基于主营产品相似度的A股市场行业分类研究

曹春晓 SAC NO：S1120520070003
杨国平 SAC NO：S1120520070002

2021年6月22日

目录

1.1 行业分类是二级市场投资的重要基础

华西证券 HUAXI SECURITIES

- 行业分类对于国民经济管理具有重要作用，同时，行业分类也是二级市场投资和研究的重要基石。
- 行业分类标准众多，一般情况可根据使用者角度不同将行业分类标准归为两种：

  - **管理型行业分类：**其目的主要在于正确反映国民经济内部的结构和发展状况，为国家宏观调控管理、各级政府及行业协会的经济管理提供标准，典型的如联合国国际标准产业分类(ISIC)、北美行业分类系统(NAICS)、国家统计局的行业分类标准和中国证监会的《中国上市公司分类指引》，均属于管理型行业分类标准。

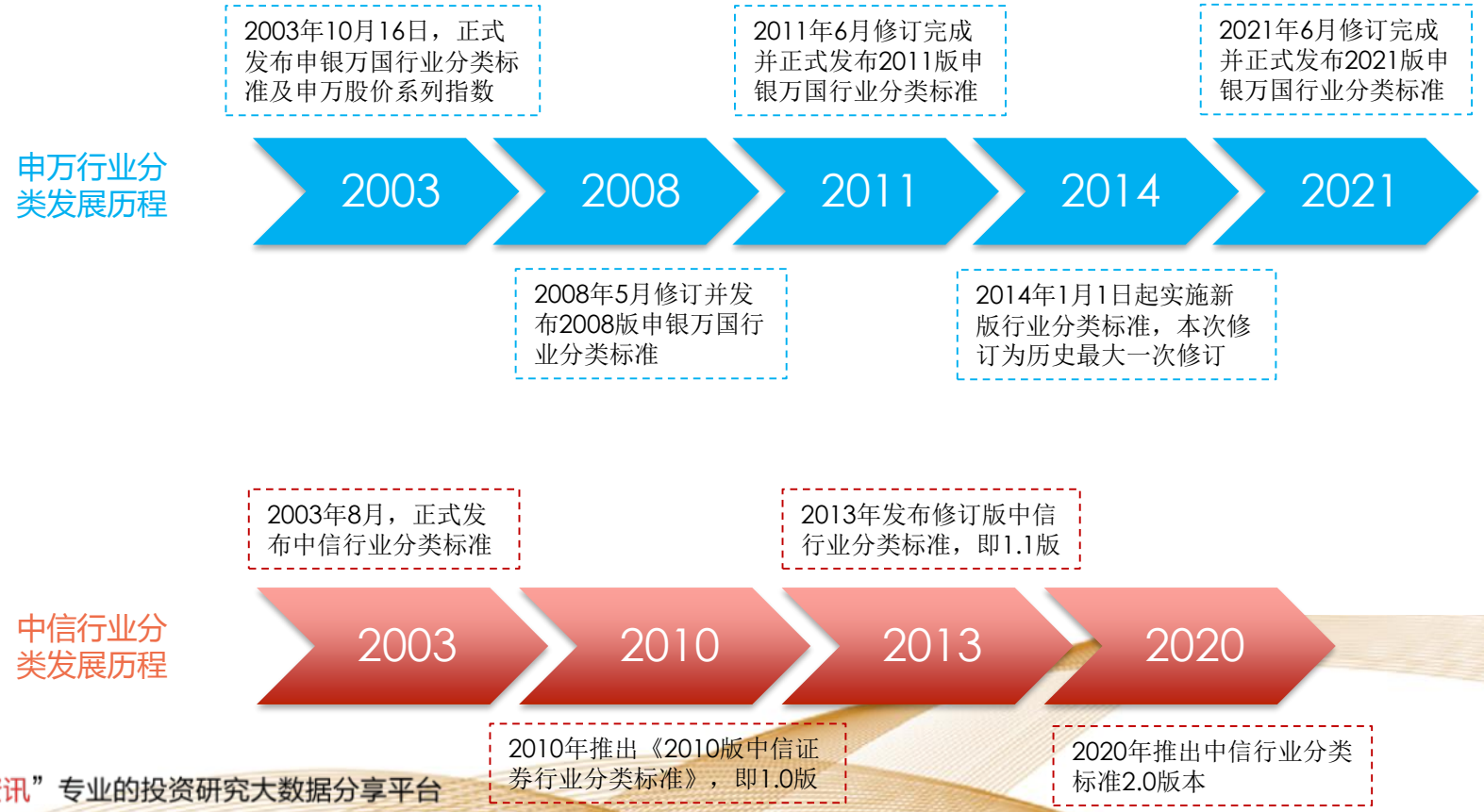
  - **投资型行业分类：**其目的主要在于为投资分析、业绩评价、资产配置等服务，比较有代表性的投资型行业分类标准如全球行业分类系统(GICS)、富时分类系统(FTSE)、MSCI行业分类、申万行业分类、中信行业分类等。

1.2 目前国内通用的行业分类体系相对固定

华西证券 HUAXI SECURITIES

- 从A股市场使用习惯来看，申万行业分类和中信行业分类是接受度最高的分类体系。尽管行业分类标准存在一定差异，但行业分类体系整体相对固定，往往数年才会调整一次。

图1：申万行业分类标准与中信行业分类标准历史变迁



资料来源：华西证券研究所整理

1.3 主营业务指标是行业分类考量的核心指标

华西证券 HUAXI SECURITIES

- 申万行业分类与中信行业分类相似，考虑的核心指标均为主营业务收入占比和主营业务利润占比，并适当考虑市场看法与投资习惯，公司未来发展规划等等，二者主要差异体现在投资收益的判断上。

图2：申万行业分类标准划分规则

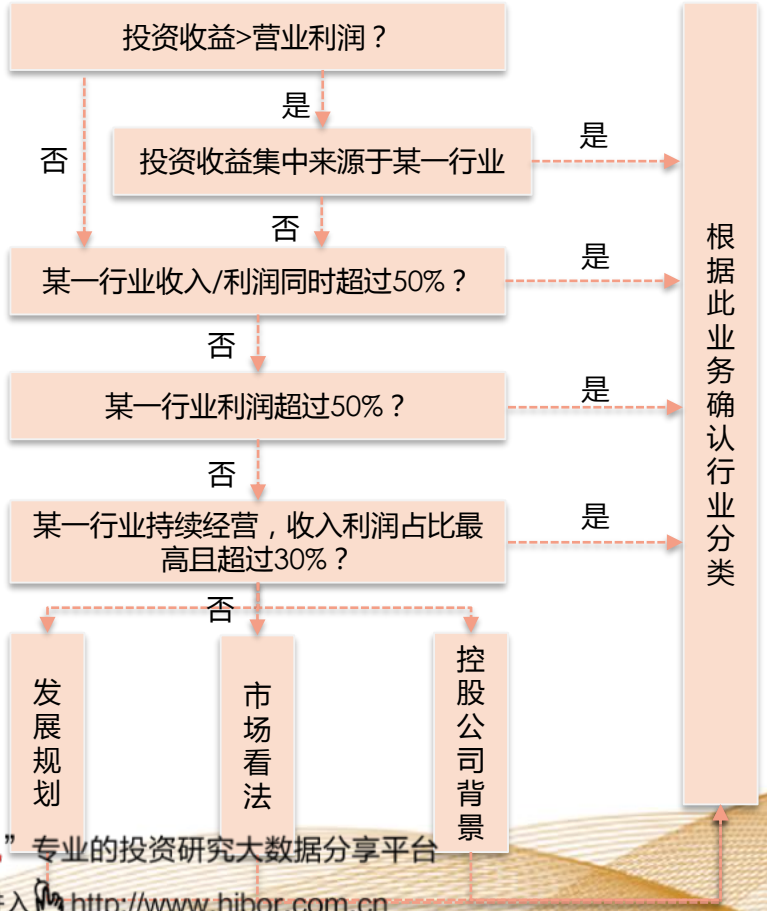


图3：中信行业分类标准划分规则

初步归类：某一行业收入/利润同时超过50%，直接归入该行业，收入利润不一致时以利润指标为准，适当考虑公司发展规划和市场看法

二次归类：某项业务收入和利润占比均最高，且均超过30%，则归入该行业，收入利润不一致时以利润指标为准，适当考虑公司发展规划和市场看法

三次归类：综合考虑公司发展规划及控股公司的背景情况分类；如果该公司没有明显的发展规划和控股公司背景，归入综合类。

例外情况：如果公司投资收益>主营利润，考虑投资收益是否集中来自于某一个行业，是则归入该行业；否则考虑公司发展规划和控股公司的背景，若没有明显的发展规划和控股公司背景，归入“综合金融-多领域控股Ⅱ-多领域控股Ⅲ”

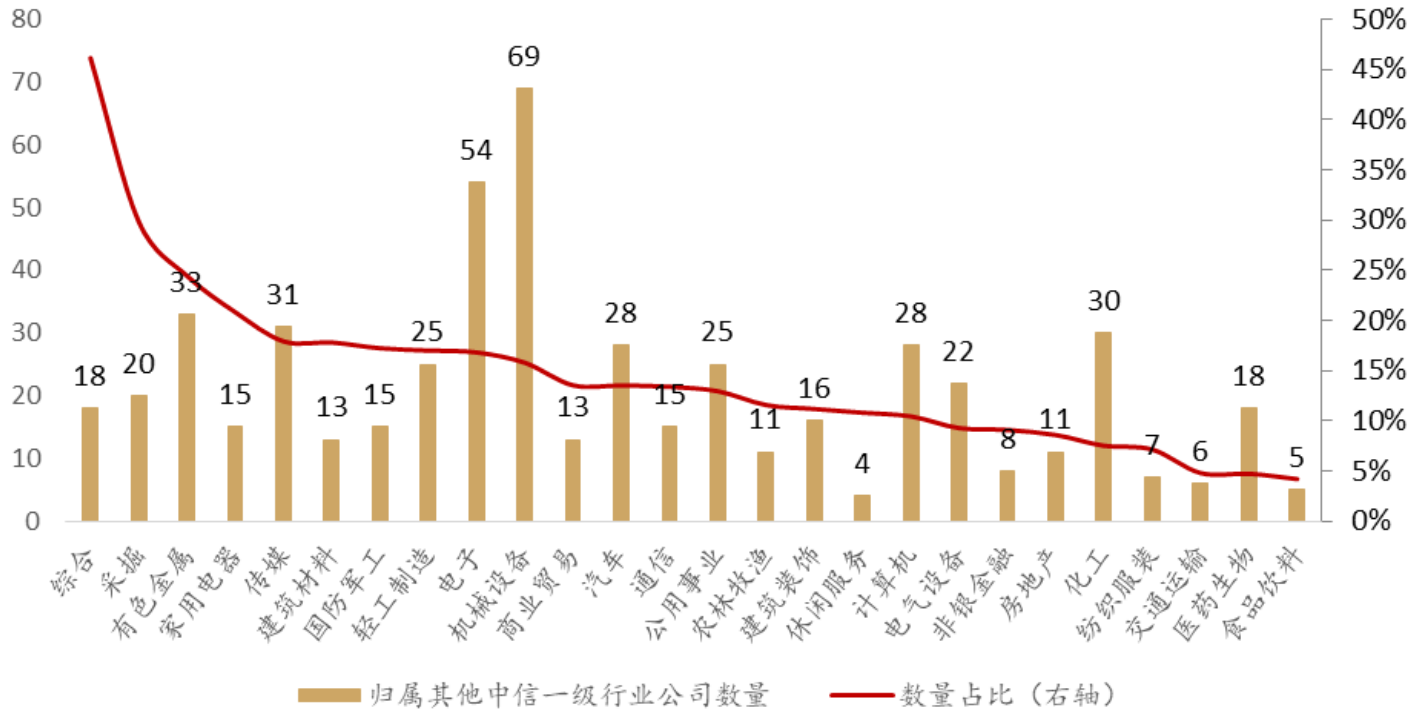
资料来源：华西证券研究所整理

1.4 申万与中信一级行业中约12%的上市公司行业分类不一致

华西证券 HUAXI SECURITIES

- 从分类结果来看，截至2021年6月16日，A股上市公司共4350家，其中以申万一级行业为基准，共有540家上市公司与中信行业分类不一致，占比12.41%。其中申万机械设备行业下共有69家公司在中信行业中划分到除机械外的其他行业，而行业归属不同的公司占比最高的是综合行业，占比达46.15%，其次是采掘和有色金属。

图4：以申万一级行业为基准，分类不一致的公司数量及行业内数量占比



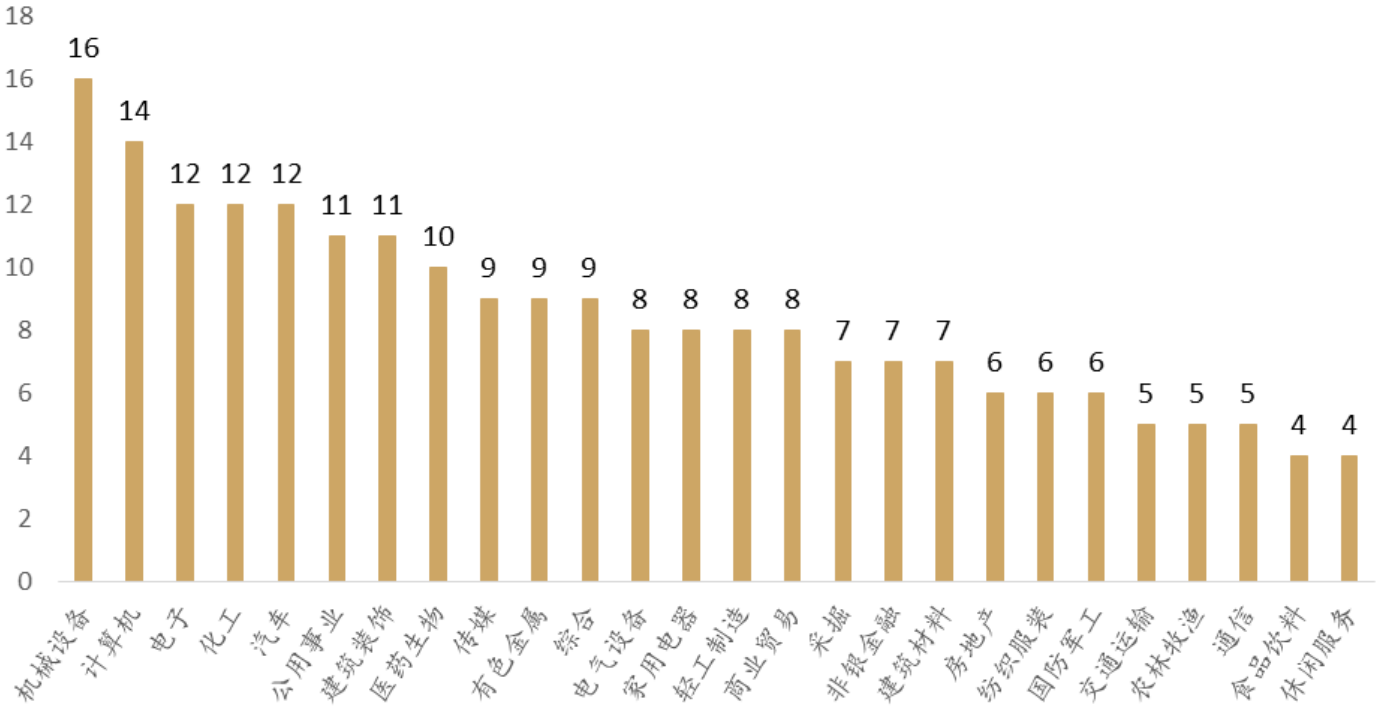
注：为统一口径，中信行业分类下综合金融并入非银金融，基础化工与石油石化合并为化工。

1.5 申万机械设备行业上市公司分散在16个中信一级行业中

华西证券 HUAXI SECURITIES

- 同样以申万一级行业为基准，统计各行业上市公司被归入中信一级行业的数量，其中归属不同行业数量最多的是机械设备，其成分股分散在16个中信一级行业中，其次是计算机、电子、化工、汽车等。

图5：以申万一级行业为基准，成分股归属中信一级行业数量



1.6 沪深300成份股内13家上市公司行业分类差异较大

华西证券 HUAXI SECURITIES

- 主要宽基指数中，沪深300成份股内有13家公司行业分类差异较大，中证500成份股内41家，创业板指成份股内8家。

表1：主要指数成份股中部分行业分类差异较大的公司

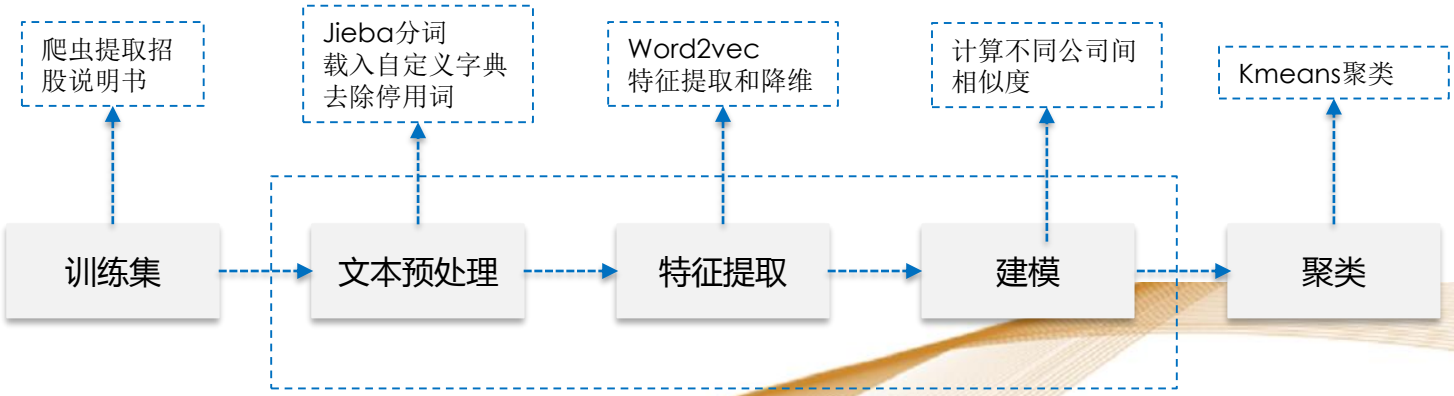
证券代码	证券简称	申万一级行业	中信一级行业	总市值（亿元）	流通市值（亿元）	是否沪深300成份股	是否中证500成份股	是否创业板指成份股
300059.SZ	东方财富	非银金融	计算机	3066.62	2539.35	是		是
601138.SH	工业富联	电子	通信	2473.81	378.29	是		
002607.SZ	中公教育	传媒	消费者服务	1350.66	350.90	是		
600745.SH	闻泰科技	电子	通信	1231.63	854.95	是		
603195.SH	公牛集团	轻工制造	家电	1119.53	140.53	是		
002841.SZ	视源股份	电子	消费者服务	835.04	490.19	是		
688012.SH	中微公司	机械设备	电子	821.98	377.13	是		
600150.SH	中国船舶	国防军工	机械	761.65	416.10	是		
002414.SZ	高德红外	电子	国防军工	616.68	459.66	是		
600176.SH	中国巨石	化工	建材	587.66	587.66	是		
300316.SZ	晶盛机电	电气设备	机械	559.15	525.18		是	是
300012.SZ	华测检测	综合	机械	547.98	497.29			是
002625.SZ	光启技术	汽车	国防军工	510.42	352.58		是	
002532.SZ	天山铝业	机械设备	有色金属	452.16	27.60		是	
002340.SZ	格林美	有色金属	电力设备及新能源	428.13	426.20		是	
600704.SH	物产中大	交通运输	综合	414.09	414.09		是	
003035.SZ	南网能源	公用事业	建筑	376.89	75.38		是	
600482.SH	中国动力	国防军工	电力设备及新能源	370.77	347.91	是		
300699.SZ	光威复材	化工	国防军工	365.70	364.29		是	是
600177.SH	雅戈尔	房地产	纺织服装	342.55	333.20		是	

目录

2.1 基于主营产品相似度的行业分类方法

- 前述投资型行业分类已经使用多年，是否已是最优的行业分类方法？特别是在当下上市公司产业链、产品、竞争环境等都面临大幅变化的情况下，固定行业分类体系是否是最佳方案值得探讨。
- Hoberg and Phillips(2016)使用文本分析的行业分类方法，对美股上市公司进行产品与业务的相似性度量，进而进行重分类研究，发现基于产品相似度的行业分类方法，一定程度上要优于传统固定行业分类方法，而且能更好的刻画行业的动态演变。
- 本文中我们将尝试从产品相似度的角度，对A股市场上市公司进行重分类研究，探讨A股市场基于主营产品相似度聚类的可行性。

图6：基于文本分析的行业聚类过程

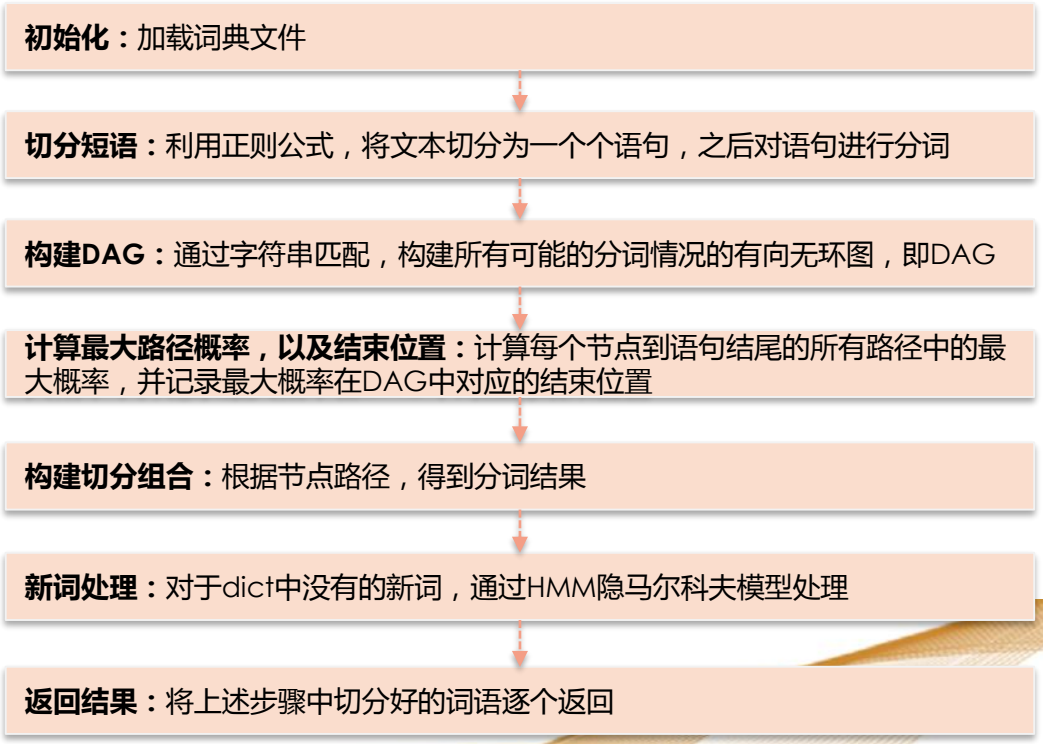


资料来源：华西证券研究所

2.2 中文文本分词工具——jieba分词

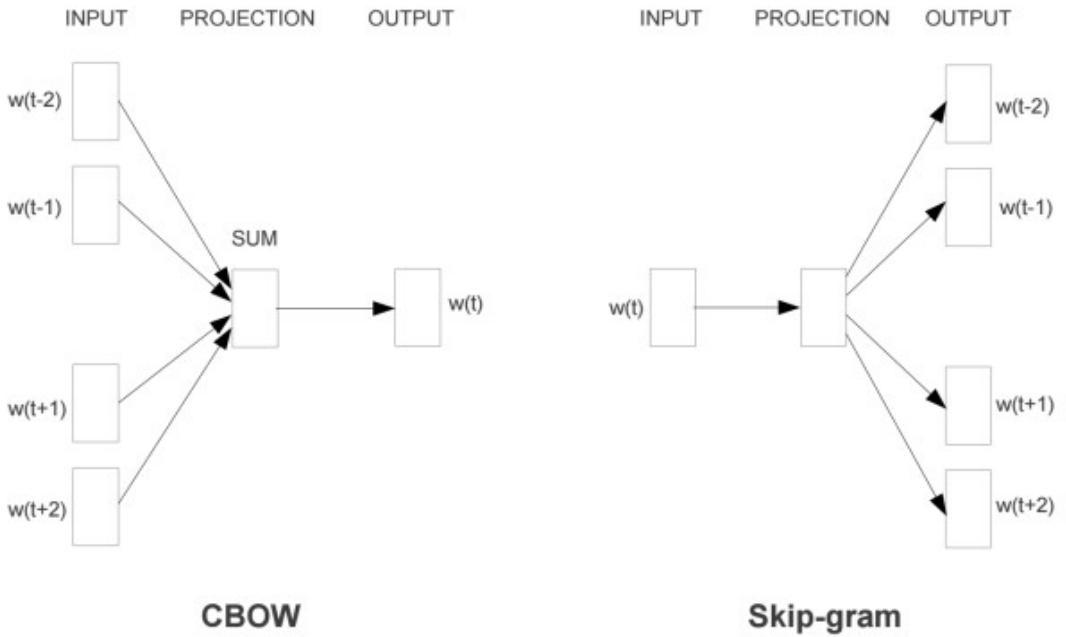
- 分词是文本处理的基础步骤，后续词向量的构建将基于分词的结果。中文文本的分词相较于英文分词难度较大，根据实现原理和特点，中文分词主要分为两类：**基于词典分词算法**和**基于统计的机器学习算法**。
- 本文使用jieba分词工具，其分词过程如下：

图7：jieba分词过程



- Word2vec是Google开源的一款将词表征为实数值向量的高效工具，其训练模式分为CBOW（词袋模型）和Skip-Gram两种。Word2vec通过训练，可以把对文本内容的处理简化为K维向量空间中的向量运算，而向量空间上的相似度可以用来表示文本语义上的相似度。
- Word2vec的实现过程是一个轻量级的神经网络，仅包括输入层、隐藏层和输出层，模型框架根据输入输出的不同，分为CBOW和Skip-gram模型。

图8：Word2vec包括CBOW和Skip-gram两种模型

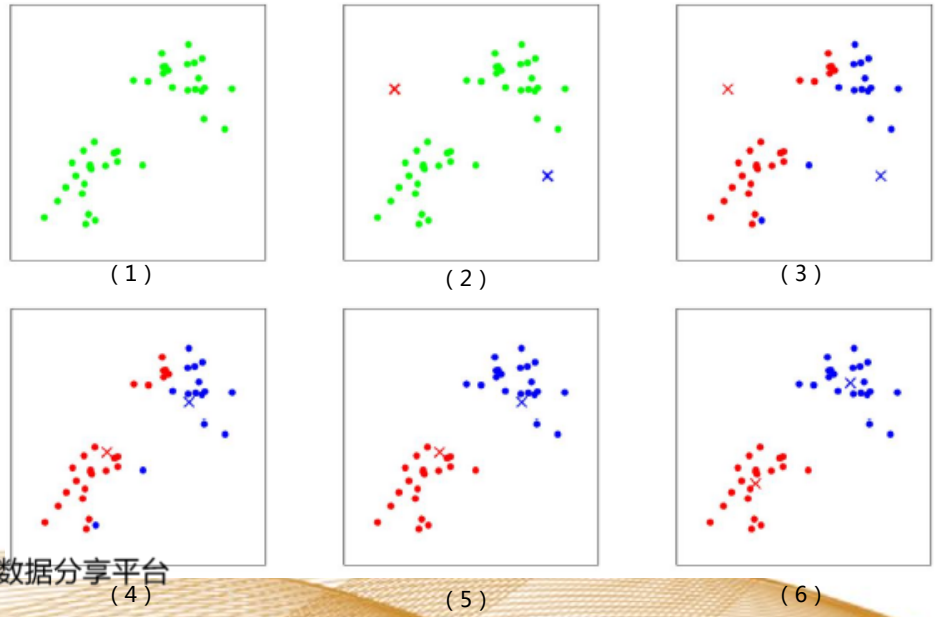


资料来源：华西证券研究所整理

2.4 公司主营相似度计算及基于相似度聚类

- 上文中我们通过Word2vec可以实现词向量的表示，而句子由词构成，因此我们可以将句子表示为矩阵结构，进而对两个句子的二维矩阵进行特征提取和降维，计算得到两个句子的相似度。
- 我们将不同公司的主营产品作为句子抽取关键词，即可用来计算不同公司的主营产品相似度。
- 得到不同公司彼此之间的相似度，即可进行行业聚类，实际投资中，一级行业较为宽泛，二级维度相对适中。目前申万二级行业104个，中信二级行业109个，为做对比我们取整设置聚类数量为100，聚类方法使用K-means聚类。

图9：K-means聚类过程示例



资料来源：华西证券研究所整理

目录

3.1 上证50指数成分股主营产品信息

华西证券 HUAXI SECURITIES

表2：上证50指数成分股主营产品信息

证券代码	证券简称	主营构成-产品1	主营构成-产品2	主营构成-产品3	主营构成-产品4	主营构成-产品5
600000.SH	浦发银行	利息收入:发放贷款及垫款:个人贷款业务	利息收入:发放贷款及垫款:公司贷款业务	利息收入:债权投资	利息收入:其他债权投资	非利息收入:手续费及佣金收入:托管及其他受托业务
600009.SH	上海机场	航空服务	其他			
600016.SH	民生银行	利息收入:发放贷款及垫款:公司贷款业务	利息收入:发放贷款及垫款:个人贷款业务	利息收入:以摊余成本计量的金融资产	利息收入:以公允价值计量且其变动计入当期损益金融资产	非利息收入:手续费及佣金收入:银行卡业务
600028.SH	中国石化	营销及分销	炼油	本部及其他	化工	勘探及开采
600030.SH	中信证券	证券经纪业务	证券投资业务	其他业务	资产管理业务	投资银行业务
600031.SH	三一重工	挖掘机械	混凝土机械	起重机械	桩工机械类	其他业务
600036.SH	招商银行	利息收入:发放贷款及垫款:个人贷款业务	利息收入:发放贷款及垫款:公司贷款业务	利息收入:债权投资	非利息收入:手续费及佣金收入:托管及其他受托业务	非利息收入:手续费及佣金收入:银行卡业务
600048.SH	保利地产	房地产	其他	其他业务		
600050.SH	中国联通	宽带及移动数据服务	数据及其他互联网应用收入	其他业务	通话及月租费收入	增值业务收入
600104.SH	上汽集团	整车业务	零部件业务	劳务及其他	金融业务	贸易
600196.SH	复星医药	药品制造与研发	医学诊断与医疗器械	医疗服务	其他业务	
600276.SH	恒瑞医药	抗肿瘤	麻醉	其他	造影剂	其他业务
600309.SH	万华化学	聚氨酯系列	石化系列	精细化学品及新材料系列	其他	其他业务
600519.SH	贵州茅台	茅台酒	系列酒	其他业务		
600547.SH	山东黄金	外购合质金	黄金	小金条	其他业务	
600570.SH	恒生电子	软件收入	硬件收入	科技园开发销售	其他业务	
600585.SH	海螺水泥	42.5级水泥	建材行业(贸易业务)	其他业务	32.5级水泥	熟料
600588.SH	用友网络	技术服务及培训	软件产品	其他业务	其他	
600690.SH	海尔智家	电冰箱	洗衣机	厨卫电器	空调器	渠道综合服务及其他
600703.SH	三安光电	化合物半导体产品	材料、废料销售	租金、物业、服务收入		
600745.SH	闻泰科技	手机及配件	半导体产品	其他业务	其他	
600837.SH	海通证券	证券经纪业务	证券投资业务	投资银行业务	管理部门及其他	融资租赁业务
600887.SH	伊利股份	液体乳	奶粉及奶制品	冷饮产品系列	其他业务	其他
600918.SH	中泰证券	证券经纪业务	信用业务	总部及其他业务	期货经纪业务	投资银行业务
601012.SH	隆基股份	太阳能组件	硅片及硅棒	电站建设及服务	其他	电力
601066.SH	中信建投	证券投资业务	投资银行业务	证券经纪业务	其他业务	投资管理业务
601088.SH	中国神华	煤炭收入	电力收入	其他业务	运输收入	煤化工收入
601138.SH	工业富联	3C电子产品	其他业务			
601166.SH	兴业银行	利息收入:发放贷款及垫款:个人贷款业务	利息收入:发放贷款及垫款:公司贷款业务	利息收入:债券投资	非利息收入:手续费及佣金收入:财务顾问业务	非利息收入:手续费及佣金收入:银行卡业务
601186.SH	中国铁建	工程承包	物流与物资贸易及其他	房地产	勘察、设计及咨询	工业制造
601211.SH	国泰君安	机构金融:机构投资者服务	证券经纪业务	国际业务	投资银行业务	投资管理
601236.SH	红塔证券	期货经纪业务	证券投资业务	信用交易业务	基金管理业务	证券经纪业务
601288.SH	农业银行	利息收入:发放贷款及垫款:公司贷款业务	利息收入:发放贷款及垫款:票据贴现业务	利息收入:以摊余成本计量的金融投资	利息收入:以公允价值计量且其变动计入其他综合收益的金融	利息收入:存放中央银行款项
601318.SH	中国平安	产险:机动车辆保险	客户贷款	寿险:分红险	寿险:传统险	寿险:长期健康险
601319.SH	中国人保	机动车辆险	个人代理	直接销售	兼业代理	意外伤害及健康险
601336.SH	新华保险	寿险:保险营销员:续期业务	寿险:分红险	寿险:长期健康险	寿险:传统险	寿险:银行保险:首年业务:首年趸缴
601398.SH	工商银行	利息收入:发放贷款及垫款:公司贷款业务	利息收入:发放贷款及垫款:个人贷款业务	利息收入:债券投资	利息收入:存放中央银行款项	利息收入:存放同业和其他金融机构款项
601601.SH	中国太保	个人寿险	寿险:个人寿险:续期业务	寿险:分红险	寿险:传统险	商业车险
601628.SH	中国人寿	人寿保险	寿险业务:续期业务	寿险业务:首年期缴	健康险业务:首年趸缴	健康险业务:续期业务
601668.SH	中国建筑	房屋建筑工程	基础设施建设与投资	房地产开发与投资	其他	设计勘察
601688.SH	华泰证券	证券经纪业务	机构服务	资产管理业务	其他业务	其他
601816.SH	京沪高铁	路网服务	客运业务	其他业务	其他	
601818.SH	光大银行	利息收入:发放贷款及垫款:个人贷款业务	利息收入:发放贷款及垫款:公司贷款业务	利息收入:债券投资	非利息收入:手续费及佣金收入:银行卡业务	利息收入:发放贷款及垫款:融资租赁业务
601857.SH	中国石油	油气销售	炼油与化工	勘探与生产	天然气与管道	其他业务
601888.SH	中国中免	商品贸易-免税商品	商品贸易-有税商品	其他业务		
603160.SH	汇顶科技	指纹识别芯片	电容触控芯片	其他芯片	其他业务	
603259.SH	药明康德	中国区实验室服务	小分子新药工艺研发及生产业务	美国区实验室服务	其他CRO服务	其他业务
603288.SH	海天味业	酱油	蚝油	其他业务	调味酱	
603501.SH	韦尔股份	CMOS图像传感器产品	半导体分销	TDDI	TVS	电源IC
603986.SH	兆易创新	存储芯片销售	微控制器	传感器	技术服务及其他	其他业务

，取自上市公司2020年年报

3.2 上证50指数成分股主营产品关键字抽取

华西证券 HUAXI SECURITIES

➢ 基于上证50指数成分股主营产品构成信息，通过jiaba分词工具抽取其关键词。

表3：上证50指数成分股主营产品信息关键词抽取

证券代码	证券简称	关键词	证券代码	证券简称	关键词
600000.SH	浦发银行	利息收入 垫款 发放贷款 非利息收入 佣金收入 投资	601066.SH	中信建投	投资银行业务 经纪业务 管理业务 投资
600009.SH	上海机场	航空服务	601088.SH	中国神华	煤化工 运输 电力 煤炭
600016.SH	民生银行	利息收入 垫款 发放贷款 金融资产 摊余成本 非利息收入 佣金收入	601138.SH	工业富联	3C 电子产品
600028.SH	中国石化	勘探 炼油 营销 开采	601166.SH	兴业银行	利息收入 非利息收入 佣金收入 垫款 发放贷款 债券投资
600030.SH	中信证券	经纪业务 资产管理 投资银行业务	601186.SH	中国铁建	物资贸易 勘察 承包 物流 房地产
600031.SH	三一重工	机械类 机械 起重机械 混凝土 挖掘	601211.SH	国泰君安	经纪业务 国际业务 投资银行业务 投资管理 金融
600036.SH	招商银行	利息收入 非利息收入 佣金收入 垫款 发放贷款 投资	601236.SH	红塔证券	交易业务 基金管理 经纪业务 期货经纪
600048.SH	保利地产	房地产	601288.SH	农业银行	利息收入 金融投资 垫款 发放贷款 票据贴现 摊余成本 综合收益 中央银行
600050.SH	中国联通	数据服务 月租费 宽带 通话 互联网 数据	601318.SH	中国平安	寿险 分红险 保险客户 健康险 产险 机动车辆
600104.SH	上汽集团	金融业务 零部件 整车	601319.SH	中国人保	健康险 机动车辆 伤害 意外
600196.SH	复星医药	医学诊断 医疗服务 医疗器械 药品	601336.SH	新华保险	寿险 分红险 首年 营销员 健康险 银行
600276.SH	恒瑞医药	造影剂 抗肿瘤 麻醉	601398.SH	工商银行	利息收入 垫款 发放贷款 债券投资 中央银行 同业 金融机构
600309.SH	万华化学	聚氨酯 化学品 精细 石化 材料	601601.SH	中国太保	寿险 分红险 车险
600519.SH	贵州茅台	茅台酒	601628.SH	中国人寿	健康险 寿险 首年 人寿保险
600547.SH	山东黄金	黄金	601668.SH	中国建筑	房地产开发 房屋建筑 勘察 投资
600570.SH	恒生电子	科技园 软件	601688.SH	华泰证券	经纪业务 资产管理
600585.SH	海螺水泥	水泥 熟料 建材行业	601816.SH	京沪高铁	路网 客运
600588.SH	用友网络	软件产品	601818.SH	光大银行	利息收入 垫款 发放贷款 债券投资 非利息收入 佣金收入 融资租赁
600690.SH	海尔智家	厨卫 空调器 电冰箱 洗衣机 电器	601857.SH	中国石油	油气 勘探 炼油 管道 天然气
600703.SH	三安光电	废料 半导体 化合物 物业 材料	601888.SH	中国中免	免税商品 有税
600745.SH	闻泰科技	配件 半导体 手机	603160.SH	汇顶科技	芯片 指纹识别 触控 电容
600837.SH	海通证券	经纪业务 投资银行业务 融资租赁	603259.SH	药明康德	实验室 CRO 中国区 分子
600887.SH	伊利股份	冷饮 奶制品 奶粉	603288.SH	海天味业	调味酱 蚝油 酱油
600918.SH	中泰证券	经纪业务 信用业务 投资银行业务 期货经纪	603501.SH	韦尔股份	CMOS TDDITVS IC 传感器 半导体 图像 电源
601012.SH	隆基股份	硅片 电站 电力	603986.SH	兆易创新	存储芯片 微控制器 传感器

，取自上市公司2020年年报

3.3 基于产品角度刻画的上证50成份股相似度矩阵

华西证券 HUAXI SECURITIES

- 基于上述关键词计算不同公司之间的相似度，从结果来看，基于主营产品文本信息的相似度矩阵对于不同行业公司具有相对较好的区分能力。

表4：上证50指数成分股主营产品相似度矩阵

股票名称	浦发银行	民生银行	招商银行	兴业银行	农业银行	工商银行	光大银行	中信证券	海通证券	中泰证券	中信建投	国泰君安	红塔证券	华泰证券	中国平安	中国人保	新华保险	中国太保	中国人寿	三安光电	闻泰科技	工业富联	汇顶科技	韦尔股份	兆易创新	贵州茅台	伊利股份	海天味业	复星医药	恒瑞医药	药明康德	中国神华	中国石油	中国石化	万华化学	恒生电子	用友网络	中国铁建	中国建筑	上海机场	京沪高铁	山东黄金	三一重工	保利地产	隆基股份	海尔智家	海螺水泥	上汽集团	中国联通	中国中免
浦发银行	1.00	0.91	0.97	0.97	0.95	0.93	1.00	0.31	0.31	0.35	0.39	0.33	0.35	0.32	0.28	0.20	0.29	0.23	0.29	0.13	0.13	0.10	0.16	0.10	0.15	-0.06	0.16	0.16	0.11	0.12	0.12	0.06	0.08	0.09	0.00	0.04	0.02	0.16	0.24	0.07	0.12	0.13	0.07	0.11	0.05	0.11	0.03	0.18	0.19	0.16
民生银行	0.91	1.00	0.87	0.87	0.93	0.86	0.88	0.24	0.28	0.29	0.26	0.24	0.24	0.26	0.25	0.17	0.23	0.18	0.24	0.15	0.06	0.08	0.14	0.10	0.12	-0.04	0.11	0.11	0.09	0.10	0.10	0.03	0.06	0.08	0.01	0.03	0.06	0.12	0.17	0.03	0.10	0.12	0.07	0.10	0.03	0.07	0.04	0.14	0.14	0.13
招商银行	0.97	0.87	1.00	1.00	0.92	0.87	0.97	0.32	0.31	0.34	0.39	0.32	0.36	0.36	0.28	0.20	0.32	0.23	0.30	0.14	0.13	0.09	0.17	0.11	0.13	-0.06	0.16	0.13	0.11	0.12	0.12	0.09	0.07	0.09	0.01	0.05	0.02	0.18	0.22	0.07	0.13	0.13	0.06	0.11	0.05	0.10	0.04	0.17	0.19	0.16
兴业银行	0.97	0.87	1.00	1.00	0.92	0.89	0.97	0.31	0.32	0.34	0.36	0.31	0.36	0.35	0.28	0.20	0.32	0.23	0.30	0.15	0.14	0.10	0.17	0.12	0.15	-0.06	0.16	0.15	0.11	0.13	0.12	0.09	0.06	0.08	0.00	0.03	0.05	0.18	0.16	0.03	0.13	0.15	0.05	0.08	0.06	0.10	0.04	0.17	0.19	0.16
农业银行	0.95	0.93	0.92	0.92	1.00	0.90	0.90	0.25	0.24	0.31	0.24	0.24	0.26	0.26	0.26	0.17	0.26	0.17	0.22	0.13	0.08	0.09	0.16	0.16	0.11	-0.08	0.09	0.07	0.09	0.08	0.11	0.09	0.06	0.08	0.05	0.05	0.02	0.15	0.19	0.02	0.09	0.15	0.06	0.13	0.06	0.09	0.05	0.15	0.16	0.11
工商银行	0.93	0.86	0.87	0.89	0.90	1.00	1.00	0.29	0.30	0.35	0.32	0.33	0.31	0.28	0.24	0.17	0.24	0.20	0.22	0.20	0.13	0.09	0.20	0.23	0.15	-0.07	0.10	0.10	0.12	0.16	0.15	0.05	0.09	0.13	0.01	0.02	0.05	0.12	0.21	-0.01	0.12	0.17	0.08	0.08	0.10	0.09	0.04	0.17	0.14	0.12
光大银行	1.00	0.88	0.97	0.97	0.90	1.00	1.00	0.31	0.36	0.37	0.35	0.34	0.35	0.33	0.27	0.20	0.27	0.22	0.27	0.20	0.12	0.10	0.20	0.15	0.16	-0.05	0.14	0.16	0.11	0.14	0.12	0.05	0.09	0.12	0.00	0.04	0.05	0.14	0.21	0.06	0.13	0.17	0.07	0.12	0.08	0.10	0.04	0.17	0.18	0.14
中信证券	0.31	0.24	0.32	0.31	0.25	0.29	0.31	1.00	0.91	0.82	0.76	0.79	0.77	1.00	0.23	0.11	0.26	0.29	0.35	0.14	0.17	0.02	0.15	0.12	0.13	-0.02	0.10	0.15	0.19	0.18	0.27	0.09	0.08	0.26	0.08	0.14	0.13	0.22	0.28	0.13	0.29	0.19	0.07	0.15	0.11	0.06	0.06	0.22	0.20	0.21
海通证券	0.31	0.28	0.31	0.32	0.24	0.30	0.36	0.91	1.00	0.95	0.76	0.83	0.74	0.82	0.22	0.15	0.24	0.26	0.26	0.20	0.19	0.06	0.15	0.13	0.13	-0.01	0.11	0.16	0.19	0.21	0.27	0.13	0.12	0.27	0.05	0.10	0.12	0.23	0.20	0.12	0.29	0.20	0.09	0.17	0.11	0.09	0.07	0.22	0.21	0.18
中泰证券	0.35	0.29	0.34	0.34	0.31	0.35	0.37	0.82	0.95	1.00	0.91	1.00	0.82	0.76	0.25	0.19	0.26	0.27	0.28	0.16	0.17	0.13	0.16	0.14	0.13	-0.04	0.10	0.15	0.22	0.20	0.29	0.13	0.12	0.32	0.09	0.09	0.13	0.22	0.21	0.14	0.28	0.17	0.15	0.16	0.12	0.11	0.07	0.28	0.22	0.26
中信建投	0.39	0.26	0.39	0.36	0.24	0.32	0.35	0.76	0.76	0.91	1.00	0.86	1.00	0.81	0.23	0.16	0.28	0.27	0.31	0.18	0.19	0.13	0.18	0.13	0.13	-0.06	0.10	0.12	0.25	0.17	0.30	0.13	0.16	0.35	0.08	0.11	0.12	0.20	0.35	0.08	0.30	0.17	0.08	0.14	0.13	0.08	0.08	0.31	0.29	0.21
国泰君安	0.33	0.24	0.32	0.31	0.24	0.33	0.34	0.79	0.83	1.00	0.86	1.00	0.73	0.73	0.23	0.14	0.26	0.24	0.26	0.18	0.15	0.08	0.14	0.16	0.12	-0.02	0.07	0.13	0.23	0.20	0.31	0.13	0.16	0.36	0.11	0.17	0.12	0.25	0.32	0.17	0.24	0.17	0.13	0.16	0.11	0.08	0.09	0.22	0.23	0.22
红塔证券	0.35	0.24	0.36	0.36	0.26	0.31	0.35	0.77	0.74	0.82	1.00	0.73	1.00	0.87	0.25	0.15	0.28	0.29	0.32	0.19	0.20	0.05	0.18	0.15	0.12	-0.04	0.11	0.12	0.26	0.15	0.30	0.16	0.10	0.33	0.09	0.12	0.09	0.25	0.24	0.17	0.30	0.20	0.18	0.14	0.13	0.12	0.09	0.25	0.29	0.27
华泰证券	0.32	0.26	0.36	0.35	0.26	0.28	0.33	1.00	0.82	0.76	0.81	0.73	0.87	1.00	0.24	0.14	0.28	0.32	0.36	0.14	0.18	0.02	0.17	0.10	0.13	-0.02	0.12	0.13	0.21	0.16	0.28	0.13	0.07	0.27	0.08	0.14	0.10	0.23	0.31	0.18	0.35	0.22	0.08	0.17	0.12	0.08	0.07	0.26	0.28	0.21
中国平安	0.28	0.25	0.28	0.28	0.26	0.24	0.27	0.23	0.22	0.25	0.23	0.23	0.25	0.24	1.00	0.67	0.85	0.74	0.81	0.14	0.18	0.20	0.14	0.17	0.15	0.11	0.21	0.17	0.19	0.18	0.28	0.16	0.13	0.17	0.14	0.10	0.13	0.16	0.17	0.16	0.24	0.11	0.19	0.03	0.08	0.21	0.08	0.19	0.22	0.17
中国人保	0.20	0.17	0.20	0.20	0.17	0.17	0.20	0.11	0.15	0.19	0.16	0.14	0.15	0.14	0.67	1.00	0.53	0.38	0.56	0.15	0.21	0.33	0.16	0.19	0.15	0.00	0.25	0.11	0.34	0.25	0.15	0.37	0.21	0.13	0.23	0.13	0.11	0.18	0.36	0.17	0.29	0.04	0.48	0.04	0.10	0.32	0.09	0.25	0.19	0.14
中国太保	0.23	0.18	0.23	0.23	0.17	0.20	0.22	0.29	0.26	0.27	0.27	0.24	0.29	0.32	0.74	0.38	0.70	1.00	0.67	0.11	0.20	0.19	0.18	0.16	0.23	0.19	0.20	0.20	0.27	0.21	0.20	0.18	0.14	0.18	0.11	0.15	0.17	0.17	0.11	0.15	0.29	0.18	0.16	0.03	0.09	0.21	0.10	0.19	0.24	0.27
中国人寿	0.29	0.24	0.30	0.30	0.22	0.22	0.27	0.35	0.26	0.28	0.31	0.26	0.32	0.36	0.81	0.56	0.95	0.67	1.00	0.15	0.20	0.06	0.11	0.16	0.11	0.13	0.21	0.15	0.20	0.15	0.26	0.10	0.09	0.20	0.11	0.10	0.10	0.15	0.14	0.16	0.22	0.15	0.10	0.06	0.04	0.21	0.04	0.13	0.24	0.20
三安光电	0.13	0.15	0.14	0.15	0.13	0.20	0.20	0.14	0.20	0.16	0.18	0.18	0.19	0.14	0.14	0.15	0.14	0.11	0.15	1.00	0.62	0.32	0.40	0.79	0.33	0.09	0.16	0.19	0.28	0.32	0.40	0.28	0.22	0.22	0.39	0.25	0.23	0.29	0.33	0.08	0.21	0.14	0.27	0.12	0.33	0.26	0.29	0.37	0.20	0.14
闻泰科技	0.13	0.06	0.13	0.14	0.08	0.13	0.12	0.17	0.19	0.17	0.19	0.15	0.20	0.18	0.18	0.21	0.13	0.20	0.20	0.62	1.00	0.54	0.51	0.75	0.44	0.10	0.31	0.23	0.24	0.23	0.25	0.30	0.25	0.20	0.28	0.32	0.25	0.18	0.16	0.05	0.10	0.13	0.29	0.03	0.40	0.38	0.19	0.54	0.33	0.19
工业富联	0.10	0.08	0.09	0.10	0.09	0.09	0.10	0.02	0.06	0.13	0.13	0.08	0.05	0.02	0.20	0.33	0.12	0.19	0.06	0.32	0.54	1.00	0.38	0.40	0.42	0.02	0.28	0.19	0.29	0.20	0.14	0.26	0.18	0.15	0.38	0.26	0.26	0.08	0.10	0.05	0.10	0.11	0.23	0.07	0.30	0.39	0.20	0.49	0.24	0.22
汇顶科技	0.16	0.14	0.17	0.17	0.16	0.20	0.20	0.15	0.15	0.16	0.18	0.14	0.18	0.17	0.14	0.16	0.14	0.18	0.11	0.40	0.51	0.38	1.00	0.59	0.60	0.05	0.15	0.19	0.16	0.32	0.26	0.18	0.13	0.16	0.22	0.39	0.28	0.10	0.04	-0.02	0.09	0.07	0.12	-0.07	0.31	0.34	0.09	0.33	0.33	0.05
韦尔股份	0.10	0.10	0.11	0.12	0.16	0.23	0.15	0.12	0.13	0.14	0.13	0.16	0.15	0.10	0.17	0.19	0.15	0.16	0.16	0.79	0.75	0.40	0.59	1.00	0.63	0.04	0.09	0.15	0.27	0.27	0.33	0.33	0.21	0.30	0.40	0.24	0.19	0.19	0.00	0.22	0.07	0.25	0.01	0.38	0.37	0.16	0.42	0.41	0.09
兆易创新	0.15	0.12	0.13	0.15	0.11	0.15	0.16	0.13	0.13	0.13	0.13	0.12	0.12	0.13	0.15	0.15	0.14	0.23	0.11	0.33	0.44	0.42	0.60	0.63	1.00	0.02	0.12	0.17	0.15	0.31	0.18	0.20	0.16	0.09	0.24	0.43	0.32	0.04	0.04	-0.04	0.11	0.05	0.15	-0.01	0.28	0.35	0.10	0.41	0.30	0.09
贵州茅台	-0.06	-0.04	-0.06	-0.06	-0.08	-0.07	-0.05	-0.02	-0.01	-0.04	-0.06	-0.02	-0.04	-0.02	0.11	0.00	0.11	0.19	0.13	0.09	0.10	0.02	0.05	0.04	0.02	1.00	0.31	0.39	0.10	0.13	0.01	0.14	0.11	0.08	0.14	-0.02	0.12	0.04	0.01	0.01	0.02	0.23	0.08	-0.03	0.09	0.22	0.18	0.05	-0.02	0.03
伊利股份	0.16	0.11	0.16	0.16	0.09	0.10	0.14	0.10	0.11	0.10	0.10	0.07	0.11	0.12	0.21	0.25	0.16	0.20	0.21	0.16	0.31	0.28	0.15	0.09	0.12	0.31	1.00	0.59	0.30	0.30	0.16	0.26	0.24	0.23	0.25	0.03	0.09	0.12	0.08	0.07	0.13	0.25	0.20	0.07	0.15	0.35	0.33	0.24	0.08	0.33
海天味业	0.16	0.11	0.13	0.15	0.07	0.10	0.16	0.15	0.16	0.15	0.12	0.13	0.12	0.13	0.17	0.11	0.15	0.20	0.15	0.19	0.23	0.19	0.19	0.15	0.17	0.39	0.59	1.00	0.18	0.38	0.16	0.17	0.22	0.21	0.21	0.12	0.19	0.09	0.10	-0.02	0.12	0.22	0.17	0.00	0.17	0.31	0.34	0.20	0.11	0.19
复星医药	0.11	0.09	0.11	0.11	0.09	0.12	0.11	0.19	0.19	0.22	0.25	0.23	0.26	0.21	0.19	0.34	0.15	0.27	0.20	0.28	0.24	0.29	0.16	0.27	0.15	0.10	0.30	0.18	1.00	0.37	0.35	0.29	0.21	0.24	0.44	0.22	0.32	0.31	0.24	0.18	0.31	0.12	0.30	0.14	0.21	0.29	0.17	0.26	0.26	0.30
恒瑞医药	0.12	0.10	0.12	0.13	0.08	0.16	0.14	0.18	0.21	0.20	0.17	0.20	0.15	0.16	0.18	0.25	0.15	0.21	0.15	0.32	0.23	0.20	0.32	0.27	0.31	0.13	0.30	0.38	0.37	1.00	0.35	0.17	0.19	0.24	0.24	0.18	0.16	0.08	0.08	-0.03	0.12	0.07	0.15	-0.03	0.21	0.28	0.15	0.19	0.14	0.07
药明康德	0.12	0.10	0.12	0.12	0.11	0.15	0.12	0.27	0.27	0.29	0.30	0.31	0.30	0.28	0.28	0.15	0.17	0.20	0.26	0.40	0.25	0.14	0.26	0.27	0.18	0.01	0.16	0.16	0.35	0.35	1.00	0.24	0.32	0.33	0.20	0.20	0.14	0.31	0.26	0.01	0.24	0.04	0.14	-0.01	0.23	0.16	0.11	0.34	0.27	0.20
中国神华	0.06	0.03	0.09	0.09	0.09	0.05	0.05	0.09	0.13	0.13	0.13	0.13	0.16	0.13	0.16	0.37	0.11	0.18	0.10	0.28	0.30	0.26	0.18	0.33	0.20	0.14	0.26	0.17	0.29	0.17	0.24	1.00	0.48	0.42	0.51	0.19	0.14	0.55	0.22	0.13	0.47	0.21	0.37	0.18	0.65	0.31	0.42	0.35	0.31	0.27
中国石油	0.08	0.06	0.07	0.06	0.06	0.09	0.09	0.08	0.12	0.12	0.16	0.16	0.10	0.07	0.13	0.21	0.10	0.14	0.09	0.22	0.25	0.18	0.13	0.33	0.16	0.11	0.24	0.22	0.21	0.19	0.32	0.48	1.00	1.00	0.54	0.25	0.18	0.37	0.41	-0.01	0.28	0.16	0.35	0.17	0.36	0.24	0.30	0.26	0.24	0.12
中国石化	0.09	0.08	0.09	0.08	0.08	0.13	0.12	0.26	0.27	0.32	0.35	0.36	0.33	0.27	0.17	0.13	0.28	0.18	0.20	0.22	0.20	0.15	0.16	0.21	0.09	0.08	0.23	0.21	0.24	0.24	0.33	0.42	1.00	1.00	0.53	0.23	0.18	0.40	0.41	0.00	0.26	0.16	0.23	0.09	0.29	0.14	0.27	0.24	0.30	0.19
万华化学	0.00	0.01	0.01	0.00	0.05	0.01	0.00	0.08	0.05	0.09	0.08	0.11	0.09	0.08	0.14	0.23	0.08	0.11	0.11	0.39	0.28	0.38	0.22	0.30	0.24	0.14	0.25	0.21	0.44	0.24	0.20	0.51	0.54	0.53	1.00	0.24	0.13	0.34	0.22	0.07	0.12	0.08	0.35	0.07	0.32	0.21	0.34	0.31	0.15	0.18
恒生电子	0.04	0.03	0.05	0.03	0.05	0.02	0.04	0.14	0.10	0.09	0.11	0.17	0.12	0.14	0.10	0.13	0.09	0.15	0.10	0.25	0.32	0.26	0.39	0.40	0.43	-0.02	0.03	0.12	0.22	0.18	0.20	0.19	0.25	0.23	0.24	1.00	0.36	0.27	0.29	0.17	0.18	-0.05	0.20	0.23	0.20	0.13	0.03	0.28	0.34	0.14
用友网络	0.02	0.06	0.02	0.05	0.02	0.05	0.05	0.13	0.12	0.13	0.12	0.12	0.09	0.10	0.13	0.11	0.11	0.17	0.10	0.23	0.25	0.26	0.28	0.24	0.32	0.12	0.09	0.19	0.32	0.16	0.14	0.14	0.18	0.18	0.13	0.36	1.00	0.12	0.11	0.00	0.16	0.09	0.08	0.15	0.19	0.10	0.07	0.29	0.22	0.18
中国铁建	0.16	0.12	0.18	0.18	0.15	0.12	0.14	0.22	0.23	0.22	0.20	0.25	0.25	0.23	0.16	0.18	0.21	0.17	0.15	0.29	0.18	0.08	0.10	0.19	0.04	0.04	0.12	0.09	0.31	0.08	0.31	0.55	0.37	0.40	0.34	0.27	0.12	1.00	0.84	0.26	0.40	0.17	0.28	0.37	0.29	0.12	0.26	0.24	0.27	0.22
中国建筑	0.24	0.17	0.22	0.16	0.19	0.21	0.21	0.28	0.20	0.21	0.35	0.32	0.24	0.31	0.17	0.36	0.14	0.11	0.14	0.33	0.16	0.10	0.04	0.19	0.04	0.01	0.08	0.10	0.24	0.08	0.26	0.22	0.41	0.41	0.22	0.29	0.11	0.84	1.00	0.17	0.31	0.14	0.31	0.34	0.23	0.12	0.28	0.18	0.19	0.14
上海机场	0.07	0.03	0.07	0.03	0.02	-0.01	0.06	0.13	0.12	0.14	0.08	0.17	0.17	0.18	0.16	0.17	0.10	0.15	0.16	0.08	0.05	0.05	-0.02	0.00	-0.04	0.01	0.07	-0.02	0.18	-0.03	0.01	0.13	-0.01	0.00	0.07	0.17	0.00	0.26	0.17	1.00	0.25	0.03	0.10	0.36	0.00	0.16	-0.02	0.07	0.10	0.19
京沪高铁	0.12	0.10	0.13	0.13	0.09	0.12	0.13	0.29	0.29	0.28	0.30	0.24	0.30	0.35	0.24	0.29	0.20	0.29	0.22	0.21	0.10	0.10	0.09	0.22	0.11	0.02	0.13	0.12	0.31	0.12	0.24	0.47	0.28	0.26	0.12	0.18	0.16	0.40	0.31	0.25	1.00	0.05	0.16	0.11	0.28	0.10	0.15	0.25	0.33	0.27
山东黄金	0.13	0.12	0.13	0.15	0.15	0.17	0.17	0.19	0.20	0.17	0.17	0.17	0.20	0.22	0.11	0.04	0.11	0.18	0.15	0.14	0.13	0.11	0.07	0.07	0.05	0.23	0.25	0.22	0.12	0.07	0.04	0.21	0.16	0.16	0.08	-0.05	0.09	0.17	0.14	0.03	0.05	1.00	0.06	0.25	0.15	0.05	0.19	0.08	0.05	0.18
三一重工	0.07	0.07	0.06	0.05	0.06	0.08	0.07	0.07	0.09	0.15	0.08	0.13	0.18	0.08	0.19	0.48	0.12	0.16	0.10	0.27	0.29	0.23	0.12	0.25	0.15	0.08	0.20	0.17	0.30	0.15	0.14	0.37	0.35	0.23	0.35	0.20	0.08	0.28	0.31	0.10	0.16	0.06	1.00	0.09	0.27	0.36	0.42	0.30	0.20	0.10
保利地产	0.11	0.10	0.11	0.08	0.13	0.08	0.12	0.15	0.17	0.16	0.14	0.16	0.14	0.17	0.03	0.04	0.02	0.03	0.06	0.12	0.03	0.07	-0.07	0.01	-0.01	-0.03	0.07	0.00	0.14	-0.03	-0.01	0.18	0.17	0.09	0.07	0.23	0.15	0.37	0.34	0.36	0.11	0.25	0.09	1.00	0.13	0.14	0.07	0.03	0.06	0.18
隆基股份	0.05	0.03	0.05	0.06	0.06	0.10	0.08	0.11	0.11	0.12	0.13	0.11	0.13	0.12	0.08	0.10	0.04	0.09	0.04	0.40	0.38	0.30	0.22	0.38	0.28	0.09	0.15	0.17	0.21	0.21	0.23	0.65	0.36	0.29	0.32	0.20	0.19	0.29	0.23	0.00	0.28	0.15	0.27	0.13	1.00	0.28	0.38	0.36	0.24	0.10
海尔智家	0.11	0.07	0.10	0.10	0.09	0.09	0.10	0.06	0.09	0.11	0.08	0.08	0.12	0.08	0.21	0.32	0.15	0.21	0.21	0.26	0.38	0.39	0.34	0.37	0.35	0.22	0.35	0.31	0.29	0.28	0.16	0.31	0.24	0.14	0.21	0.13	0.10	0.12	0.12	0.16	0.10	0.05	0.36	0.14	0.28	1.00	0.17	0.37	0.19	0.13
海螺水泥	0.03	0.04	0.04	0.04	0.05	0.04	0.04	0.06	0.07	0.07	0.08	0.09	0.09	0.07	0.08	0.09	0.07	0.10	0.04	0.29	0.19	0.20	0.09	0.16	0.10	0.18	0.33	0.34	0.17	0.15	0.11	0.42	0.30	0.27	0.34	0.03	0.07	0.26	0.28	-0.02	0.15	0.19	0.42	0.07	0.38	0.17	1.00	0.23	0.13	0.12
上汽集团	0.18	0.14	0.17	0.17	0.15	0.17	0.17	0.22	0.22	0.28	0.31	0.22	0.25	0.26	0.19	0.25	0.14	0.19	0.13	0.37	0.54	0.49	0.33	0.42	0.41	0.05	0.24	0.20	0.26	0.19	0.34	0.35	0.26	0.24	0.31	0.28	0.29	0.24	0.18	0.07	0.25	0.08	0.30	0.03	0.36	0.37	0.23	1.00	0.24	0.20
中国联通	0.19	0.14	0.19	0.19	0.16	0.14	0.18	0.20	0.21	0.22	0.29	0.23	0.29	0.28	0.22	0.19	0.25	0.24	0.24	0.20	0.33	0.24	0.33	0.41	0.30	-0.02	0.08	0.11	0.26	0.14	0.27	0.31	0.24	0.30	0.15	0.34	0.22	0.27	0.19	0.10	0.33	0.05	0.20	0.06	0.24	0.19	0.13	0.24	1.00	0.12
中国中免	0.16	0.13	0.16	0.16	0.11	0.12	0.14	0.21	0.18	0.26	0.21	0.22	0.27	0.21	0.17	0.14	0.18	0.27	0.20	0.14	0.19	0.22	0.05	0.09	0.09	0.03	0.33	0.19	0.30	0.07	0.20	0.27	0.12	0.19	0.18	0.14	0.18	0.22	0.14	0.19	0.27	0.18	0.10	0.18	0.10	0.13	0.12	0.20	0.12	1.00

3.4 基于主营产品角度寻找相似度较高的上市公司

华西证券 HUAXI SECURITIES

- 以中芯国际为例，其主营业务主要为集成电路圆晶代工，与其主营产品相似度较高的公司如下：

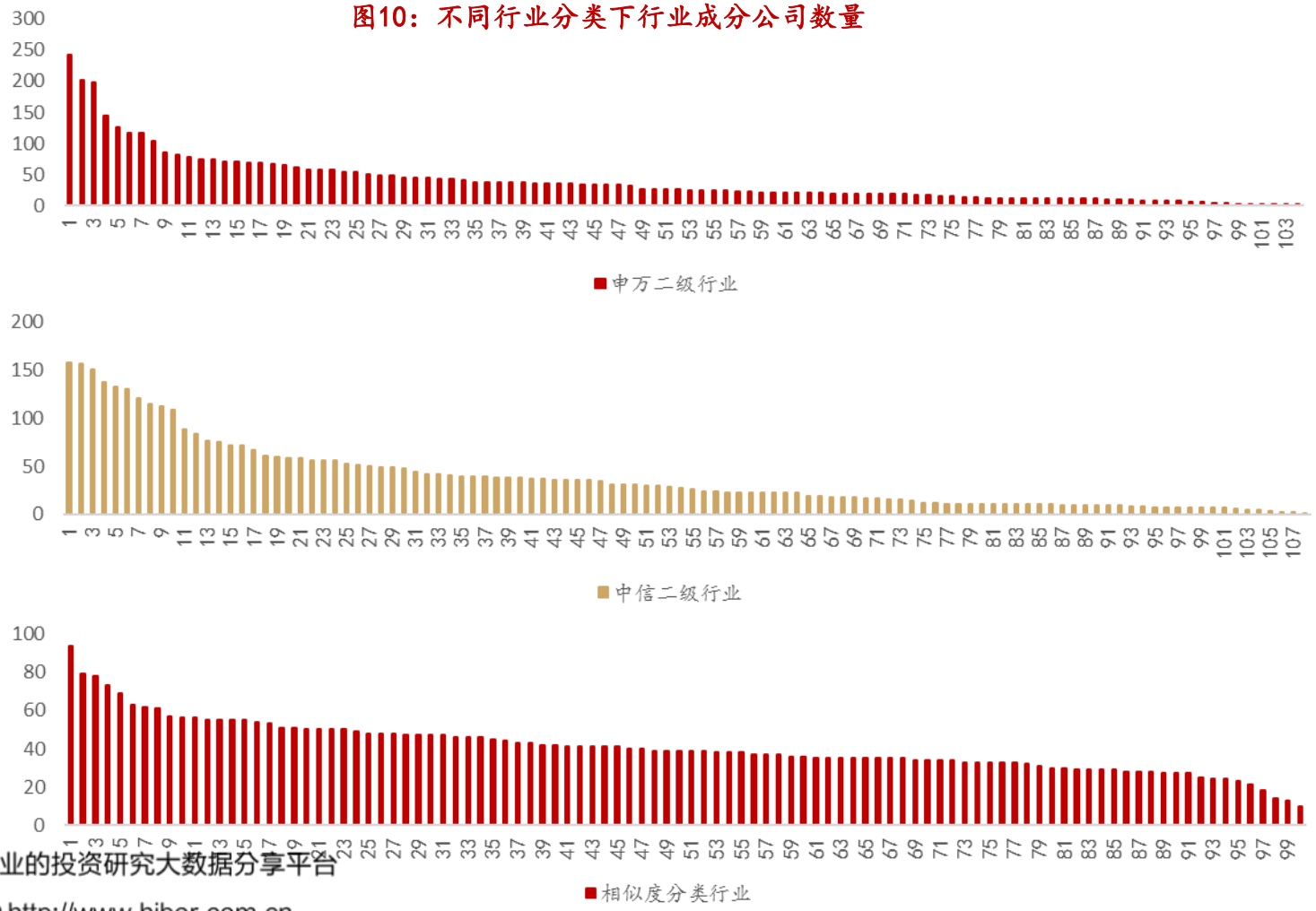
表5：与中芯国际主营业务相似度较高的公司

股票代码	股票名称	产品相似度	申万二级行业	申万三级行业	中信二级行业	中信三级行业
688981.SH	中芯国际	100.00%	半导体	集成电路	半导体	集成电路
002156.SZ	通富微电	77.58%	半导体	集成电路	半导体	集成电路
002185.SZ	华天科技	72.44%	半导体	集成电路	半导体	集成电路
300101.SZ	振芯科技	72.44%	航天装备Ⅱ	航天装备Ⅲ	其他军工Ⅱ	其他军工Ⅲ
600460.SH	士兰微	72.44%	半导体	集成电路	半导体	分立器件
688536.SH	思瑞浦	70.35%	半导体	集成电路	半导体	集成电路
600584.SH	长电科技	65.42%	半导体	集成电路	半导体	集成电路
600536.SH	中国软件	64.50%	计算机应用	IT服务	计算机软件	行业应用软件
688588.SH	凌志软件	64.50%	计算机应用	软件开发	计算机软件	行业应用软件
002180.SZ	纳思达	64.08%	计算机设备Ⅱ	计算机设备Ⅲ	半导体	集成电路
300708.SZ	聚灿光电	64.08%	光学光电子	LED	光学光电	LED
603005.SH	晶方科技	64.08%	半导体	集成电路	半导体	集成电路
605111.SH	新洁能	64.08%	半导体	分立器件	半导体	分立器件
688008.SH	澜起科技	64.08%	半导体	集成电路	半导体	集成电路
688018.SH	乐鑫科技	64.08%	半导体	集成电路	计算机设备	专用计算机设备
688368.SH	晶丰明源	64.08%	半导体	集成电路	半导体	集成电路
300456.SZ	赛微电子	63.73%	半导体	集成电路	其他电子零组件Ⅱ	其他电子零组件Ⅲ
300139.SZ	晓程科技	62.83%	半导体	集成电路	半导体	集成电路
300373.SZ	扬杰科技	62.49%	半导体	分立器件	半导体	分立器件
605358.SH	立昂微	62.49%	半导体	集成电路	半导体	集成电路
600171.SH	上海贝岭	60.47%	半导体	集成电路	半导体	集成电路
002449.SZ	国星光电	59.77%	光学光电子	LED	光学光电	LED
002008.SZ	大族激光	59.63%	其他电子Ⅱ	其他电子Ⅲ	消费电子	消费电子设备
002432.SZ	九安医疗	59.37%	医疗器械Ⅱ	医疗器械Ⅲ	其他医药医疗	医疗器械
000625.SZ	长安汽车	59.36%	汽车整车	乘用车	乘用车Ⅱ	乘用车Ⅲ
600198.SH	*ST大唐	58.98%	半导体	集成电路	通信设备制造	系统设备
603893.SH	瑞芯微	58.98%	半导体	集成电路	半导体	集成电路

3.5 基于产品相似度的行业分类公司数量相对均衡

- 我们对全部上市公司进行基于产品相似度的行业重分类，行业个数设定为100，如下图所示，相较于传统行业分类（二级维度），各行业数量相对更加均衡。

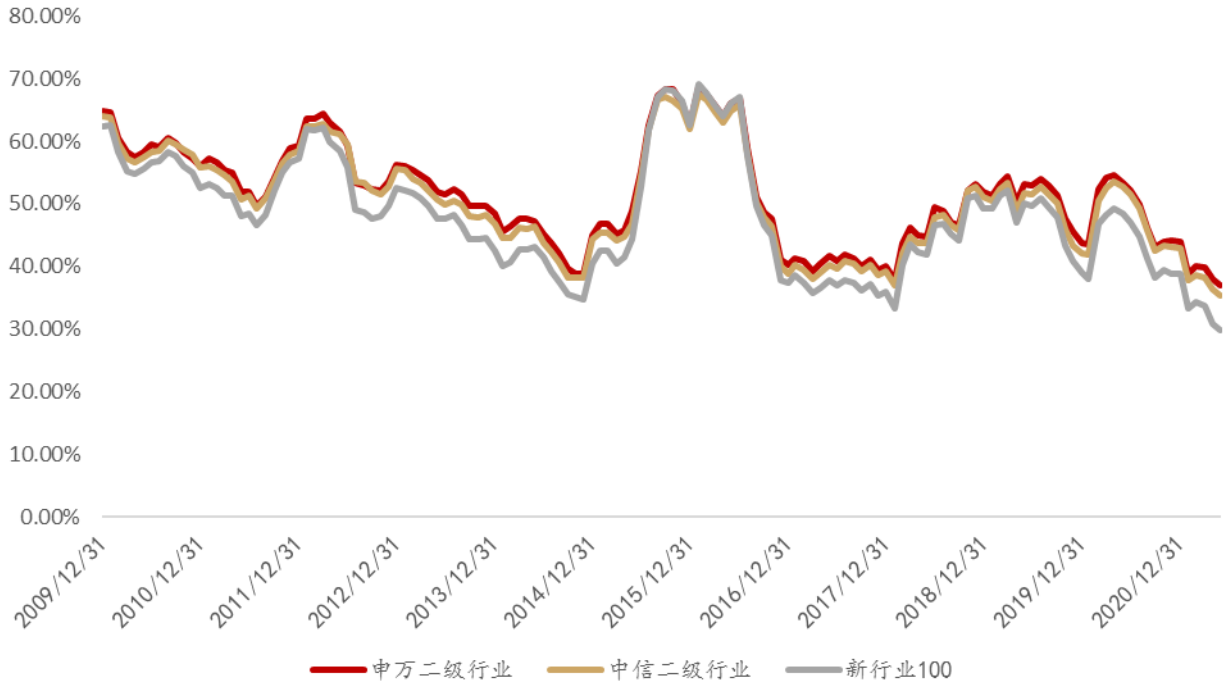
图10：不同行业分类下行业成分公司数量



3.6 基于产品的行业分类一定程度上可以实现最小化行业内差距

- 行业分类的目标之一是最大化行业间差距，最小化行业内差距，我们分别计算申万二级行业、中信二级行业以及通过文本分析方法得到的聚类行业内公司股价相关性，可以看到，基于文本分析的行业分类方法，股价相关性与传统行业分类较为接近。

图11：基于主营业务产品信息聚类的行业分类股价相关性较高

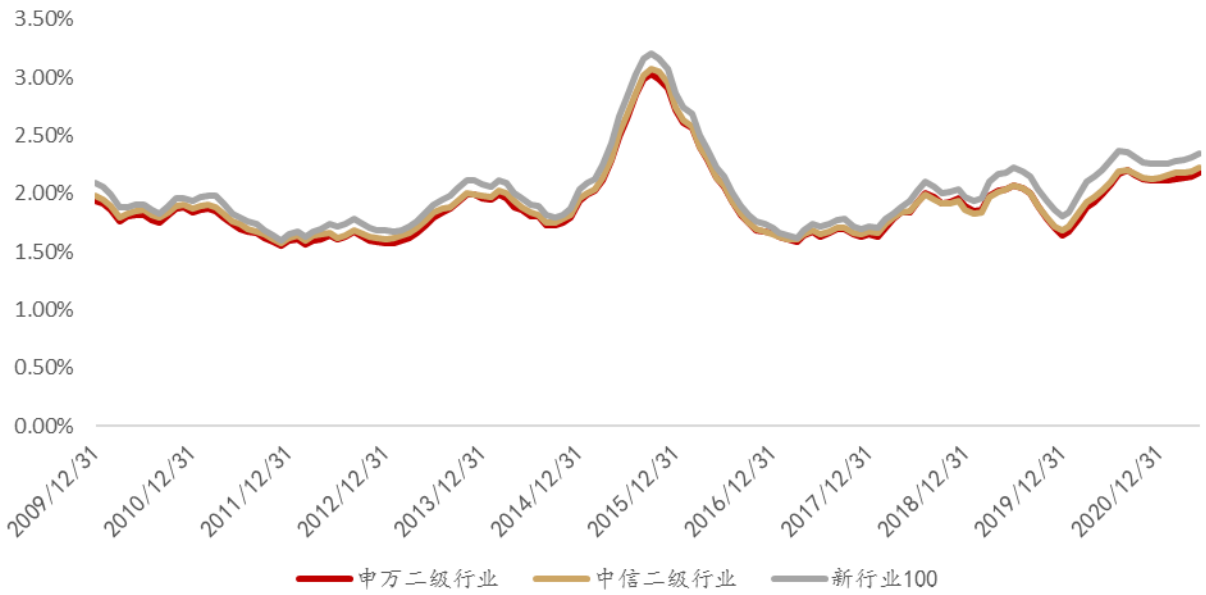


3.7 收益率标准差角度新行业分类结果也与通用行业接近

华西证券 HUAXI SECURITIES

- 此外，行业内股票横截面收益率标准差也可以衡量行业内股票表现的一致程度，我们取最近六个月的历史均值，可以看到新行业分类结果与申万和中信行业分类结果也较为接近。

图12：基于主营业务产品信息聚类的行业分类股价收益率标准差较低



3.8 基于新行业分类的指数增强策略测试

华西证券 HUAXI SECURITIES

- 我们进一步测试将该行业分类应用于中证500指数增强策略上的效果。选股范围为全市场非ST股票，因子组合、风险约束等均相同，唯一区别在于股票行业分类分别采用申万二级行业和文本聚类100行业，市值、行业约束要求严格中性。
- 从新行业分类与申万二级行业对比来看，结果较为接近，新行业分类下年化超额收益有一定提升（从18.67%提升至19.21%），但信息比率有所下降（3.73下降至3.54）。

图13：不同行业分类标准下中证500增强组合超额收益曲线

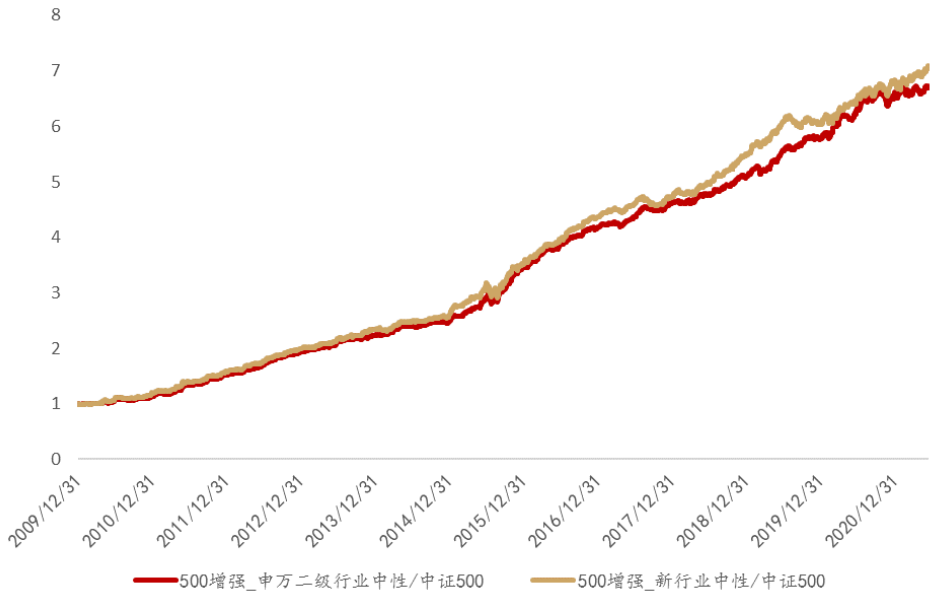


表6：不同行业分类标准下中证500增强组合分年度表现

年份	绝对收益		超额收益	
	500增强_申万二级	500增强_新行业	500增强_申万二级	500增强_新行业
2010年	24.57%	29.88%	14.50%	19.81%
2011年	-10.96%	-11.72%	22.87%	22.10%
2012年	27.31%	26.94%	27.03%	26.66%
2013年	34.72%	37.28%	17.84%	20.39%
2014年	56.10%	55.29%	17.09%	16.28%
2015年	99.25%	93.79%	56.13%	50.68%
2016年	-1.50%	1.48%	16.27%	19.25%
2017年	10.61%	7.73%	10.81%	7.94%
2018年	-26.67%	-22.14%	6.65%	11.18%
2019年	43.91%	39.00%	17.53%	12.61%
2020年	37.69%	35.81%	16.82%	14.94%
2021年	7.55%	9.77%	1.91%	4.14%
年化收益	22.25%	22.79%	18.66%	19.21%
信息比率			3.73	3.54

3.9 小结

- 本报告使用文本分析方法对A股上市公司行业分类进行了重新划分，初步研究结果显示，基于主营产品相似度的行业划分方法具有一定的可行性，不论是行业内股价相关性，还是将其应用于指数增强策略，新行业分类表现均与现有行业分类较为接近。
- 该划分方法基于上市公司披露的主营产品进行相似度聚类，由于主营业务明细在A股上市公司报表中仅在附注中披露，且不同公司披露详细程度不一致，因此存在一定的误归类可能。
- 语料的丰富程度对于模型准确度有较大影响，而大训练样本在模型训练过程中耗时较长，因此合适的语料文件一定程度上可能会影响分类结果。
- 不同词向量模型下的分类结果可能存在一定差异，本文中我们仅测试了较为常用的Word2vec模型的结果，模型存在较大可优化的空间。

目录

风险提示

量化模型基于历史数据计算，未来可能存在失效风险。

分析师承诺

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资评级	说明
以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过15%
	增持	分析师预测在此期间股价相对强于上证指数在5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过15%
行业评级标准		
以报告发布日后的6个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层

网址：http://www.hx168.com.cn/hxzq/hxindex.html

免责声明

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。 。
本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。