

DOI:10.3969/j.issn.1000-9760.2019.06.016

基于甲状腺肿瘤研究热点的双聚类分析*

张庆 苏丽 虞浩然

(济宁医学院医学信息工程学院,日照 276826)

摘要 **目的** 探讨双向聚类分析方法在文本主题识别中的可行性,揭示甲状腺肿瘤领域临床研究和诊疗策略等相关研究热点,为临床相关人员提供参考,以便提高我国甲状腺肿瘤整体的诊治水平。**方法** 文章以 PubMed 收录的甲状腺肿瘤相关文献作为数据来源,提取文献中的主要主题词进行频次统计,截取高频主题词,生成高频主题词-论文矩阵,根据高频主题词的共现情况进行聚类分析,根据分析得到的聚类树图和对应的类标签文献,分析当前该领域的研究现状。**结果** 当前甲状腺肿瘤领域的主要研究热点为甲状腺肿瘤的影像学诊断、遗传学、切除术等 5 个主题方向。**结论** 通过双向聚类分析并对文献进行梳理,对认识甲状腺肿瘤的研究现状,了解其最新的诊疗方法是必要且现实可行的。

关键词 甲状腺肿瘤;双聚类;可视化;聚类分析;研究热点

中图分类号:G250.252 R736.1 文献标识码:A 文章编号:1000-9760(2019)12-445-05

Analysis of hotspots in the field of thyroid tumor research based on the biclustering method

ZHANG Qing, SU Li, YU Haoran

(School of Medical Information Engineering, Jining Medical University, Rizhao 276826, China)

Abstract: Objective The paper mainly explored the feasibility of biclustering method analysis in text topic recognition in order to reveal the clinical research progress and diagnosis and treatment strategy in this field, and provide reference for clinical related personnel, so as to improve the overall diagnosis and treatment level of thyroid tumors in China. **Methods** The high frequency subject terms were extracted from the PubMed on thyroid neoplasms field to generate the matrix of high frequency subject terms and their source papers. The co-occurrence of high frequency subject terms in a same paper was analyzed by clustering analysis. The hotspots in thyroid neoplasms were analyzed according to the clustering analysis of high frequency subject terms and their corresponding class labels. **Results** The hotspots in thyroid neoplasms were diagnostic imaging, genetics, and thyroidectomy and so on. **Conclusion** Through biclustering method analysis and literature review, it was necessary and feasible to understand the current situation of thyroid tumor research and the latest diagnosis and treatment methods.

Keywords: Thyroid neoplasms; Biclustering method; Visualization; Cluster analysis; Research hotspots

双聚类算法是在 1971 年由 Hartigan 提出的概念,又被称作同时聚类^[1],是对数据矩阵中的样本和变量同时进行聚类,即在对象及其属性两个方向上同时聚类,针对对象及其属性提取它们的联合信息,发现潜在的布局模式,双聚类算法在应用上比许多单向传统聚类方法更具有优势^[2]。

当前,我国甲状腺癌已经跃居女性恶性肿瘤的前 5 位,并成为近年来发病率增长最快的恶性肿瘤^[3]。基于此,本文对近 10 年甲状腺肿瘤的相关文献进行高频主题词共现聚类分析,探讨双向聚类分析方法在文本主题识别中的可行性,为临床相关人员提供参考,以便于提高我国甲状腺肿瘤整体的诊治水平。

* [基金项目] 济宁医学院大学生创新训练计划项目 (cx 2019028)

1 数据来源与方法

1.1 数据来源

文章的数据来源于 PubMed 数据库。检索方式采用主题词加权检索,检索策略为“Thyroid Neoplasms” [Majr] AND (“2014/01/01” [PDAT] : “2018/12/31” [PDAT]),检索到文献 9075 篇。

1.2 方法

以 XML 格式套录检索结果,利用书目共现分析系统 BICOMB^[4] 统计并抽取文献记录中的主要主题词与副主题词,按照出现频次由高到低进行排序,截取频次 ≥ 125 的 45 个主题词/副主题词作为高频词。其中出现频次最高的前 4 位主题词/副主题词分别为: 甲状腺肿瘤/病理学 (Thyroid Neoplasms/pathology)、甲状腺肿瘤/诊断 (Thyroid Neoplasms/diagnosis)、甲状腺肿瘤/外科 (Thyroid Neoplasms/surgery)、甲状腺肿瘤/遗传学 (Thyroid Neoplasms/genetics)。该 45 个高频词占有与甲状腺肿瘤相关主题词的累计比例为 42.50%。

根据高频词在每篇文献记录中出现的情况,对高频主题词进行统计,形成高频词词篇矩阵。将此篇矩阵输入 gCLUTO 软件,采用系统聚类法对所得词篇矩阵进行聚类分析,聚类分析的结果可以反映出高频词之间的亲疏关系,根据高频主题词的聚类结果以及主题词之间的语义关系,总结出甲状腺肿瘤领域的研究热点。

2 聚类结果

对高频主题词进行双聚类分析,经过多次聚类,发现类群数量为 5 时,山丘图的聚类效果较为理想。聚类结果见图 1,其中左侧是可视化矩阵,横向树代表文献,纵向树代表主题词和副主题词,每一类群的方格中颜色越深代表出现的频次越高;右下侧是每一类群对应的高频主题词;右上侧是对应的可视化山丘图,其中山峰体积越大,代表所含文献越多,峰顶的颜色代表了标准差,红色代表类内研究主题相似度高,蓝色代表类内研究主题相似度低。

首先,根据每一类高频词的含义以及这些主题词之间的语义关系,总结出每一类主题词所代表的研究热点,即甲状腺肿瘤的研究热点,例如主题词 Thyroid Neoplasms/drug therapy (甲状腺肿瘤/药物治疗) 和 Antineoplastic Agents/therapeutic use (抗肿瘤药物/治疗应用) 距离较近,关系密切,首先聚成一类;Thyroid Neoplasms/etiology (甲状腺肿瘤病因学) 再与前面两个词合成一类,依此类推。通过分析这些主题词的语义关系就能得出它们所代表的类团含义标签,综合各个类别的类标签可以得出该主题的研究热点。其次,利用 gCLUTO 软件计算各类成员对聚类贡献率的指标 (描述度 Descriptive 和区分度 Discriminating),选择对每一类形成贡献最大的来源文献作为表示该类内容的类标签文献^[5]。通常选取描述度分值最高者作为该类的类标签文献,确定类标签文献后再对文献内容进一步分析,进而阐释该类研究方向的具体内容。

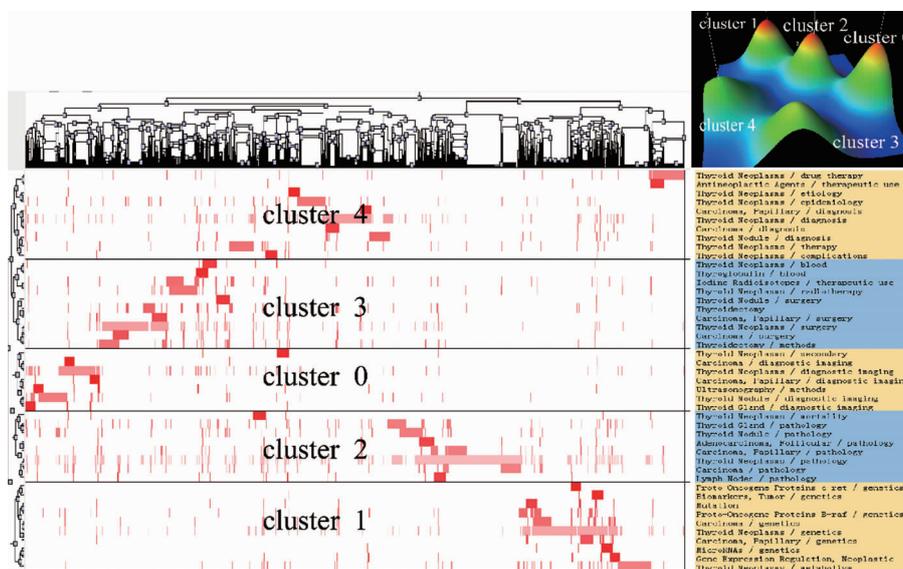


图 1 甲状腺肿瘤领域高频主题词的共现聚类图

3 分析

通过对 MEDLINE 收录的甲状腺肿瘤研究文献的高频主题词和副主题词进行共现聚类分析(图 1),可将该领域的研究热点分为 5 个主题。

3.1 甲状腺肿瘤影像学诊断方法应用(Cluster 0)

主题词有 Thyroid Neoplasms/secondary, Carcinoma/diagnostic imaging, Thyroid Neoplasms/diagnostic imaging, Papillary Carcinoma/diagnostic imaging, Ultrasonography/methods, Thyroid Nodule/diagnostic imaging, Thyroid Gland/diagnostic imaging。该类中山丘图峰顶颜色为红色,说明研究主题的相似度极高。主要研究内容为甲状腺乳头状癌、甲状腺结节的影像学诊断,如借助于 F-18 氟脱氧葡萄糖正电子发射断层扫描、放射性碘分子成像等进行影像学检查。各种先进的成像技术可以早期发现病变,并有助于作出临床决策,确保患者获得最佳的医学治疗效果。例如,甲状腺乳头状癌患者术前通过进行甲状腺包膜连续性和肿瘤轮廓的超声检查,可以区分宏观甲状腺外延伸(ete)和微观 ete 以及非 ete 和 ete,从而更好地进行手术^[6];镓-68 (GA-68)是一种以生长抑素受体(SSTR),特别是 2 亚型(SSTR2)为靶点的放射性标记肽成像方式。评估 GA-68 Dotatate 正电子发射断层扫描/计算机断层扫描(PET/CT)检测到的偶然不典型甲状腺摄取的频率和临床意义,在 GA-68 Dotatate PET/CT 扫描中,偶然出现局灶性甲状腺摄取异常的患者应进一步进行临床评估,以排除甲状腺癌的诊断^[7]。超声下甲状腺结节大小作为恶性肿瘤和最终病理大小的预测因子分析^[8]。

3.2 甲状腺肿瘤的基因突变及其遗传学(Cluster 1)

该类涉及的主题词包括 Proto-Oncogene Proteins c-ret/genetics, Tumor Biomarkers/genetics, Mutation, Proto-Oncogene Proteins B-raf/genetics, Carcinoma/genetics, Thyroid Neoplasms/genetics, Papillary Carcinoma/genetics, MicroRNAs/genetics, Neoplastic Gene Expression Regulation, Thyroid Neoplasms/metabolism。该类中山丘图峰顶颜色为红色,说明研究主题的相似度极高。主要研究内容为甲状腺肿瘤(甲状腺乳头状癌)的基因突变及其遗传学,包括信号通路的遗传和表观遗传变化。从分子机制角度阐释甲状腺肿瘤不同组织学类型与亚型的发

病机制,为制定更有效的甲状腺癌治疗策略提供广阔的前景。例如对同一患者(PTC/MTC)发生的乳头状(PTC)和髓质性(MTC)甲状腺癌进行遗传分析,分析 PTC/MTC 中 ret、braf 和 ras 癌基因突变情况,研究发现参与 PTC 和 MTC 发病机制的癌基因的经典激活突变可能与它们的同时发生无关^[9];微 RNA (micromas, microna)的解除是乳头状甲状腺癌发病机制中的一个重要事件,研究这些小 RNA 亚群对治疗乳头状甲状腺癌会有所帮助^[10]。

3.3 甲状腺肿瘤的病理学相关研究(Cluster 2)

包括的主题词有 Thyroid Neoplasms/mortality, Thyroid Gland/pathology, Thyroid Nodule/pathology, Follicular Adenocarcinoma/pathology, Papillary Carcinoma/pathology, Thyroid Neoplasms/pathology, Carcinoma/pathology, Lymph Nodes/pathology。该类中山丘图峰顶颜色为红色,说明研究主题的相似度极高。主要研究内容包括甲状腺肿瘤(甲状腺结节、甲状腺滤泡癌、乳头状甲状腺癌)的病理学特征。如具有良性超声特征的甲状腺癌根据结节大小具有不同的淋巴结转移风险和组织学亚型,较大的结节与较高的淋巴结转移(LNM)患病率相关^[11];甲状腺滤泡癌(FTC)的临床病理及预后特点^[12];细针穿刺活检(FNAB)是甲状腺结节评价的基础。在大多数情况下, FNAB 可以区分良性和恶性疾病。丹麦甲状腺外科数据库(THYKIR)记录了术前甲状腺 FNAB 和最终组织学。本研究的目的是根据丹麦标准评估可疑甲状腺 FNAB 患者的恶性肿瘤风险,并确定恶性组织学的可能预测因素^[13];评价甲状腺细针抽吸中可疑细胞学结果与组织学结果的相关性^[14]。

3.4 甲状腺肿瘤的切除术及术后的碘放射性同位素消融治疗(Cluster 3)

该类包括主题词 Thyroid Neoplasms/blood, Thyroglobulin/blood, Iodine Radioisotopes/therapeutic use, Thyroid Neoplasms/radiotherapy, Thyroid Nodule/surgery, Thyroidectomy, Papillary Carcinoma/surgery, Thyroid Neoplasms/surgery, Carcinoma/surgery, Thyroidectomy/methods。主要研究内容为甲状腺肿瘤的切除术方法及术后的碘放射性同位素消融治疗。主要集中于针对不同组织学类型的甲状腺肿瘤,采用不同的术式进行治疗,术后的辅助治疗及其治疗效果的评价。如术后低剂量非刺激甲状腺球蛋白作为备用碘消融术的标准^[15];单剂

量放射性碘治疗中晚期甲状腺乳头状癌甲状腺全切除术的疗效及对血清甲状腺球蛋白的影响^[16]；甲状腺球蛋白水平和甲状腺球蛋白/促甲状腺激素比值可预测¹³¹I 在分化型甲状腺癌中的消融/治疗成功率^[17]。

3.5 甲状腺肿瘤诊断、治疗及流行病学与病因学研究(Cluster 4)

该类涉及的主题词包括 Thyroid Neoplasms/drug therapy, Antineoplastic Agents/therapeutic use, Thyroid Neoplasms/etiology, Thyroid Neoplasms/epidemiology, Papillary Carcinoma/diagnosis, Thyroid Neoplasms/diagnosis, Carcinoma/diagnosis, Thyroid Nodule/diagnosis, Thyroid Neoplasms/therapy, Thyroid Neoplasms/complications。主要研究内容为甲状腺肿瘤诊断、治疗及流行病学与病因学研究。该类研究主题相似度不高。如甲状腺结节的鉴别诊断,通过更大样本验证 miRN-95 和-190 的血清表达水平是一种准确和无创的鉴别甲状腺结节的方法^[18]；X-连锁凋亡抑制剂作为甲状腺乳头状癌预后标志物和治疗靶点的作用^[19]；甲状腺髓样癌(MTC)治疗的临床特点、分子机制和当前分子和靶向治疗策略^[20]。

4 讨论

双聚类分析是对数据矩阵中的行和列同时进行聚类的一种聚类方法,可实现在对象及其属性两个方向上的同时聚类。在文本聚类中,部分词语仅在部分文档中显著出现,从而利用这些高频词汇可以推测文本主题,因此,本文应用双向聚类方法对国外甲状腺肿瘤相关文献进行聚类分析,总结出甲状腺肿瘤的研究热点主要集中在以下方面。1)甲状腺肿瘤的影像学诊断及其它诊断方法的研究。研究内容为借助各种先进的成像技术或者无创鉴别方法早期发现病变,确保患者获得最佳的医学治疗效果。2)甲状腺肿瘤治疗的相关研究。主要研究内容为针对不同组织学类型的甲状腺肿瘤切除手术方法,术后的辅助治疗及其治疗效果的评价。3)甲状腺肿瘤的病理学相关研究。研究内容包括甲状腺肿瘤(甲状腺结节、甲状腺滤泡癌、乳头状甲状腺癌)的病理学特征。4)甲状腺肿瘤(甲状腺乳头状癌)的基因突变及其遗传学,包括信号通路的遗传和表观遗传变化。从分子机制角度阐释甲状腺肿瘤不同组织学类型与亚型的发病机制,为制定更有效的甲状腺癌治疗策略提供广阔的前景。

通过双向聚类分析并对文献进行梳理,对认识甲状腺肿瘤的研究现状,了解其最新的诊疗方法是必要且现实可行的。

参考文献:

- [1] Hartigan JA. Direct clustering of a data matrix[J]. Publications of the American Statistical Association, 1972, 67(337): 123-129. DOI: 10.1080/01621459.1972.10481214.
- [2] 李向森. 基于双聚类和可视化的重金属污染领域研究热点探析[J]. 公共卫生与预防医学, 2016(6): 67-70.
- [3] 本刊编辑部. 《中国肿瘤临床》文章荐读: 甲状腺肿瘤专栏[J]. 中国肿瘤临床, 2018, 45(23): 1205.
- [4] 崔雷, 刘伟, 闫雷, 等. 文献数据库中书目信息共现挖掘系统的开发[J]. 现代图书情报技术, 2008, 24(8): 70-75.
- [5] 史航, 高雯璐, 崔雷, 等. 生物医学文本挖掘研究热点分析[J]. 中华医学图书情报杂志, 2016, 25(2): 27-33.
- [6] Rajeev PAhmed S, Ezzat TM, et al. The number of positive lymph nodes in the central compartment has prognostic impact in papillary thyroid cancer[J]. Langenbecks Arch Surg, 2013, 398(3): 377-382. DOI: 10.1007/s00423-012-1041-6.
- [7] Rim JH, Chong S, Ryu HS, et al. Feasibility study of ultrasonographic criteria for microscopic and macroscopic extra-thyroidal extension based on thyroid capsular continuity and tumor contour in patients with papillary thyroid carcinomas[J]. Ultrasound Med Biol, 2016, 42(10): 2391-2400. DOI: 10.1016/j.ultrasmedbio.2016.06.014.
- [8] Cavallo A, Johnson DN, White MG, et al. Thyroid nodule size at ultrasound as a predictor of malignancy and final pathologic size[J]. Thyroid, 2017, 27(5): 641. DOI: 10.1089/thy.2016.0336.
- [9] Nockel P, Millo C, Keutgen X, et al. The rate and clinical significance of incidental thyroid uptake as detected by Gallium-68 DOTATATE positron emission tomography/computed tomography[J]. Thyroid, 2016, 26(6): 831-835. DOI: 10.1089/thy.2016.0174.
- [10] Ciampi R, Romei C, Pieruzzi L, et al. Classical point mutations of RET, BRAF, and RAS, oncogenes are not shared in papillary and medullary thyroid cancer occurring simultaneously in the same gland[J]. J Endocrinol Invest, 2017, 40(1): 55-62. DOI: 10.1007/s40618-016-0526-5.

(下转第 452 页)

潜移默化的德育教育作用,丰富高校德育功能,弥补不足。实现其德育功能的不断提升,为培育学生做出应有的贡献。

参考文献:

- [1] 胡慧. 高校图书馆德育功能研究[D]. 武汉: 华中师范大学, 2014.
- [2] 吴徐年. 中外公共图书馆运动的共同特征[J]. 图书馆理论与实践, 2006(2): 76-77, 103. DOI: 10. 3969/j. issn. 1005-8214. 2006. 02. 034.
- [3] 林燕. 高校图书馆与大学生的思想政治教育[J]. 中国成人教育, 2011(8): 43-45.
- [4] 孙颖, 李婧, 王彤, 等. 高校图书馆参与网络舆情的德育功能探析[J]. 图书馆工作与研究, 2014(6): 26-30. DOI: 10. 3969/j. issn. 1005-6610. 2014. 06. 006.
- [5] 唐邦勤, 王梅. 高校图书馆思想政治工作职能的供给侧分析[J]. 图书馆情报工作, 2017, 61(S1): 26-29.
- [6] 洪筱萍. 高校图书馆在大学生思想政治教育中的作用与实现途径[J]. 学校党建与思想教育, 2010(16): 65-66. DOI: 10. 3969/j. issn. 1007-5968. 2010. 06. 025.

- [7] 樊健. 论大学图书馆的育人功能[J]. 大学图书馆学报, 2013, 31(5): 32-35. DOI: 10. 3969/j. issn. 1002-1027. 2013. 05. 006.
- [8] 杨淑珍, 秦瑞革, 曲一歌. “互联网+”时代高校图书馆思想政治教育探析[J]. 学校党建与思想教育, 2017(23): 76-78.
- [9] 沈宏斌. 网络环境下高校图书馆主页建设中的德育职能浅探[J]. 图书馆, 2009(5): 121-122. DOI: 10. 3969/j. issn. 1002-1558. 2009. 05. 043.
- [10] 耿艾莉. 基于 WEB2. 0 的高校图书馆思想政治教育方法创新研究[J]. 教育与职业, 2011(36): 185-186. DOI: 10. 3969/j. issn. 1004-3985. 2011. 36. 084.
- [11] 毕毅敏. 网络环境下高校图书馆对大学生德育教育的实现途径[J]. 中国成人教育, 2009(9): 58-59.
- [12] 付正芳, 张红伟, 刘盼英. 新阅读时代读者信息需求调查及资源建设对策[J]. 医学信息学杂志, 2014, 35(9): 69-73. DOI: 10. 3969/j. issn. 1673-6036. 2014. 09. 017.

(收稿日期 2019-03-27)

(本文编辑: 林琳)

(上接第 448 页)

- [11] Shin DY, Lee YK, Kim KJ, et al. Thyroid cancers with benign-looking sonographic features have different lymph node metastatic risk and histologic subtypes according to nodule size[J]. Endocr Pathol, 2014, 25(4): 378-384. DOI: 10. 1007/s12022-014-9327-6.
- [12] Yu XF, Wang WB, Teng XD, et al. Clinicopathological and prognostic analysis of follicular thyroid carcinoma in a single institute over a 15-year period[J]. Eur J Surg Oncol, 2014, 40(7): 869-874. DOI: 10. 1016/j. ejso. 2014. 01. 012.
- [13] Egset AV, Holm C, Larsen SR, et al. Risk of malignancy in fine-needle aspiration biopsy in patients with thyroid nodules[J]. Dan Med J, 2017, 64(2).
- [14] Baynes AL, Rio AD, Mclean C, et al. Fine-needle aspiration of the thyroid: correlating suspicious cytology results with histological outcomes[J]. Ann Surg Oncol, 2014, 21(5): 1653. DOI: 10. 1245/s10434-013-3448-x.
- [15] Mourão GF, Rosario PW, Calsolari MR. Low postoperative nonstimulated thyroglobulin as a criterion to spare radioiodine ablation[J]. Endocr Relat Cancer, 2015, 23(1): 47. DOI: 10. 1530/ERC-15-0458. pii: A5320.
- [16] Hughes DT, Miller BS, Cohen MS, et al. Outcomes of total thyroidectomy with therapeutic central and lateral neck dissection with a single dose of radioiodine in the treatment of regionally advanced papillary thyroid cancer

and effects on serum thyroglobulin[J]. Ann Surg Oncol, 2014, 21(5): 1647-1652. DOI: 10. 1245/s10434-013-3467-7.

- [17] Trevizam PGC, Tagliarini José Vicente, Castilho EC, et al. Thyroglobulin levels and thyroglobulin/thyrotropin ratio could predict the success of the ablative/therapeutic 131I in the differentiated thyroid cancers[J]. Endocr Res, 2016; 1-7. DOI: 10. 3109/07435800. 2016. 1173056.
- [18] Pilli T, Cantara S, Marzocchi C, et al. Diagnostic value of circulating microRNA-95 and-190 in the differential diagnosis of thyroid nodules: a validation study in 1000 consecutive patients[J]. Thyroid, 2017, 27(8): 1052-1057. DOI: 10. 1089/thy. 2017. 0035.
- [19] Hussain AR, Bu R, Ahmed M, et al. Role of X-linked inhibitor of apoptosis as a prognostic marker and therapeutic target in papillary thyroid carcinoma[J]. J Clin Endocrinol Metab, 2015, 100(7): E974-E985. DOI: 10. 1210/jc. 2014-4356.
- [20] Rajabi S, Hedayati M. Medullary thyroid cancer: clinical characteristics and new insights into therapeutic strategies targeting tyrosine kinases[J]. Mol Diagn Ther, 2017(2): 1-14. DOI: 10. 1007/s40291-017-0289-5.

(收稿日期 2019-10-15)

(本文编辑: 石俊强)