

基于层次聚类的云南省生物医药产业 专利发展状况研究

杨 森, 吴建德

(昆明理工大学 知识产权发展研究院, 昆明 650093)

摘要:为了更好的分析云南省生物医药产业专利发展状况,以国家知识产权局专利数据库中收录的 1997—2016 年间的云南省生物医药领域专利信息为数据源,选取专利申请量、授权量、成长率和有效率作为层次聚类变量,对云南省生物医药产业专利近 20 年发展态势、重点技术领域 IPC 分类号和专利高产申请人采取层次聚类,给出聚类结果并分析。聚类结果与云南省生物医药产业目前整体发展状况相对吻合,且未来前景乐观。

关键词:聚类分析;生物医药;专利;发展状况

中图分类号:G306;R1 **文献标志码:**A **文章编号:**1671-1807(2017)12-0035-06

生物医药产业是 21 世纪最为活跃、影响最为深远的新兴产业,也是我国七大战略性新兴产业之一。生物医药是关系着国计民生和促进社会健康发展的特殊产业,因此具有重大的战略意义。云南省在生物医药产业发展中拥有着独特的天然优势,是生物资源、天然药物和民族医药资源最丰富的省份之一。在国家相关政策的鼓励下,云南省委省政府也相继出台了《云南省生物医药和大健康产业发展规划(2016—2020 年)及三年行动计划(2016—2018 年)》,推进生物医药产业创新升级和快速发展。生物医药产业具有研发周期长、成本高、风险大等特点,因此研发成果更迫切需要知识产权保护。

近年来,国内学者基于不同角度和方法对云南省生物医药产业的研究日益增多。本文则针对云南省生物医药产业的专利发展状况进行研究。专利作为技术信息最有效的载体,涵盖了全球 90% 以上最新的技术情报,因此对生物医药产业专利发展状况研究,更能揭示出生物医药产业的发展态势。专利数据分析是指对相关专利进行检索、去噪、统计,使之成为可以分析处理的数据。本文采用层次聚类方法,将生物医药产业专利数据归纳为一系列有意义的簇,对聚类结果进行分析。

本文所采用的层次聚类分析与简单的归类分析

不同之处在于:一般的归类分析往往是以其中一个变量作为指标将对象进行分类,而层次聚类分析是可以同时选择多个变量作为指标将对象划分类群,使研究结果更具整体性。

1 数据检索与研究工具

1.1 数据检索

本文研究对象是云南省生物医药产业专利数据,专利数据来源于国家知识产权局专利数据库。为了更加细致和全面地得到本文研究所需的专利数据,选择 IPC 分类号和主题关键词相结合的检索策略,时间跨度为 1997—2016 年,即一个发明专利权的有效生命周期 20 年。

根据生物医药的解释,可以将生物医药理解为是以基因工程、细胞工程或抗体工程技术生产的,源自生物体内,用于体内诊断、治疗或预防的药物,主要包括基因工程药物、疫苗、多肽和核酸类药物等,广义理解的生物医药也包括化学药、天然药和中药等传统医药领域^[1]。同时结合云南省在生物医药领域的独特资源,确定检索方向,进一步构建检索式,最终制定出本文研究所需的 IPC 分类号和关键词,见表 1。

1.2 研究工具

专利数据、相关数据计算由 Office、Excel 完成,专利数据的聚类分析由 SPSS 20.0 完成。

收稿日期:2017-09-15

作者简介:杨森(1989—),男,河北邯郸人,昆明理工大学知识产权发展研究院,硕士研究生,研究方向:数字信息管理与保护;吴建德(1979—),男,云南保山人,昆明理工大学知识产权发展研究院,教授、博士生导师,博士,研究方向:工业大数据分析、复杂工业过程检测、控制与优化。

表 1 生物医药领域专利检索式构建

检索要素	说明
IPC 分类号	A61、C07H、CO7K、C12M、C12N、C12P、C12Q、C12R、G01N33
主关键词	基因、细胞、酶、蛋白、肽、血清、核酸、氨基酸、核苷酸、维生素、介素、疫苗、菌苗、抗体、抗生素、干扰素、抑制剂、凝血剂、交联剂、诊断试剂、制剂、试剂、克隆、移植、重组、转基因、制备方法、制备办法、构建方法、天然药、中药、民族药、三七、灯盏花、天麻、石斛、滇重楼、云当归、白及、滇龙胆、水蛭、美洲大蠊、云茯苓、云木香、黄草乌、黄精、珠子参、砂仁

2 研究方法

2.1 层次聚类

层次聚类是一种被广泛使用的经典聚类方法，可分为凝聚的层次聚类和分裂的层次聚类。考虑到本文研究数据的特点，选用凝聚的层次聚类方法。凝聚的层次聚类方法如下：

- 1)将每个对象归为一类，计算临近度矩阵；
- 2)合并最接近的两个类；
- 3)更新临近度矩阵，以反映新的类与旧的类之间的临近度；
- 4)重复第 2 步与第 3 步，直到最后合并成为一个类为止。

算法中的临近度计算即为类与类之间的亲疏程度，根据计算出的亲疏程度从而将其进行归类。本文采用欧几里得距离作为临近度的计算方法，将每个聚类成员看作 n 维空间的一个点，在 n 维空间中，点与点的距离越近，其相似程度越高，更有可能聚为一类。下面给出具体欧几里德距离公式：

$$d(x,y)=\sqrt{\sum_{k=1}^n(x_k-y_k)^2}$$

(1)

公式(1)中， n 是维数， x_k 和 y_k 分别是 x 和 y 的第 k 个属性值。

2.2 聚类成员与聚类变量的确定

本文主要从云南省生物医药产业专利数据的年度发展状况、重点技术领域 IPC 分类号的构成、专利高产申请人三方面来分析，这三方面的状况可以作为某个地区专利发展态势的分析要素。因此，聚类成员确定为此 3 项。

本文选用一般专利分析中常用到的分析指标作为聚类变量：专利申请量、专利授权量、专利成长率和专利有效率。下面给出这些指标的含义、计算方法及相关说明。

专利申请量和专利授权量为根据检索式检索出 1997—2016 年间的云南省生物医药产业专利数据统计值。

专利成长率为当年授权专利数量与上一年授权专利数量的百分比，它可显现技术创新随时间的变化增加或减缓。专利均成长率为 20 年的专利成长率平均值。

专利有效率为截止当年最后一日的有效专利数量与截止当年最后一日所有授权专利数量的百分比。发明专利被授予专利权后，在专利权有效期内需缴纳费用，而且费用随时间的推移越来越高，所以长期有效的专利一定存在其技术价值。因此专利有效率这一指标就变得尤为重要。

2.3 数据预处理和产生样本空间

本文所指专利即为发明专利，因为根据检索式检索出的 1997—2016 年专利数量，共有发明专利申请 5 626 件，实用新型专利申请 176 件，实用新型数量过少，且发明专利较实用新型专利具有更高的技术价值，所以这里暂且不把实用新型专利统计在内。

数据预处理是将检索出的专利数据中存在的一些不相关数据删除，起到数据去噪的效果，并将保留下来的数据进行相关指标计算和规范化处理，生成可被 SPSS 20.0 处理的样本空间，见表 2、3、4。

表 2 1997—2016 年生物医药产业专利年度状况

聚类编号	年度	专利申请量/件	专利授权量/件	专利成长率/%	专利有效率/%
1	1997	23	13	41.9	15.3
2	1998	21	13	100.0	30.6
3	1999	47	29	223.0	24.1
4	2000	111	53	182.7	13.2
5	2001	90	53	100.0	26.4
6	2002	131	71	133.9	30.9
7	2003	190	110	154.9	36.3
8	2004	203	126	114.5	42.8
9	2005	225	128	101.5	40.6
10	2006	246	146	114.0	44.5
11	2007	181	98	67.1	52.0
12	2008	260	130	132.6	61.5
13	2009	263	141	108.4	55.3
14	2010	422	248	175.8	64.9
15	2011	332	199	80.2	73.3
16	2012	410	225	113.0	69.7
17	2013	467	260	115.5	86.9
18	2014	551	204	78.4	89.7
19	2015	766	55	26.9	89.0
20	2016	866	2	3.6	100.0

表 3 生物医药产业专利技术领域 IPC 状况

聚类编号	IPC 分类号	专利申请量/件	专利授权量/件	专利均成长率/%	专利有效率/%
1	A61	3 984	1 684	100.7	68.1
2	C07H	156	86	138.9	69.7
3	C07K	256	95	110.7	65.2
4	C12M	37	21	46.3	53.4
5	C12N	1 165	444	107.8	58.5
6	C12P	193	76	102.0	52.6
7	C12Q	466	147	195.0	59.2
8	C12R	531	212	115.6	64.1
9	G01N33	204	75	161.0	69.0

表 4 主要生物医药产业专利高产申请人状况

聚类编号	申请人	专利申请量/件	专利授权量/件	专利均成长率/%	专利有效率/%
1	昆明理工大学	516	167	112.8	48.0
2	中国科学院昆明植物研究所	313	193	125.0	80.3
3	云南大学	228	100	102.0	28.5
4	中国科学院昆明动物研究所	184	86	107.9	54.7
5	云南农业大学	167	53	63.8	40.3
6	中国医学科学院医学生物学研究所	129	60	111.9	100.0
7	昆明制药集团股份有限公司	119	81	67.8	100.0
8	云南师范大学	89	42	84.7	50.0
9	云南中烟工业有限责任公司	86	15	10.0	93.3
10	云南民族大学	73	37	97.1	45.9
11	云南省烟草农业科学研究院	62	26	40.9	96.1
12	云南中医学院	51	10	22.5	100.0
13	云南省药物研究所	55	24	67.1	95.8
14	云南白药集团股份有限公司	55	33	63.4	93.9
15	昆明医科大学	51	14	24.6	92.8
16	昆药集团股份有限公司	37	1	5.0	100.0
17	云南烟草科学研究院	33	20	46.6	60.0
18	昆明振华制药厂有限公司	35	22	46.3	100.0

由于各变量存在不同量纲、数据存在不同数量级,所以为消除变量间的权重差异,使各数据之间具有可比较性,在聚类前还需要对数据进行标准化变换。这里采取标准正态变换方法,把原始数据转换为标准 Z 分数(Z scores),其计算公式为:

$$x_{ij} = \frac{x_{ij} - \overline{x_j}}{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \overline{x_j})^2}$$

(2)

根据公式(2),先将需标准化处理的变量的每个变量值 x_{ij} 减去该变量的平均值 $\overline{x_j}$,继而再除以该变量的标准差。标准化后的变量值平均值为 0,方差为 1。

3 聚类结果分析

3.1 年度发展状况结果分析

将表 1 的样本空间进行层次聚类,聚类结果采用

树状谱系图展示,如图 1。

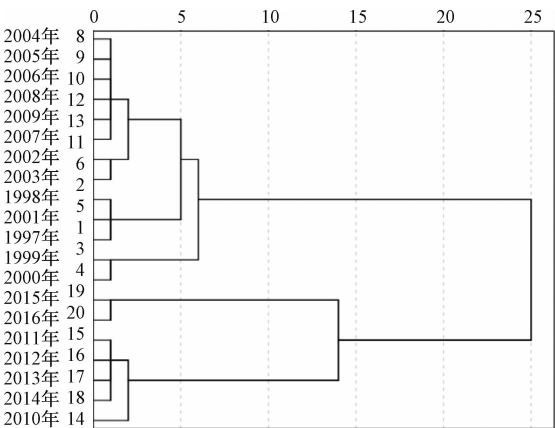


图 1 年度发展状况聚类谱系图

由图 1、表 1 的结果可见,根据每一年专利申请

量、专利授权量、专利成长率及专利有效率四个变量的不同,将 20 年的发展状况聚为 4 类。

2015 年、2016 年聚为第 1 类。这两年生物医药产业专利申请量较之其他年份有着巨大的增幅,分别为 766 件和 866 件。专利申请量的大幅增加意味着生物医药企业、高校和科研院所所在生物医药领域研发成果不断,且越来越重视申请专利来保护技术成果,提高自身竞争力。同时,云南省委省政府出台的相关鼓励和引导政策效果也较为明显。受发明专利授权周期长等因素影响,专利授权量、专利成长率较之其他年份有着不小的减幅。从专利授权量和专利成长率的减少来看,目前云南省生物医药产业在创新能力、核心技术等方面仍存在不足,未能将产业基础和资源优势完全加以利用。生物医药企业、高校和科研院所应利用云南省在生物医药领域特有的优势,加大研发力度,进一步提高技术创新能力,增多增强自己的核心技术。

2010—2014 年这连续的 5 年聚为第 2 类。可以看出,这 5 年中专利申请量、专利授权量、专利成长率、专利有效率虽有所波动,但整体保持增长态势。云南省在“十二五”期间,将生物医药产业作为可持续发展的战略产业而重点加以培养,以实现生物医药产业跨越式发展,且专利数量已成为衡量研发成果的一个重要标准。据云南省官方数据显示,2010 年—2014 年云南省生物医药产业经济总量保持了年均 25% 以上增长,这一经济增长速率与专利数据的各项指标均增长率保持一致。可以看出,专利数量和质量的稳步提升,意味着云南省自“十二五”以来,生物医药产业发展态势良好,生物医药产业已俨然成为云南省经济转型发展的战略性新兴产业之一。

2003—2009 年这连续的 7 年聚为第 3 类。第 3 类与第 2 类的数据指标比较而言,各数据指标略低,但仍以较好的态势逐步增长。事实上,从 2003 年起,云南省经过经济产业结构调整,开始提出将生物医药产业逐步实现跨越式发展。这 7 年的发展,为生物医药产业后继进一步跨越式发展奠定了产业基础。生物医药企业、高校和科研院所对知识产权的保护和利用逐步有了认识,以专利技术作为其核心竞争力已成为趋势。

1997—2002 年这连续的 6 年聚为第 4 类,在这早期的 6 年间,各数据指标均比较低,但是每一年也有小幅增长的趋势,专利作为核心竞争力已经开始步入生物医药产业。

3.2 技术领域 IPC 分类号状况结果分析

通过分析云南省生物医药产业专利主要 IPC 分类号,可以找出关键技术领域相对集中的地带,从而了解云南省在生物医药产业研发的重点方向。将表 2 的样本空间进行层次聚类,聚类结果如图 2。

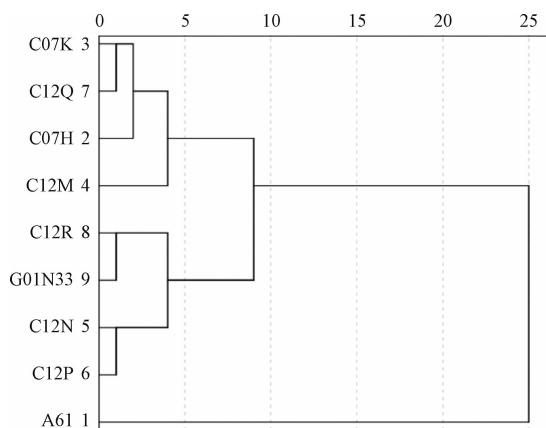


图 2 IPC 分类号状况聚类谱系图

由图 2、表 2 的结果可见,系统将技术领域 IPC 分类号聚为 3 类。

第 1 类 IPC 分类号为 A61,在《国际专利分类表》中,A61 代表医学或兽医学、卫生学。从其高达 3 984 件专利申请量、1 984 件专利授权量表明 A61 这一技术领域在云南省生物医药产业中占据着重要位置。医学或兽医学、卫生学是生物医药领域重要的组成部分,也是与我们日常生活最贴近的领域。因此,生物医药企业、高校和科研院所在这一个基础广泛的技术领域范围内技术成果丰硕且产业基础踏实。

第 2 类 IPC 分类号为 C12P、C12N、G01N33、C12R,在《国际专利分类表》中,这 4 类可以概括为微生物、酶、微生物和酶组合物、特殊方法研究或分析材料、发酵或使用酶的方法获得的组合物。这一类技术领域相关专利的专利成长率均在 100% 以上,表现出较高的成长率,显现出这一技术领域目前发展比较迅速,技术创新也集中在这一类技术领域。聚类结果也从另一方面体现出云南已经开始利用其独特优势做大做强天然药、中药(民族药),这一类技术领域与其紧密相关。

第 3 类 IPC 分类号为 C07H、C07K、C12M、C12Q,在《国际专利分类表》中,这 4 类可以概括为核苷、核酸、核苷酸、肽、微生物装置、酶学、包含酶或微生物的检测方法等领域。这一类技术领域各数据指标表现出平稳增长的趋势,同时也显现出云南省已经

着力发展现代生物制药技术。

3.3 专利高产申请人状况结果分析

专利高产申请人代表着生物医药领域技术研发的中坚力量,其技术研发能力在很大程度上将推进云南省生物医药产业的发展。将表 3 的样本空间用进行层次聚类,聚类结果如图 3。

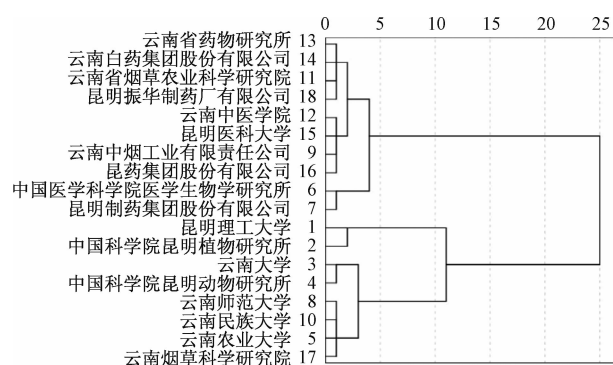


图 3 专利高产申请人状况聚类谱系图

由图 3、表 3 的结果可见,聚类结果将云南省生物医药领域专利申请量前 18 位的申请人进行了层次聚类,聚类结果简单划分的话可以看作 3 类。

昆明理工大学、中科院昆明植物研究所聚为第 1 类,这两位申请人分别为高校和科研院所。从表 3 也不难看出,这两申请人在专利申请量、专利授权量上均大幅度领先其他申请人。其专利均成长率、和专利有效率也趋于较高水平。所以无论从专利的数量或者质量上来看,这两者在云南省生物医药产业技术研发方面走在前列,其知识产权保护意识也较强。

云南大学、中科院昆明动物研究所、云南农业大学、云南民族大学、云南师范大学、云南烟草科学研究院这 7 位申请人聚为第 2 类。从聚类结果可以看出,第 2 类的 7 位申请人中,高校数量比重大,各项数据指标呈现平稳增长趋势,同时也体现出生物医药领域相关技术多集中在高校和科研院所,尚未形成以企业为主体的技术创新体系。

剩下的申请人聚为第 3 类。第 3 类的申请人中生物医药企业数量居多,如在全国生物医药企业都有着很高声誉的云南白药集团股份有限公司、昆药集团股份有限公司。这一类申请人虽然专利数量不多,但

是其专利有效率却要比第 1、2 类高出很多百分点,专利有效率均高达 90% 以上,所以这一类申请人的专利技术价值和商业价值较高。所以在以生物医药企业为主体的第 3 类中,应进一步增加企业核心专利技术,将专利技术与产品开发、生产相结合,打造企业优势产品和名牌产品。

4 结语

本文采用层次聚类对云南省生物医药产业专利发展状况进行研究,聚类结果与当前云南省生物医药产业整体发展状况相对吻合。当前,云南省生物医药产业专利发展状况良好,创新能力不断增强,重点领域创新成果不断涌现。随着国家实施“一带一路”战略,为云南省生物医药产业发展带来了难得的机遇。云南省委省政府也在不断完善政策体系,加大政策支持力度,推进优势、重点生物医药领域技术创新,把生物医药产业打造成为引领云南经济发展的支柱产业。云南在生物医药领域的优势确实非常明显,有资源就会有无限的可能,只要拥有足够先进的科学技术,便可以从这些宝库中研发出更多专利成果,在生物医药领域云南可以说是大有可为。

参考文献

- [1] 姜莉莉. 基于专利数据分析的云南生物医药企业专利战略研究[D]. 昆明: 云南大学, 2013.
- [2] 王天哥, 王金苗, 袁红梅. 基于专利维度的我国生物医药核心技术的识别与分析[J]. 情报杂志, 2016, 35(4): 112—117.
- [3] 郭良玉, 王旭杰. 基于聚类分析的中药复方专利年度发展状况研究[J]. 中草药, 2012, 43(10): 2083—2088.
- [4] 刘洋, 瞿卫军, 黄庆, 肖云鹏, 石昱, 曹津燕. 专利评价指标体系(二)——运用专利评价指标体系中的指标进行数据分析[J]. 知识产权, 2004, 14(83): 29—34.
- [5] 夏太寿, 王园磊, 田丽丽. 基于专利分析的江苏省生物医药发展现状与对策分析[J]. 中国生物工程杂志, 2016, 36(8): 123—130.
- [6] 双凡. 云南生物产业专利分析与预警[M]. 哈尔滨: 哈尔滨工程大学出版社, 2016.
- [7] 甘邵宁, 曾志华. 专利信息分析管理与应用[M]. 北京: 知识产权出版社, 2015.
- [8] 缪元武. 基于层次聚类的数据分析[D]. 合肥: 安徽大学, 2013.

(下转第 166 页)

发掘,具有一定的研究意义和现实价值。

参考文献

- [1] 季辉. 中小企业融资难的原因及对策[J]. 科技和产业, 2005, 5(5): 93—95.
- [2] 刘澄, 武鹏, 郝丹杰. 基于模糊集理论的科技型小微企业信用等级评估研究[J]. 科技进步与对策, 2013, 30(18): 170—176.
- [3] 王帅, 杨培涛, 黄庆雯. 基于多层次模糊综合评价的中小企业信用风险评估[J]. 财经理论与实践, 2014(5): 13—17.
- [4] CANALES R, NANDA R. A darker side to decentralized banks: market power and credit rationing in SME lending[J]. Journal of Financial Economics, 2012, 105(2): 353—366.
- [5] 李道波. 科技型企业财务预警实证研究[J]. 绵阳师范学校学报, 2005, 24(6): 16—21.
- [6] 杨楠. 成长型科技创业企业的信用评价研究[J]. 重庆大学学报: 社会科学版, 2015, 21(2): 69—74.
- [7] 楼裕胜. 基于模糊神经网络的企业信用风险评估模型研究[J]. 中南大学学报: 社会科学版, 2013(5): 21—25.
- [8] 李秉祥. 基于模糊神经网络的企业财务危机非线性组合预测方法研究[J]. 管理工程学报, 2005, 19(1): 19—23.
- [9] 庞素琳. 概率神经网络信用评价模型及预警研究[J]. 系统工程理论与实践, 2005, 25(5): 43—48.
- [10] 张洪潮, 雒国斌. 科技型小微企业集聚式发展研究[J]. 企业经济, 2014(6): 86—90.
- [11] 申华. 基于数据挖掘的个人信用评分模型开发[D]. 厦门: 厦门大学, 2009.
- [12] 初俊博. 改进自适应神经网络控制策略及仿真研究[D]. 大连: 大连海事大学, 2008.
- [13] 孙英人. 企业科技信用评价研究[D]. 大连: 大连理工大学, 2008.
- [14] 陈森林. 前向神经网络的分类能力与训练算法的研究[D]. 西安: 西安电子科技大学, 2003.
- [15] 贺颖, 张目, 甘罗佳. 基于熵权法的科技型中小企业信用评价研究[J]. 科技创业月刊, 2013, 26(7): 64—66.

Chinese Science and Technology SME Credit Risk Evaluation Model

——The empirical analysis based on MLP neural network

WEN Xiao-ni, HAN Xin-rui

(School of Economics & Management, Xidian University, Xi'an 710126, China)

Abstract: This thesis aimed at the science and technology SME, firstly established a multi-level evaluation index system of credit risk, and use the t test and factor analysis to simplify the index system, then on the basis of the index system establishes a credit risk evaluation model of the science and technology SME based on the MLP neural network. Finally, use 66 high-tech listed companies as the research sample for empirical analysis, the test results show that the model established in this paper for small and mid-sized enterprise credit risk evaluation model has a high accuracy.

Key words: science and technology finance; neural networks; credit risk; evaluation model

(上接第 39 页)

Research on Patent Development of Biomedical Industry in Yunnan

Province Based on Hierarchical Clustering

YANG Sen, WU Jian-de

(Yunnan Research Institute of Intellectual Property Rights Development, Kunming University of Science and Technology, Kunming 650093, China)

Abstract: In order to analyze the patent development of bio pharmaceutical industry in Yunnan Province, hierarchical clustering analysis of patent data of biological medicine in Yunnan province from 1997 to 2016 was carried out in the patent database of the State Intellectual Property Office, taking the patent application quantity, authorization quantity, growth rate and efficiency are the hierarchical cluster variables, give the results of cluster analysis, development trend of biotechnology industry patents in Yunnan Province in recent 20 years, key technical areas and high yield applicants. The clustering results are in good agreement with the overall development of the biomedical industry in Yunnan Province, and the future prospect is optimistic.

Key words: cluster analysis; biomedicine; patent; development status