

中山大学本科生课程作业

# 养老保险课程论文



SUN YAT-SEN UNIVERSITY

中山大學

嶺南(大學)學院

LINGNAN (UNIVERSITY) COLLEGE

姓名：李思远

学号：20327077

学院：岭南学院

专业：金融学

## 目录

摘要 .....	4
1. 引言及研究现状 .....	4
1.1 选题背景与研究意义 .....	4
1.2 研究现状 .....	5
1.2.1 个人养老金政策梳理 .....	5
1.2.2 互联网舆论研究 .....	6
2. 数据采集与清洗 .....	7
2.1 数据爬取 .....	7
2.2 数据清洗 .....	7
2.3 知乎平台概览 .....	8
3. 主题分析 .....	10
3.1 基于词云工具 .....	10
3.2 基于 LDA 算法 .....	12
3.2.1 LDA 算法介绍 .....	12
3.2.2 LDA 实现 .....	12
4. 舆论关切与政策建议 .....	18
4.1 关切：作为公共政策的个人养老金 .....	18
4.2 关切：作为金融产品的个人养老金 .....	18
4.3 关切：政策国别比较 .....	19
参考文献 .....	19

## 图表目录

图表 1：爬取知乎数据情况.....	7
图表 2：2020 年知乎用户性别比例情况.....	9
图表 3：2020 年知乎用户年龄分布情况.....	9
图表 4：2020 年知乎用户地域分布情况.....	10
图表 5：全部数据整体的词云图.....	11
图表 6：考虑影响力的词云图.....	11
图表 7：多维缩放的话题图.....	13
图表 8：主题一、二和三的关键词及其在整体中的比重.....	14
图表 9：“美国”的条件主题分布.....	16
图表 10：第四类主题下部分回答节选.....	16
图表 11：“新冠”的条件主题分布.....	17
图表 12：第五类主题下部分回答节选.....	17

# 舆论对个人养老金政策有何关切

## —基于中文问答社区知乎的数据

李思远 20327077 金融学

### 摘要

本文爬取、分析知乎平台下关于个人养老金政策的问答文本数据，并应用 LDA 算法对文本主题进行提取，以期了解网络舆论对个人养老金政策的关切，并将这些关切概括为个人养老金的公共政策属性和金融产品属性，为完善个人养老金政策的构建与执行、回应舆论关切提供了建议。

**关键字：**

## 1.引言及研究现状

### 1.1 选题背景与研究意义

民意是公共政策的基石，网络舆论是民意网络化呈现，是公共政策的晴雨表。随着互联网信息技术的飞速发展，民意互联网表达渠道越来越畅通，加之公众的民主意识不断增强，通过网络参政议政的愿望越来越强烈，影响公共政策决策的社会力量和渠道越来越多。运用网络舆情监测分析手段了解公共政策应然的状态，助使公共政策朝着合乎民意的方向发展，发挥其实应的价值<sup>[1]</sup>。另一方面，政策的实施也离不开民众的配合。政策在实施过程中，如果能够了解民意之关切，知晓施行中的阻力和助力，对民生福祉大有裨益。

根据灼识咨询数据，知乎是中国第五大综合在线内容社区，也是最大的中文在线问答社区<sup>1</sup>。知乎的用户基础较大，且有其独特画像，是研究某一政策在

---

<sup>1</sup> 新股 IPO | 知乎招股中，第一家以双重主要上市方式回港上市的中概互联网公司！.天风国际证券 . 天风国际微信公众号 . 2022/04/11.url=[https://mp.weixin.qq.com/s?\\_\\_biz=MzI5MTczNzQwMg==&mid=2247497486&idx=3&sn=d](https://mp.weixin.qq.com/s?__biz=MzI5MTczNzQwMg==&mid=2247497486&idx=3&sn=d)

特定群体中相关舆情的有效平台。

本文将爬取若干知乎平台中以个人养老金为话题的高活跃度问题，对问题下的用户回答进行文本分析，研究知乎平台上用户对个人养老金政策的看法与观点，了解其关切，并形成一定的政策建议。

## 1.2 研究现状

### 1.2.1 个人养老金政策梳理

国务院于 1991 年颁布了《关于企业职工养老保险制度改革的决定》，首次提出了养老体系“三大支柱”的概念，指出要逐步建立基本养老保险、企业补充养老保险和个人商业养老保险相结合的养老保险制度。据国家统计局数据，截至 2021 年底，我国养老金规模合计 10.8 万亿，养老体系仍以第一支柱、第二支柱为主，第三支柱尚处于起步发展阶段。第一支柱基本养老保险参保人数 10.3 亿人，累计结存 6.4 万亿。第二支柱方面，共有 11.75 万户企业建立企业年金，参加职工 2875 万人，企业年金投资运营规模 2.6 万亿；职业年金投资运营规模 1.8 万亿。

2022 年 4 月 21 日，国务院办公厅印发《关于推动个人养老金发展的意见》，推动发展适合中国国情、政府政策支持、个人自愿参加、市场化运营的个人养老金，实现养老保险补充功能。个人养老金制度是我国社会保障体系建设顶层设计中的重要制度安排。2022 年 11 月 4 日，人社部等五部门联合印发《个人养老金实施办法》，搭建了整体制度框架；2022 年 11 月 4 日，财政部、国家税务总局联合发布《关于个人养老金有关个人所得税政策的公告》，阐明了税收优惠安排；2022 年 11 月 4 日，证监会发布《个人养老金投资公开募集证券投资基金业务管理暂行规定》，2022 年 11 月 18 日，银保监会印发《商业银行和理财公司个人养老金业务管理暂行办法》，2022 年 11 月 22 日，银保监会发布《关于保险

公司开展个人养老金业务有关事项的通知》，丰富了个人养老金的投资范围<sup>2</sup>。

在现行政策标准下，个人养老金实行个人账户制度，缴费完全由参加人个人承担，实行完全积累。参加人通过个人养老金信息管理服务平台，建立个人养老金账户。参加人可以用缴纳的个人养老金在符合规定的金融机构或者其依法合规委托的销售渠道购买金融产品。个人养老金资金账户实行封闭运行，其权益归参加人所有，除另有规定外不得提前支取。

### 1.2.2 互联网舆论研究

关于互联网舆论的研究有两个大类，一类研究集中在对概念的内涵和外延的定义与探索，另一类则注重以实际的算法和数据进行分析。以定量方法辅助定性研究者亦有之。考虑到中文语境中“舆情”、“舆论”虽有微妙的差别，但在很多时候其研究方法乃至概念本身可以互换，本文并未对此二者进行严格的区分。

2009年，曾润喜将网络舆情定义为由于各种事件的刺激而产生的通过互联网传播的人们对于该事件的认知、态度、情感和行为倾向，并认为为构建高效的网络舆情管控工作机制，必须要提高舆情重要性认识、制定应急预案、利用网络舆情规律、完善网络监控体制、改善国家社会关系以及加强理论应用研究等。同年，黄晓斌和赵超分析了文本挖掘技术的主要功能，提出网络舆情信息挖掘分析模型，并以实例说明文本挖掘在网络舆情分析中的应用<sup>[2]</sup>。同样在2009年，王伟等根据对网络舆情分析的需求，构建出基于聚类的网络舆情热点发现及分析系统。通过对样本网页文本的特征提取，构建向量空间模型，使用OPTICS算法获取网页热点簇，根据热点簇特征向量对网页进行二次聚类，从而获取关于舆情的时间演变模式，为相关领域研究提供决策支持<sup>[4]</sup>。2012年，张一文等将贝叶斯网络建模方法和网络舆情态势评估相结合，同时基于贝叶斯网络三个重要特点——复杂关联关系表示能力、概率不确定表示能力以及因果推理能力，提出基于贝叶斯网络建模的网络舆情态势评估方法。通过对关键指数

---

<sup>2</sup> 资料来源为中国国政府网对应通知.url=<http://www.gov.cn/>

据进行仿真和学习,建立网络舆情态势评估模型,从而对网络舆情态势进行有效评估和预测<sup>[4]</sup>。2014年,孙玲芳等利用遗传算法优化BP神经网络的初始权值与阈值,构建了基于BP神经网络和遗传算法的网络舆情危机预警模型。最后,通过仿真实验,结合5个具体案例对该模型进行了验证与分析。证明了网络舆情预警指标体系与遗传BP神经网络模型是有效可行的<sup>[5]</sup>。

关于网络舆情的研究,其定性结论在过去十数年中逐步拓展、细化,其定量研究所采用的方法亦随计算机技术的进步和算法的迭代而更新。现代优化算法、人工智能算法逐渐多见于舆情分析的研究之中,并发挥重要作用。

## 2. 数据采集与清洗

### 2.1 数据爬取

为规避知乎的反爬虫机制,使用Python的Selenium库操作Firefox浏览器进行数据爬取。爬取范围为知乎“个人养老金”话题下回答数超过30的问题,并逐个处理讨论个人养老金却没有正确归类的话题。取得数据概况如下。

图表 1: 爬取知乎数据情况

序号	问题	回答数	文本量(字)
1	五部门联合发布个人养老金实施办法,有哪些信息值得关注具有哪些积极意义?	386	133,963
2	个人养老金实施办法发布明确规定个人养老金资金账户内的资产都可以继承,具体情况如何?	172	71,226
3	个人养老金有必要买吗?	1003	252,599
4	个人养老金有参与的必要吗?	38	15,865
5	30岁了该不该存个人养老金?	480	139,649
合计		2079	613,302

### 2.2 数据清洗

为避免因编码问题带来的困扰,对数据中的特殊符号进行剔除。

为方便后续研究,对数据进行统一的分词处理。通过Python的Jieba库进行

分词，在分词过充中设置中文常用停用词，停用词来自于哈尔滨工业大学智能技术与自然语言处理实验室。

考虑评估每一项回答的影响力。在爬取回答内容时，粉丝数量、点赞数量、评论数也被一齐收集。点赞数量较多的回答显然能够反映更多的民意，但评论数量可能不是回答反应民意程度的有效测度，因为并非所有评论都对回答表示支持，而通过情感识别、关键词识别等方式分析某一评论是否支持该评论所属的回答的可行性较差，准确率较低。这一方面的研究比较空白，本文自行设定了回答权重的模型，模型中的自变量和权重来自本文作者的经验与直觉。

$$\text{影响力} = 1 + 0.5 * \text{赞同数} + 0.3 * \text{评论数} + 0.0001 * \text{粉丝数}$$

在后续处理中，对影响力取自然对数，乘以 10 并向下取整，将得到的数作为对应回答的额外频数。

考虑到知乎的推送机制，已关注用户并不经常出现在推荐列表上，而是需要进入关注界面浏览，且数据中亦出现过粉丝数与赞同数相关性较差的情况，也即高粉丝数的用户经常得到较为有限的赞同，我们为粉丝数赋予较低的权重。

如若对回答的影响力进行深究，或许需要先假设以影响力为因变量、以赞同数、评论数和粉丝数等数据为自变量的模型是存在的，而相关系数的确认则需要对公司的用户画像、用户习惯进行深入的研究，这并非本文所属领域内的议题，亦不在本文作者能力和时间范围之内，故而本文将本文提出的模型作为先验的结论加以接受，并期望在未来会有新闻传播、社会学等领域的研究者解决影响力模型的问题。

## 2.3 知乎平台概览

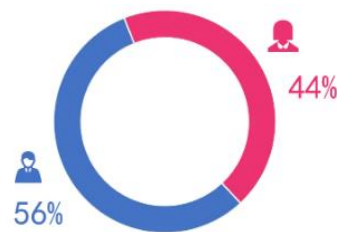
知乎是一个高品质的问答、知识分享社区，以及创作者聚集的原创内容平台。该平台拥有独特而专业的社区氛围和高质量的内容积累，吸引了大量关注互联网、科技、商业、影视、时尚等领域的用户，致力于分享知识、经验和见解。自 2011 年 1 月正式上线至 2020 年，知乎 APP 的月活跃用户数达到了 6700



万<sup>3</sup>，使用设备数达到 4.85 亿，每日搜索量达到了 2545 万次。2021 年是知乎上线十周年，其产品口号更是“有问题，就会有答案”。

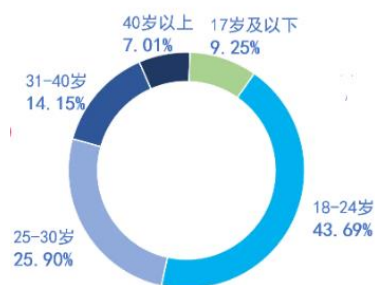
知乎的主要用户可分为两类：创作者和普通用户，创作者类型包括业内精英、自媒体从业者和“知+用户”等；普通用户类型则包括知识型中产阶级、注重品质生活的追求者以及新生代用户等。因为知乎用户角色转化存在闭环，即普通用户可以转化为创作者。此外，知乎大力激励普通用户成为创作者。

从性别、年龄和地域等多个维度来看，知乎的用户构成呈现多元化分布的趋势，并且以年轻化为主要特征，呈现出高消费的态势。我们选取 2020 年的数据进行静态分析。



图表 2：2020 年知乎用户性别比例情况

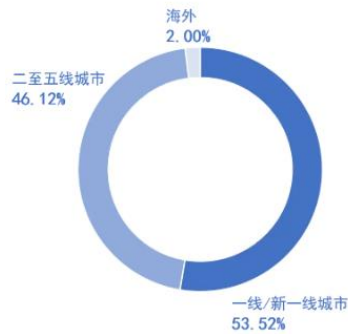
在性别方面，2020 年知乎男性用户占比 56%，女性用户占比 44%，性别分布相对均衡。这可能是由于知乎的内容不仅涵盖了知识和科普类的内容，还包括亲子、时尚、娱乐等吸引女性的板块内容，因此不倾向于特定性别，具有普适性的用户特征。



图表 3：2020 年知乎用户年龄分布情况

<sup>3</sup> 数据来源：未特殊注明均来自于当年财报。

在年龄方面，2020年，年龄在30岁以下的Z世代年轻人占据了高达78.8%的比例，30岁以上的用户占比为21.1%，其中25-30岁的社会中坚和24岁以下的“新新人类”是知乎的核心用户群体，社区氛围呈现年轻化的趋势。



图表 4：2020 年知乎用户地域分布情况

在地域方面，53.52%的用户分布在一二线城市，这些用户收入水平高于全国均值。

综上所述，知乎的主流用户是知识中产阶层和新新人群。这些用户处于高度自主汲取知识的阶段，有着较高的付费意愿和接受新事物的能力，社交需求强烈，用户行为活跃，其言行具有较高的研究价值，知乎用户对个人养老金政策的观点和态度能够在一定程度上反应部分中国年轻知识分子对养老金政策的态度。这一群体在个人养老金政策颁布的伊始正处在自身职业生涯的初期到前中期，是购买个人养老金产品的重要力量。

### 3. 主题分析

#### 3.1 基于词云工具

为对数据整体形成一个直观的印象，利用 Python 的 Wordcloud 库进行词云绘制。选择出现频数前 20 的词汇，删去“个人养老金”、“个人养老金账户”和“养老保险”等与政策的名称本身相关但对主题意义不大的词语，得到词云图如下。



图表 5：全部数据整体的词云图

可以看到，知乎用户在个人养老金政策相关问题下的回答中，出现频率最高的几个词是退休、投资、产品、社保和收入等。可以猜测知乎用户关心个人养老金政策对退休年龄、退休后的生活的潜在影响，将个人养老金看作一个投资渠道，并关注相应的风险与收益。

运用本文在 2.2 数据清洗章节中阐述的权重方法，绘制下一张词云图。



图表 6：考虑影响力的词云图

所得结果与未使用权重时相比未见明显区别。

## 3.2 基于 LDA 算法

### 3.2.1 LDA 算法介绍

文本相似度计算是各种文本挖掘技术的基石，有了文本相似度的定义就有了各种文本比较的理论依据。比如信息检索任务就可以看成是检索文本与被检索文本之间的一种相似度度量；而文本分类和文本聚类技术则运用了文本之间的离散度或者相似度概念；在文本自动摘要研究中，需要记住文本中句子之间的相似度问题；在自动问答系统中也需要计算问题与答案之间的相似度。向量空间模型（VSM）是信息检索领域最为经典的分析模型之一。其中基于 TF-IDF 的向量空间模型文本相似度计算方法是使用最广泛的文本相似度计算方法，这种方法以词在文本中出现频率以及在文本集中出现该词的频率来表征，但 SVM 不能完全建模自然语言的复杂性问题，比如语义的问题<sup>[7]</sup>。

隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型是一种对离散数据集（如文档集）建模的概率主题模型，是一种对文本数据的主题信息进行建模的方法，能够对语义进行一定程度的识别。LDA 模型通过对文档进行一个简短的描述，保留本质的统计信息，有助于高效地处理大规模的文档集。它有 3 层生成式贝叶斯网络结构，基于这样一种前提假设：文档是由若干个隐含主题构成，而这些主题是由文本中若干个特定词汇构成，忽略文档中的句法结构和词语出现的先后顺序<sup>[6]</sup>。

### 3.2.2 LDA 实现

通过 Gensim 实现 LDA 算法。Gensim 是一款开源的第三方 Python 工具包，用于从原始的非结构化的文本中，无监督地学习到文本隐层的主题向量表达。它支持包括 TF-IDF, LSA, LDA, 和 word2vec 在内的多种主题模型算法，支持流式训练，并提供了诸如相似度计算，信息检索等一些常用任务的 API 接口<sup>4</sup>。

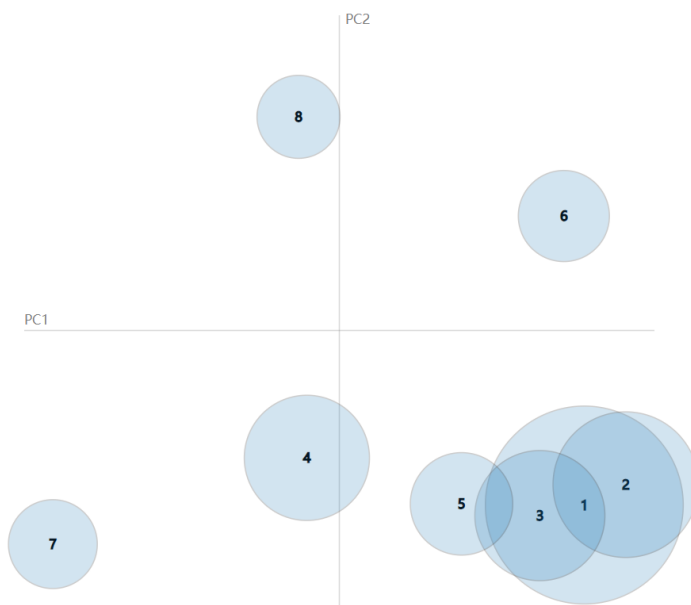
将分词后的数据导入 Gensim 设置的接口，生成词典，并通过 doc2bow 稀疏

---

<sup>4</sup> 官方文档：<https://github.com/apacheecn/gensim-doc-zh/>

向量生成语料库，生成 `gensim.models.ldamodel.LdaModel` 对象以训练 LDA 模型。

通过 `pyLDAvis`<sup>5</sup>对 LDA 模型的结果进行可视化。



图表 7：多维缩放的话题图

考虑到分类的效果，本文预设主题数量为 8 个。上图的八个气泡对应到 LDA 模型产生的 8 个主题，气泡的大小表示主题出现的频率，并佐以数字编号。通过多维尺度分析，`pyLDAvis` 提取出样本的主成分做维度，将主题分布到这两个维度上，主题相互之间的位置远近，就表达了主题之间的接近性。气泡距离采用的是 JSD 距离，可认为是主题间的差异度，气泡有重叠说明这两个主题里的特征词有交叉。

为平衡 LFDA 模型结果的普遍性与与特殊性，`pyLDAvis` 设置了可以调整的  $\lambda$  值。如果  $\lambda$  接近 1，那么在该主题下更频繁出现的词，跟主题更相关；如果  $\lambda$  越接近 0，那么该主题下更特殊、更独有（exclusive）的词跟主题更相关。考虑到本文数据来源均为讨论个人养老金的问题之下的回答，其回答间总的文本相似性会比较突出，因此应该尝试一些更小的  $\lambda$  值，否则无法体现不同主题之间的区分度。然而，过小的  $\lambda$  值会带来关键词过分生僻的问题，使其对回答真正

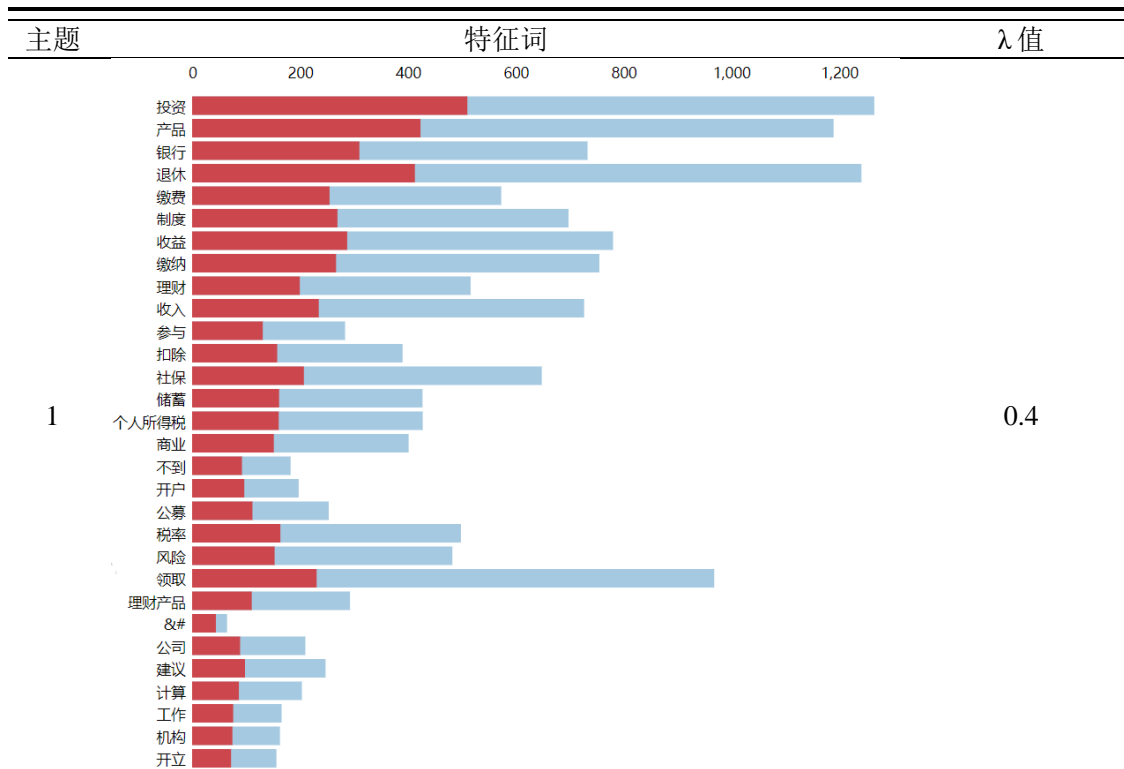
<sup>5</sup> 官方文档：<https://pyldavis.readthedocs.io/en/latest/readme.html>

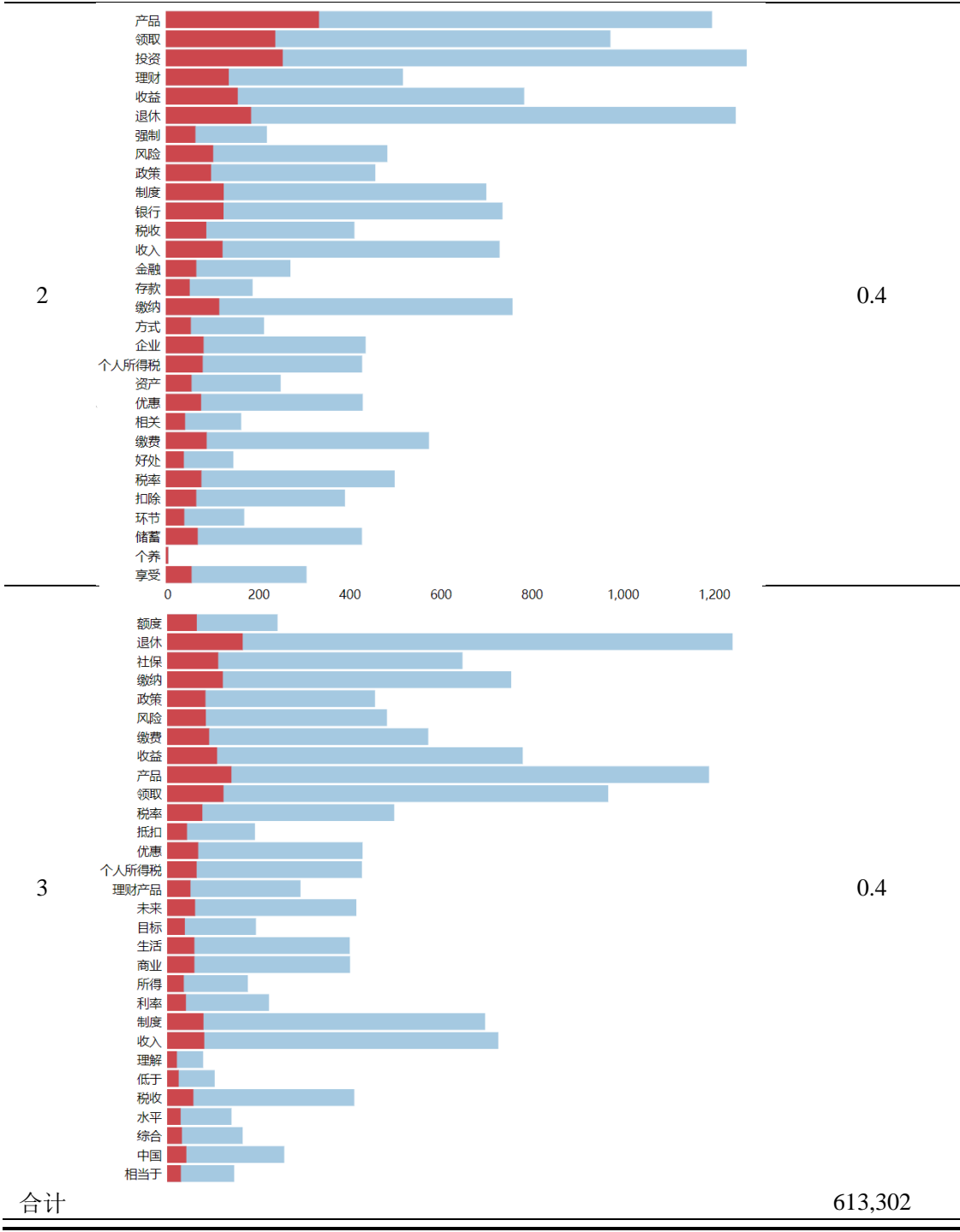
主题之测度的准确性受到怀疑。本文发现，设置  $\lambda$  值为 0.4 时提取的关键词效果最好。

观察图表 4 可知，主题一、主题二和主题三有明显的交集，且主题一对主题二、主题三形成近似于包含的关系，而主题二与主题三之间的交集较小。

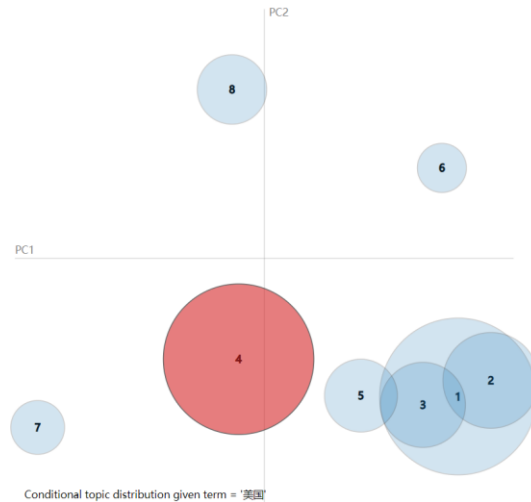
图表 5 列示了在  $0.4\lambda$  下各主题的关键词情况。缴费、制度、个人所得税、储蓄和收益是主题一、二和三的共同关切，而在二和三之间，主题二更关注理财、投资，主题三更关注社保。

图表 8：主题一、二和三的关键词及其在整体中的比重





第四类主题的独特关切在“美国”。



图表 9：“美国”的条件主题分布

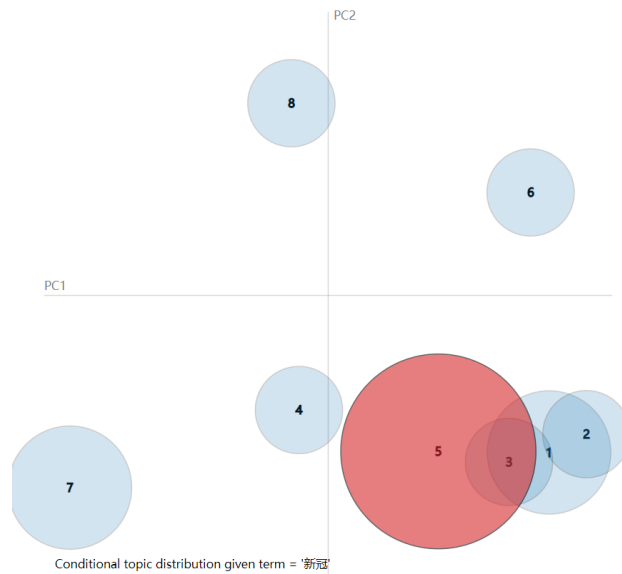
查阅此类主题下的部分回答，推测这一类回答多提及美国养老金政策、美国资本市场情况并与中国的政策和经济环境加以比较。

图表 10：第四类主题下部分回答节选

序号	回答节选	涉及内容	赞同数	评论数
1	至于养老金本身的理财收益，人家美国股市什么水平？排除战争这种特殊因素，长期看基本是连年上涨，自然可以养老金入市。但你看 A 股，准入过于复杂	美国资本市场	112	4
2	美国 401k 计划是一种缴费确定型（DC）计划，实行个人账户积累制，其建立需符合一定条件。	美国养老金政策	119	94
3	再说一个兑付期限问题，美国的 401K 计划，规定了 60 岁就可以给付，而当前的个人养老金只有一句模糊的“依照法定退休年龄”，而不巧的是，我们刚被修改了退休年龄	美国养老金政策	7,364	639
4	这设计对标是似美国 401K，而 401K，就是把你类似退休金的钱直接投资在资本市场的共同基金中，但可怜的是它可能都跑不赢被动的标普 500 指数！	美国养老金政策	20	0
5	不过当股市规模足够庞大的时候，是可以合理预期政策出台的，毕竟美国股市也是三大泡沫之一，托底股市有时候是必选项	美国资本市场	25	8
6	如在美国，起初拥有第二支柱养老金的个人不被允许参加个人养老金，后来有所放松，但是仍然明确，家庭年收入达到一定标准的，也不允许参加个人养老金缴费。	美国养老金政策	2	0



第五类主题的特殊性在于对新冠、感染的关注。



图表 11：“新冠”的条件主题分布

这一类回答多关注新冠病毒感染及其反复感染对人均寿命预期的影响，以及疫情带来的社会面风险。很多答主认为由于后遗症和反复感染，自己的预期寿命会大幅降低，以至于购买个人养老保险成为一项生命内净现值小于 0 的行为。所幸这一部分悲观预期得到的浏览和支持较少，影响力较弱。

图表 12：第五类主题下部分回答节选

序号	回答节选	涉及内容	赞同数	评论数
1	中青年完全不要买。65 岁退休，以中国的劳动强度，必有基础病，相当多人过 5 到 10 年就会死于新冠。	预期寿命	178	30
2	新冠确实是有后遗症的，不多，确实有...我觉得如果你才 20 多，不是我诅咒你，自己想想你活到 65 岁，需要感染多少轮。	预期寿命	20	0
3	请参考新冠疫情管控措施的变化，你品你细品	政策稳定性	107	0
4	聊到新冠转眼间新冠来势汹汹，为了老百姓生命财产的安全实行了为期 3 年的封控隔离，这导致经济下行...自己钱装在自己口袋里还是安全一些	疫情增加社会面风险	5	3
5	各位老哥应该听过之前别的保险公司卖的新冠险吧。现在中招的人多了就有人向保险公司寻求赔偿，结果保险公司说你得的是奥密戎不是新冠拒绝赔偿。现在你可以再考虑	合同诈骗、政策稳定性	4	0

---

个人养老金的事情了。

---

---

## 4.舆论关切与政策建议

本文首先通过词云的方法对样本整体的关切进行了研究。通过上文应用 LDA 模型对数据进行的文本聚类，本文从中提取出八个主题，并对其中五个可解释性较强的主题进行了分析，研究了不同的主题及其关注的特殊性。综合上述研究的结果，本文总结出知乎社区用户对个人养老金政策的关注情况如下。

### 4.1 关切：作为公共政策的个人养老金

词云图和 LDA 模型产生的前三个主题中对“社保”、“政策”和“制度”等词汇的关注反复提醒我们，个人养老金政策是养老保险三根支柱的一部分，网络舆论关注个人养老金作为公共政策，作为养老保险第三支柱的效益和社会影响。具体而言，关注个人养老金在实践上对养老体验的优化作用，以及相应的社会总体稳定性和福利的上升。

在这一方面，政策制定者和执行者可以注意以下三点：，做好第三支柱制度顶层设计，制定适合第三支柱发展的各项政策；做好第三支柱平台建设，搭建第三支柱基本公共服务平台；做好第三支柱监管工作，建立个人养老金运行监管体系。这些行动将提高个人养老金政策的公信力与可持续性。此外，个人养老金有望推动储蓄养老向投资养老的转变。个人养老金政策可有效集合个人养老资金，改善家庭资产配置，通过合理的市场化投资运营实现保值增值，同时为资本市场增加持有长期资产的机构投资者，而发挥养老资金长期属性，获取合理收益，促进国家金融结构调整和实体经济增长<sup>[10]</sup>。

### 4.2 关切：作为金融产品的个人养老金

同样是在全样本词云图和 LDA 模型的前三个主题中，“投资”、“产品”和“税率”等词语体现出网络舆论对个人养老金的金融产品属性的关切。事实上，对个

人养老金实际上的收益率的关注是所有回答中占比最大的<sup>6</sup>。

决策者处理此方面的问题，应当着重强调个人养老金账户的支取自由程度、账户中可以购买的理财产品的风险与收益等。作为超长期的投资，且没有自由支取的权限，以朴素的经济理论看，由于机会成本的损失，个人养老金账户上的投资应该在承担有限的风险的前提下拥有超过定期存款的期望收益。随着我国经济、文化的发展，我国居民的受教育程度逐渐提高，经济学、金融学知识的普及率也在相应地增加，如果不能拿出实际的超额收益，会越来越难吸引到公民的投资。个人养老金账户的税收优惠政策是其区别于一般投资的额外收益，这也为知乎平台网络舆论所关注，这项收益能够为公民带来的实际效用亟待精算领域的研究。

### 4.3 关切：政策国别比较

西方发达国家在个人养老金上的实践比较丰富，但国内个人养老金的实践和研究均起步较晚，制度领域在逐渐结合自己的实际国情进行探索和尝试。在此背景下，中国的个人养老金政策配用于和西方发达国家相关政策进行比较是一种必然。制定政策时参考国外先进政策的做法，并在政策宣发时一并讲解国外政策现状以及使政策契合我国国情的思考，有利于满足公民对政策制定思路的知情权。

## 参考文献

- [1] 邵德奇,冯超,王丽萍.网络舆情对公共政策影响的研究[J].安徽师范大学学报(自然科学版),2022,45(05):409-414.DOI:10.14182/J.cnki.1001-2443.2022.05.001.
- [2] 曾润喜.网络舆情管控工作机制研究[J].图书情报工作,2009,53(18):79-82.

---

<sup>6</sup>这里应该有一个基于主题词表的定量分析，但没有时间细化了

- [3] 王伟,许鑫.基于聚类的网络舆情热点发现及分析[J].现代图书情报技术,2009(03):74-79.
- [4] 黄晓斌,赵超.文本挖掘在网络舆情信息分析中的应用[J].情报科学,2009,27(01):94-99.
- [5] 孙玲芳,周加波,林伟健,候志鲁,许锋.基于 BP 神经网络和遗传算法的网络舆情危机预警研究[J].情报杂志,2014,33(11):18-24.
- [6] 王振振,何明,杜永萍.基于 LDA 主题模型的文本相似度计算[J].计算机科学,2013,40(12):229-232.
- [7] 黄承慧,印鉴,侯昉.一种结合词项语义信息和 TF-IDF 方法的文本相似度量方法[J].计算机学报,2011,34(05):856-864.
- [8] 张军,张丽,沈凡凡,谭海,何炎祥.RoBERTa 融合 BiLSTM 及注意力机制的隐式情感分析[J].计算机工程与应用,2022,58(23):142-150.
- [9] 魏来,王伟洁.基于互动仪式链理论的视频网站弹幕信息情感分析——以 Bilibili 健康科普类视频为例 [J]. 情报理论与实践,2022,45(09):119-126.DOI:10.16353/j.cnki.1000-7490.2022.09.016.
- [10] 董克用,施文凯.加快建设中国特色第三支柱个人养老金制度:理论探讨与政策选择[J].社会保障研究,2020(02):3-12.