

所属行业: 电子信息

企业资质: 100强企业

100强企业 请选择 获得时间: 选择日期

人事信息

企业总人数: 500 人 领军人数: 60 人

大专及以上人数: 486 人 本科及以上人数: 368 人

硕士及以上人数: 276 人 博士及以上人数: 30 人

经营数据 单位: 万元

年份	总资产	净利润	营业收入	税金	研发投入
2021	20000.00	0.00	16900.00	0.00	1000.00
2020	18600.00	0.00	0.00	0.00	700.00
2019	9670.00	0.00	0.00	0.00	600.00

知识产权

知识产权总数: 300 个 专利总数: 230 个

图7 企业完善信息页面

全部 国家级 省级 市级 区县级

企业政策报告 最新政策

- [省级]关于组织申报2022年度省知识产权战略推进计划项目的通知** 匹配度: 100% **申报期内**

发文单位: 江苏省知识产权局

所属地区: 江苏省

截止日期: 2022/02/15
- [省级]关于组织申报2022年第一批江苏省产学研合作项目的通知** 匹配度: 100% **未开始**

发文单位: 江苏省科技厅

所属地区: 江苏省

截止日期: 2022/02/28
- [省级]关于申报推荐2021年度省级专精特新小巨人企业和组织开展有关...** 匹配度: 100% **已过期**

发文单位: 江苏省工信厅

所属地区: 江苏省

截止日期: 2021/11/10
- [省级]关于组织申报2022年江苏省战略性新兴产业和服务业标准化试点...** 匹配度: 100% **申报期内**

发文单位: 江苏省市场监督管理局

所属地区: 江苏省

截止日期: 2022/01/20
- [省级]省发展改革委关于开展现代服务业高质量发展领军企业培育工...** 匹配度: 100% **申报期内**

发文单位: 省发展改革委

所属地区: 江苏省

截止日期: 2021/12/30
- [省级]省市场监管局关于组织申报2022年度江苏省地方标准项目的通知** 匹配度: 100% **申报期内**

发文单位: 省市场监管局

1. 关于组织申报2022年度省知识产权战略推进计划项目的通知

2. 关于组织申报2022年苏州市科技发展计划(碳达峰碳中和科技支撑重点专项)项目的通知

3. 关于组织申报2022年第一批江苏省产学研合作项目的通知

4. 关于组织申报2021年南京市工业企业技术装备投入财政奖补项目(第二批)的通知

5. 关于申报推荐2021年度省级专精特新小巨人企业和组织开展有关工作的通知

6. 关于组织开展2021年度省市级新型研发机构新建备案工作的通知

7. 关于组织申报2022年江苏省战略性新兴产业和服务业标准化试点项目的通知

8. 省发展改革委关于开展现代服务业高质量发展领军企业培育工程第一批领军企业申报推荐工作的通知

图8 企业获取的匹配政策列表页面

本文系统采用了基于深度学习的属性抽取方法,能自动抽取政策条件属性,并根据用户填入的企业信

息,提供相匹配的政策信息.该系统的特色应用体现在以下几点:(1)自动采集分析,具备数据采集功能,便于

各类政策汇总分析与管理; (2) 精准匹配, 系统操作简单高效, 精准匹配用户需求, 自动更新推荐结果; (3) 政

策自动推送, 为新政策执行中提高覆盖面和执行率等提供了强时效性的辅助工具.

匹配度: 62.5%

政策指标	指标详情	企业情况
地区	江苏省	符合
企业性质	社会团体 或者 科技企业	符合
注册资金	大于100万元	不符合
企业总人数	大于200个	符合
企业类型	国资 或者 民营 或者 外资	不符合
所属领域	电子信息 或者 装备制造 或者 新材料	符合
领军人数	大于3个	符合

图9 政策匹配详情页展示

企业政策匹配使用情况

企业名称: 政策等级: 使用时间: 年-月-日 年-月-日

使用企业名称	匹配政策等级	匹配政策数量	使用时间
江苏...公司	省级	95	2021-12-28
江苏...公司	国家级	27	2021-12-28
江苏...公司	全部	124	2021-12-28
淮安...公司	全部	124	2021-12-28
淮安...公司	国家级	27	2021-12-28
淮安...公司	区县级	0	2021-12-28
淮安...公司	市级	6	2021-12-28
淮安...公司	省级	95	2021-12-28
淮安...公司	国家级	27	2021-12-28
淮安...公司	全部	124	2021-12-28

图10 企业使用情况

5 总结

本文针对企业用户查找匹配相关政策不方便的问题, 通过研究基于政策属性抽取的匹配方法, 设计实现了政策匹配系统, 形成了一套根据企业自身实际条件去获取相应匹配政策的系统流程, 极大提高了企业获取政策并实现政策匹配的效率. 本文的属性抽取方法有效利用了 BERT 模型自有的双向 Transformer 结构, 可以更好地表征上下文语义信息. 在下一步研究中, 希望通过扩充数据集训练算法模型, 进一步提高本文方法对政策文本分析的能力; 此外, 可以考虑加入注意力机制对于一些相似的属性进行分析抽取, 对于关系类的提取可以尝试在模型标注的基础上利用规则和模板的方法. 对于企业用户的信息录入方面, 后期研究的方向可以考虑通过工商部门自动获取一些企业的信息以

减少录入项. 通过以上方面进一步研究和改进, 可以进一步提高企业用户的使用效率.

参考文献

- 1 钟登华. 新工科建设的内涵与行动. 高等工程教育研究, 2017, (3): 1-6.
- 2 李梓涵昕, 朱桂龙. 产学研合作中的主体差异性对知识转移的影响研究. 科学学研究, 2019, 37(2): 130-138.
- 3 万晓. 产学研融合背景下高职院校科技成果转化效率及其提升路径研究. 鄂州大学学报, 2021, 28(2): 72-74.
- 4 张颖, 蒋浩, 乔培宇. 互联网+产学研合作智能服务平台的生态服务模式. 计算机系统应用, 2019, 28(4): 69-75. [doi: 10.15888/j.cnki.csa.006834]
- 5 罗化, 郭丽芳. 基于区块链技术的产学研信息共享模式研究. 管理现代化, 2020, 40(6): 42-45.

- 6 裴雷, 孙建军, 周兆韬. 政策文本计算: 一种新的政策文本解读方式. 图书与情报, 2016, (6): 47-55.
- 7 马海群, 张斌. 我国政策计量研究: 方法与模型. 数字图书馆论坛, 2019, (5): 2-8.
- 8 宋伟, 夏辉. 地方政府人工智能产业政策文本量化研究. 科技管理研究, 2019, 39(10): 192-199. [doi: [10.3969/j.issn.1000-7695.2019.10.027](https://doi.org/10.3969/j.issn.1000-7695.2019.10.027)]
- 9 Dedaić MN. Politics as text and talk: Analytic approaches to political discourse. *International Politics*, 2004, 41(2): 286-287. [doi: [10.1057/palgrave.ip.8800052](https://doi.org/10.1057/palgrave.ip.8800052)]
- 10 赵洪, 王芳, 王晓宇, 等. 基于大规模政府公文智能处理的知识发现及应用研究. 情报学报, 2018, 37(8): 805-812. [doi: [10.3772/j.issn.1000-0135.2018.08.007](https://doi.org/10.3772/j.issn.1000-0135.2018.08.007)]
- 11 邓文红. 基于知识管理的办公系统智能化研究 [博士学位论文]. 成都: 西南交通大学, 2013.
- 12 陆川, 吴波, 邓华, 等. 一种政策大数据智能匹配系统及匹配方法: 中国, 201710803785. X, 2017-09-08.
- 13 李发明. 一种用于企业服务的政策智能匹配系统的构建方法: 中国, 202011353435.6, 2020-11-26.
- 14 Su JL. A hierarchical relation extraction model with pointer-tagging hybrid structure. <https://github.com/bojone/kg-2019>. (2020-05-16)[2022-03-06].
- 15 屠可伟, 李俊. 句法分析前沿动态综述. 中文信息学报, 2020, 34(7): 30-41. [doi: [10.3969/j.issn.1003-0077.2020.07.003](https://doi.org/10.3969/j.issn.1003-0077.2020.07.003)]
- 16 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 4171-4186.
- 17 刘明吉, 王秀峰, 黄亚楼. 数据挖掘中的数据预处理. 计算机科学, 2000, 27(4): 54-57. [doi: [10.3969/j.issn.1002-137X.2000.04.015](https://doi.org/10.3969/j.issn.1002-137X.2000.04.015)]

(校对责编: 孙君艳)