

NSFC 计算机图像与视频处理领域项目关键词分析*

侯素娟¹, 赵瑞珍², 宋 苏²

¹(山东师范大学 信息科学与工程学院, 山东 济南 250014)

²(国家自然科学基金委员会 信息科学部, 北京 100085)

通讯作者: 赵瑞珍, E-mail: zhaorz@nsfc.gov.cn



摘要: 关键词能够反映出一份项目申请书的主要研究内容. 统计了国家自然科学基金计算机图像与视频处理领域 2014 年~2018 年申请与资助项目关键词, 并分别从关键词标引量、关键词词频等方面进行分析, 探讨其与资助率的关系. 最后, 运用定量的方法, 透过高频关键词的内容变化, 分析近 5 年来的计算机图像与视频处理领域的研究热点变化.

关键词: 国家自然科学基金; 计算机图像与视频处理; F020502; 关键词; 资助率
中图法分类号: TP399

中文引用格式: 侯素娟, 赵瑞珍, 宋苏. NSFC 计算机图像与视频处理领域项目关键词分析. 软件学报, 2019, 30(9): 2918–2924. <http://www.jos.org.cn/1000-9825/5849.htm>

英文引用格式: Hou SJ, Zhao RZ, Song S. Keywords Statistics and analysis of applications and grants in the field of computer image and video processing of the National Natural Science Foundation of China. Ruan Jian Xue Bao/Journal of Software, 2019, 30(9): 2918–2924 (in Chinese). <http://www.jos.org.cn/1000-9825/5849.htm>

Keywords Statistics and Analysis of Applications and Grants in the Field of Computer Image and Video Processing of the National Natural Science Foundation of China

HOU Su-Juan¹, ZHAO Rui-Zhen², SONG Su²

¹(School of Information Science and Engineering, Shandong Normal University, Ji'nan 250014, China)

²(Department of Information Sciences, National Natural Science Foundation of China, Beijing 100085, China)

Abstract: Keywords can reflect the main research content of a project application. In this study, the keywords of applications and grants of the National Natural Science Foundation of China in the field of computer image and video processing during 2014–2018 are firstly counted, and then they are analyzed from various perspectives, such as keyword quantity, keyword frequency, and the relationship between them and subsidizing rate. Lastly, the change of research hotspots in the field of computer image and video processing is also discussed based on the frequency of the keywords by using quantitative method.

Key words: National Natural Science Foundation of China; computer image and video processing; F020502; keyword; subsidizing rate

国家自然科学基金申请书通常反映相关领域目前最新进展和成果. 申请书一般都附有关键词, 这些关键词能够迅速、准确地反映出一份申请书的主题内容和重点, 也是函评阶段计算机辅助指派系统进行专家匹配的重要因素之一. 对某个研究领域较长时间段的大量基金申请书进行关键词的词频和趋势规律分析, 有助于揭示本领域研究的热点、总体内容特征、内容之间的内在联系、学术研究的发展脉络与发展方向等^[1]. 文献[2]对国家自然科学基金自动化领域 1986 年~2017 年项目申请和资助数据进行了大量分析, 发现自动化领域基金资助的相关领域反映了国内外研究前沿; 同时发现, 热点领域理论研究的比重大于应用基础研究.

近些年, 随着网络和存储技术的快速发展, 多媒体数据量呈爆炸式增长, 图像和视频数据的研究已成为计算

* 收稿时间: 2019-02-22; 修改时间: 2019-03-25; 采用时间: 2019-04-04; jos 在线出版时间: 2019-05-22

CNKI 网络优先出版: 2019-05-22 15:26:15, <http://kns.cnki.net/kcms/detail/11.2560.TP.20190522.1525.008.html>

机应用领域研究的热点.2019 年,国家自然科学基金委员会信息科学部明确将“社交媒体大数据分析处理”作为优先支持的研究领域之一^[3].目前,计算机学科(对应一级代码 F02)共设有 65 个三级代码.近 5 年,计算机图像与视频处理领域(对应三级代码 F020502)的项目申请量和项目资助量在 65 个三级代码中均排第一位,具体地,在申请方面,2014 年~2018 年度各类项目总量共计 2 445 项,占计算机应用(F0205)项目总量的 27.4%,占整个计算机学科项目总量的 10.4%;在资助方面,2014 年~2018 年度各类项目总量共计 646 项,占计算机应用项目总量的 30.09%,占整个计算机学科项目总量的 11.5%.因此,本文选取 2014 年~2018 年计算机图像与视频处理领域的项目关键词进行统计和分析.

1 关键词标引量与资助率

在进行项目申报时,国家自然科学基金委员会要求每个项目标引 5 个以内(含 5 个)的关键词.我们统计了 2014 年~2018 年计算机图像与视频处理领域的项目申请量及其标引的关键词数量(本文统计的关键词数量包含了重复出现的关键词)、项目资助量及其标引的关键词数量,见表 1.

Table 1 Number of projects and keywords in the field of computer image and video processing from 2014 to 2018

表 1 2014 年~2018 年计算机图像与视频处理领域的项目数量和关键词数量

年份	申请项目		资助项目	
	项目数量	关键词数量	项目数量	关键词数量
2014	452	1 958	137	596
2015	503	2 174	122	529
2016	496	2 175	131	575
2017	533	2 354	145	641
2018	461	2 025	111	503
共计	2 445	10 686	646	2 844

根据表 1 中的结果统计,2014 年~2018 年申请项目 2 445 项,标引关键词 10 686 个,篇均关键词 4.37 个.资助项目 646 项,标引关键词 2 844 个,篇均关键词 4.40 个.

同时,我们还统计了计算机图像与视频处理领域关键词标引量(关键词标引量:申请书中所标注的关键词的个数)及其对应的项目申请量,如图 1 所示.

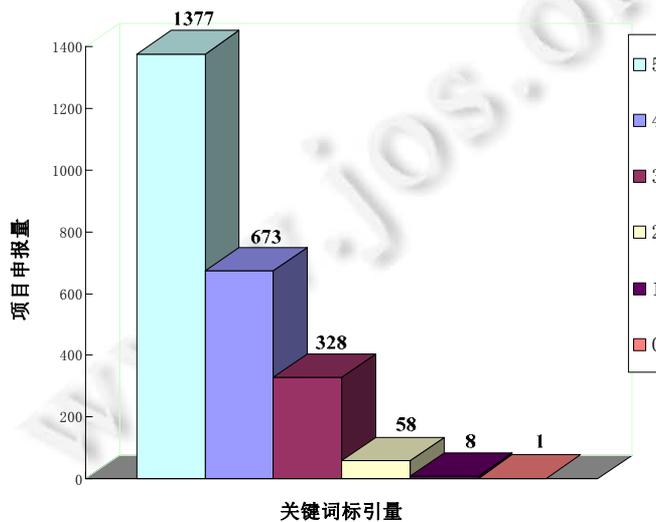


Fig.1 Number of keywords and corresponding project application

图 1 关键词标引量及其项目申请量

由图 1 可以看出:大多数申请项目的关键词标引量集中在 4~5 个,其中,标注了 4 个和 5 个关键词的项目数

量分别有 673 项和 1 377 项,分别占总项目的 27.53%和 56.32%,二者之和所占比例达 83.85%.

此外,笔者还统计了计算机图像与视频处理领域关键词标引量及其对应的项目资助占比,如图 2 所示,其中,括号内的数字为关键词标引量对应项目数量的占比.

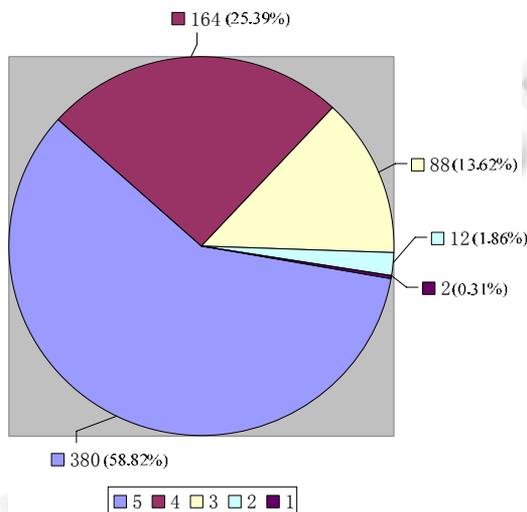


Fig.2 Number of keywords and corresponding project funding rate

图 2 关键词标引量及其项目资助占比

由图 2 可以看出,近 5 年资助项目的关键词标引量具有以下特点.

- (1) 大多数获资助项目的关键词标引量集中在 4~5 个,其中,标注 4 个和 5 个关键词的项目共计 544 项,占项目总数的 84.21%;标引 5 个关键词的项目资助占比最高,相对于同为 5 个关键词的项目申请占比 (56.32%)高 2.5%;
- (2) 结合图 1 数据,可以计算出标引 1~5 个关键词的项目资助率分别为 25%,20.69%,26.83%,24.37%,27.6%,其中,标引 5 个关键词的项目获得资助的比例相对较高.

以上分析表明,获资助申请书其关键词标引量绝大多数在 4 个或以上,当申请书的关键词标引量为 5 时,获资助比例较高.

2 关键词词频与资助率

2.1 关键词词频统计

词频用于定义关键词在某一研究领域中出现的频次高低^[4].我们对 2014 年~2018 年计算机图像与视频处理领域申请及获资助项目的关键词词频进行统计,表 2 中列出排序在前 10 的关键词及其词频(注:申请及获资助项目的关键词总量分别为 10 686 和 2 844).

由表 2 的统计结果可以看出:

- (1) 从关键词反映出的申请项目的研究热点与专家的认可度基本一致,例如:申请项目和资助项目词频排序在前 2 位的均依次为“深度学习”和“特征提取”.然而,两者其他关键词的研究热度和专家认可度并不一致,如“图像理解”在资助项目的关键词词频排序中列第 3 位,而在申请项目的关键词词频排序中列第 7 位;
- (2) 高频词确实反映了计算机图像与视频处理领域的热点:a) 随着 AlphaGo 战胜人类围棋世界冠军,人工智能迎来第三次热潮,深度学习成为本领域备受关注的研究内容,这在表 2 的词频统计结果中得到了充分体现;b) 图像特征提取是图像和视频研究的关键,特征提取的效果直接决定着图像和视频应用

的性能.如何从原始图像中提取具有较强表示能力的特征,是计算机图像与视频处理的一个研究热点.这从表 2 中也得到体现.

Table 2 Top 10 keyword frequency in the field of computer image and video processing from 2014 to 2018

表 2 2014 年~2018 年计算机图像与视频处理领域项目关键词词频 TOP10

排序	资助项目		申请项目	
	关键词	词频	关键词	词频
1	深度学习	92	深度学习	253
2	特征提取	47	特征提取	188
3	图像理解	38	目标跟踪	138
4	目标跟踪	36	目标检测	122
5	稀疏表示	30	图像分割	112
6	图像识别	30	图像识别	108
7	图像分割	26	图像理解	108
8	目标检测	24	稀疏表示	103
9	计算机视觉	23	计算机视觉	81
10	压缩感知	17	图像检索	65

2.2 关键词词频与资助率

我们对计算机图像与视频处理领域的项目出现频次较高的关键词的资助率进行统计,并将统计结果按照资助项目的关键词词频高低进行排序,表 3 给出了资助项目词频排序在前 10 的关键词的资助率统计结果.

Table 3 Keyword frequency and project funding rate in the field of computer image and video processing from 2014 to 2018

表 3 2014 年~2018 年计算机图像与视频处理领域的项目关键词词频与关键词的资助率

关键词	资助项目		申请项目		关键词的资助率 (%)
	词频	排序	词频	排序	
深度学习	92	1	253	1	36.36
特征提取	47	2	188	2	25
图像理解	38	3	108	7	35.19
目标跟踪	36	4	138	3	26.09
稀疏表示	30	5	103	8	29.13
图像识别	30	6	108	6	27.78
图像分割	26	7	112	5	23.21
目标检测	24	8	122	4	19.67
计算机视觉	23	9	81	9	28.4
压缩感知	17	10	55	13	30.91

注:关键词的资助率=(关键词在资助项目中出现的频次/关键词在申请项目中出现的频次)×100

从表 3 可以看出:

- (1) 词频高的项目,其资助率不一定高.例如:“图像理解”在资助项目的关键词词频排序中列第 3 位,在申请项目的词频排序中列第 7 位,其资助率较高;“目标检测”在资助项目的关键词词频排序中列第 8 位,在申请项目的关键词词频排序中列第 4 位,其资助率较低;
- (2) 涉及“深度学习”和“图像理解”的项目相对容易得到同行专家的认可.

3 研究热点变化

词频分析法是利用词频来确定该领域研究热点和发展动向的计量学方法^[4].申请书中关键词词频的高低,可以反映本领域的研究热点和主要研究方向,由关键词词频分布特征可以分析出本领域研究的集中程度.为了分析近 5 年计算机图像与视频处理领域研究热点的变化,笔者运用词频分析法从词频方面分别对申请项目和资助项目的关键词词频进行逐年统计,并列出了词频排序在前 10 的关键词(见表 4).由表 4 的统计结果可以看出:

- (1) 对于词频最高的关键词,申请项目和资助项目两者完全匹配.例如:2014 年,申请项目和资助项目词频

最高的关键词均为“计算机视觉”;2015年~2018年,关键词“深度学习”的词频排序持续保持第一.目前,深度学习发展迅速,其研究价值和潜力正不断地被挖掘,在图像检索、语音识别、人脸识别、机器翻译等领域均取得了突出的应用效果;

- (2) 对于其他关键词,申请项目和资助项目两者的词频排序基本不匹配,具体表现在同一关键词不仅在同一年度的申请和资助的词频排序各不相同,而且不同年份的词频排序变化波动也较为明显.以“图像检索”关键词为例,图3给出其近5年在申请和资助两方面词频排序的变化.

Table 4 Top 10 annual keyword frequency of the field of computer image and video processing from 2014 to 2018

表 4 2014年~2018年计算机图像与视频处理领域项目关键词 TOP10 逐年统计

	排序	资助	申请
2014年	1	计算机视觉	计算机视觉
	2	稀疏表示	特征提取
	3	特征提取	图像分割
	4	视频分析	稀疏表示
	5	图像分析	数字图像处理
	6	图像分割	图像分析
	7	视频编码	目标跟踪
	8	深度学习	视频分析
	9	三维重建	视频编码
	10	目标跟踪	图像识别
2015年	排序	资助	申请
	1	深度学习	深度学习
	2	图像理解	目标跟踪
	3	特征提取	特征提取
	4	目标跟踪	图像分割
	5	图像分割	目标检测
	6	稀疏表示	稀疏表示
	7	视频压缩	行为识别
	8	图像处理	图像检索
	9	图像融合	图像理解
10	图像识别	图像识别	
2016年	排序	资助	申请
	1	深度学习	深度学习
	2	图像理解	特征提取
	3	特征提取	目标跟踪
	4	目标跟踪	图像理解
	5	图像重构	目标检测
	6	压缩感知	图像分割
	7	图像分割	图像识别
	8	特征描述	图像去噪
	9	视频压缩	特征描述
10	图像去噪	压缩感知	
2017年	排序	资助	申请
	1	深度学习	深度学习
	2	图像识别	图像理解
	3	图像理解	目标检测
	4	目标跟踪	图像识别
	5	目标检测	特征提取
	6	特征提取	目标跟踪
	7	稀疏表示	卷积神经网络
	8	迁移学习	图像分割
	9	行为识别	稀疏表示
10	图像检索	图像融合	

Table 4 Top 10 annual keyword frequency of the field of computer image and video processing from 2014 to 2018

表 4 2014 年~2018 年计算机图像与视频处理领域项目关键词 TOP10 逐年统计

	排序	资助	申请
	2018 年	1	深度学习
	2	特征提取	特征提取
	3	卷积神经网络	图像理解
	4	目标跟踪	目标跟踪
	5	图像理解	图像识别
	6	图像特征	目标检测
	7	生成对抗网络	卷积神经网络
	8	图像分割	图像特征
	9	视频压缩	稀疏表示
	10	图像分辨率增强	特征描述

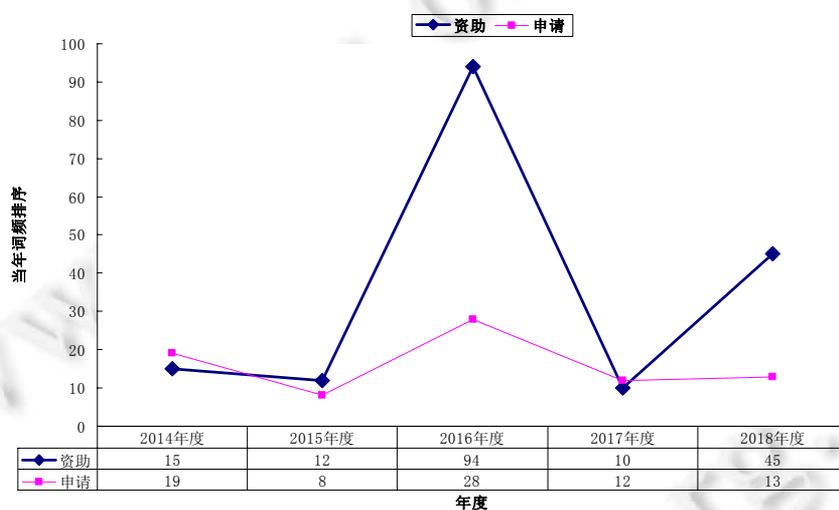


Fig.3 Change of Word Frequency Ranking towards Image Retrieval from 2014 to 2018

图 3 关键词图像检索在 2014 年~2018 各年度的词频排序变化

4 总 结

本文从申请项目和资助项目两个方面,对 2014 年~2018 年计算机图像与视频处理领域的关键词进行统计,分别对关键词的标引量和词频进行分析,并探讨了它们与资助率的关系,最后,透过热频关键词的变化分析了计算机图像与视频处理领域研究热点的变化.可以看出,“深度学习”在计算机图像与视频处理领域已持续成为大家关注和感兴趣的研究课题.

致谢 感谢国家自然科学基金委科学传播中心徐卫国老师在论文撰写和修改过程中给予的悉心指导和帮助.

References:

- [1] Li WL, Yang ZG. Analysis on frequencies of keywords in Chinese information science journals papers. Information Science, 2005,23(1):68-70,143 (in Chinese with English abstract).
- [2] Deng F, Song S, Liu K, Wu GZ, Fu J. Data and research hotspot analyses of national natural science foundation of china in automation field. Acta Automatica Sinica, 2018,44(2):377-384 (in Chinese with English abstract).

- [3] National Natural Science Foundation of China. National Natural Science Fund Guide to Programs 2019. Beijing: Science Press, 2019 (in Chinese).
- [4] Ma FC, Zhang Q. Comparative analysis of knowledge management literature between China and overseas: A bibliometric analysis. Journal of the CHINA Society for Scientific and Technical Information, 2006,25(2):163-171 (in Chinese with English abstract).

附中文参考文献:

- [1] 李文兰,杨祖国.中国情报学期刊论文关键词词频分析.情报科学,2005,23(1):68-70,143.
- [2] 邓方,宋苏,刘克,吴国政,付俊.国家自然科学基金自动化领域数据分析与研究热点变化.自动化学报,2018,44(2):377-384.
- [3] 国家自然科学基金委员会.2019年度国家自然科学基金项目指南.北京:科学出版社,2019.
- [4] 马费成,张勤.国内外知识管理研究热点——基于词频的统计分析.情报学报,2006,25(2):163-171.



侯素娟(1984—),女,山东成武人,博士,副教授,CCF 专业会员,主要研究领域为计算机图像,视频处理.



宋苏(1963—),男,博士,教授,主要研究领域为自适应控制,人工智能和复杂系统理论,计算机理论与应用.



赵瑞珍(1975—),男,博士,教授,CCF 专业会员,主要研究领域为计算机图像,视频处理.