第一节

科技论文概述

内容提要

1. 为什么要写科技论文
2. 科技论文种类及结构
3. 科技论文出版及传播
4. 科技论文检索
5. 科技论文影响力评价

为什么要写科技论文

- 科技论文
  - 自然科学领域中用语言文字撰写的原始科研成果并以公开出版的科学文章。
- 为什么要写科技论文？
  - 同学术同行和社会共享你的原始研究成果
  - 综述其他人的研究
- 科技论文的重要性
  - 科技论文对于现代科技的发展和演化非常重要。每个科学工作者的工作和成就都是建立在他人的工作基础之上。（站在巨人的肩膀上）
  - 科学的道路上没有平坦的大路可走，只有那些在崎岖小路上攀登不畏劳苦的人，才有希望到达光辉的顶点。--马克思
  - 二十大提出的“教育强国、科技强国、人才强国”三位一体的发展思路和战略部署。

科技论文的基本属性

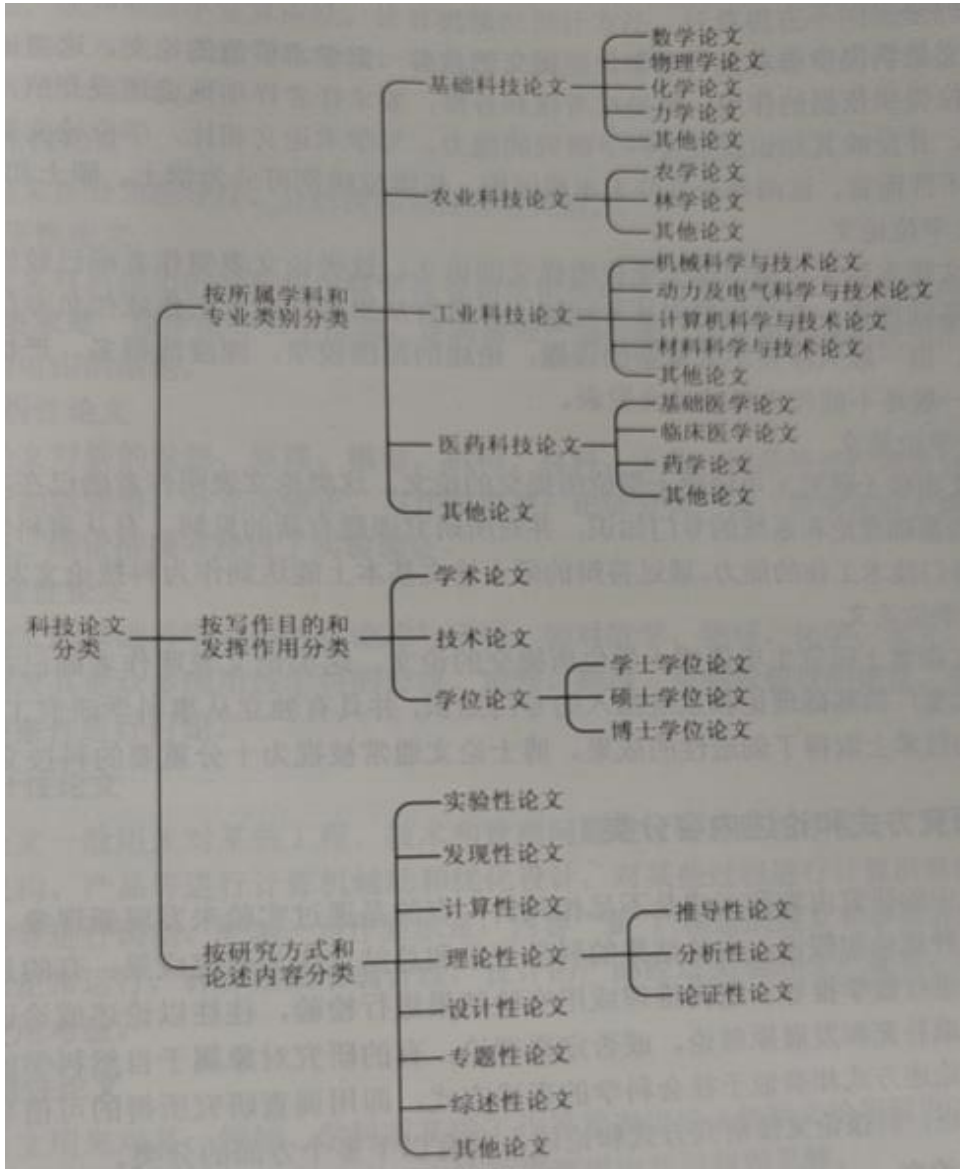
- 科学性
  - 论文的内容必须表述的科学，已事实为依据。
- 创新性
  - 新发现、新理论、新方法、新算法、新思想、新系统等，必须不同于，同时要优于现有的成果。
- 逻辑性
  - 论文表达的内容要正确（不能宣传伪科学）。
  - 文章结构脉络清晰，结构严谨，层次分明，推断合理，前后照应，自成体系。
- 通达性
  - 用词规范，文字表达通顺流畅，通理达意。符合大多数读者的阅读习惯和思考方式。

内容提要

1. 为什么要写科技论文
2. 科技论文种类及结构
3. 科技论文出版及传播
4. 科技论文检索
5. 科技论文影响力评价

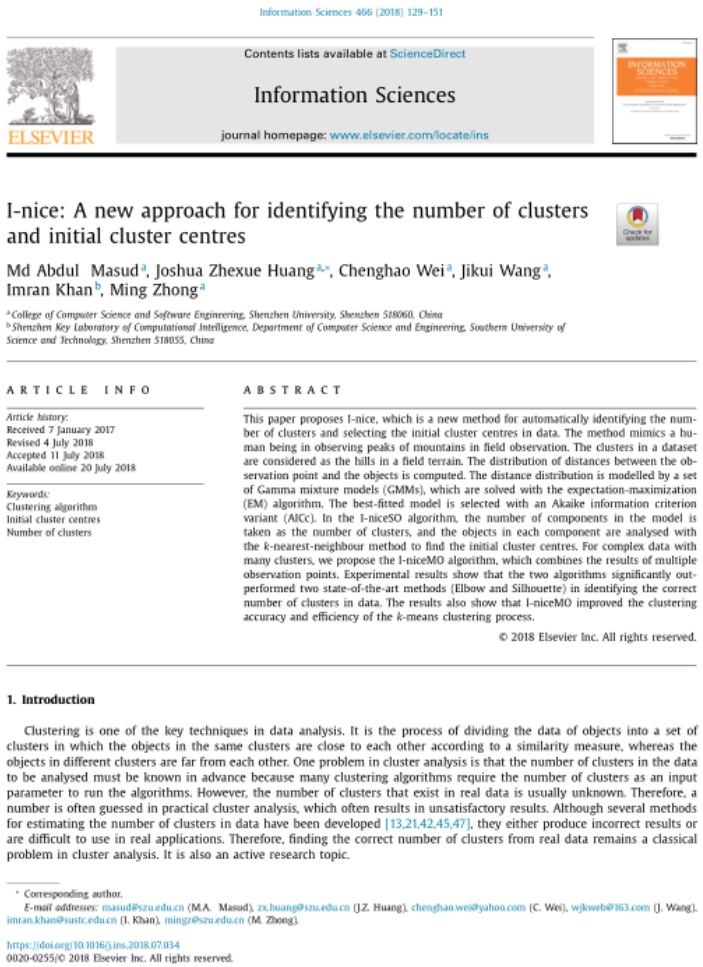
科技论文的类型

- 研究论文
- 研究快报
  - Letters, Communications, correspondence
- 研究简报
  - Notes，Erratum （勘误）
- 指导性论文 （Tutorial Paper ）
  - A tutorial paper provides novel and original insights and reflection on the use of one or several methods of modeling, design, analysis or synthesis in an acceptable format that can be used for guided and self-instruction.



研究论文

- 科研人员、研究生发表科研成果的主要手段。
- Scientific papers are the major means for researchers and graduate students to publish and disseminate their research results to academic communities and the public.
- **Publish or perish**.
  - A phrase coined to describe the pressure in academia to rapidly and continually publish academic work to sustain or further one's career. (Wikipedia)



Information Sciences 466 (2018) 129–151

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

I-nice: A new approach for identifying the number of clusters and initial cluster centres

Md Abdul Masud[a], Joshua Zhexue Huang[a,*], Chenghao Wei[a], Jikui Wang[a], Imran Khan[b], Ming Zhong[a]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[b] Shenzhen Key Laboratory of Computational Intelligence, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 7 January 2017
Revised 4 July 2018
Accepted 11 July 2018
Available online 20 July 2018

Keywords:
Clustering algorithm
Initial cluster centres
Number of clusters

ABSTRACT

This paper proposes I-nice, which is a new method for automatically identifying the number of clusters and selecting the initial cluster centres in data. The method mimics a human being in observing peaks of mountains in field observation. The clusters in a dataset are considered as the hills in a field terrain. The distribution of distances between the observation point and the objects is computed. The distance distribution is modelled by a set of Gamma mixture models (GMMs), which are solved with the expectation-maximization (EM) algorithm. The best-fitted model is selected with an Akaike information criterion variant (AICc). In the I-niceSO algorithm, the number of components in the model is taken as the number of clusters, and the objects in each component are analysed with the *k*-nearest-neighbour method to find the initial cluster centres. For complex data with many clusters, we propose the I-niceMO algorithm, which combines the results of multiple observation points. Experimental results show that the two algorithms significantly outperformed two state-of-the-art methods (Elbow and Silhouette) in identifying the correct number of clusters in data. The results also show that I-niceMO improved the clustering accuracy and efficiency of the *k*-means clustering process.

1. Introduction

Clustering is one of the key techniques in data analysis. It is the process of dividing the data of objects into a set of clusters in which the objects in the same clusters are close to each other according to a similarity measure, whereas the objects in different clusters are far from each other. One problem in cluster analysis is that the number of clusters in the data to be analysed must be known in advance because many clustering algorithms require the number of clusters as an input parameter to run the algorithms. However, the number of clusters that exist in real data is usually unknown. Therefore, a number is often guessed in practical cluster analysis, which often results in unsatisfactory results. Although several methods for estimating the number of clusters in data have been developed [13,21,42,45,47], they either produce incorrect results or are difficult to use in real applications. Therefore, finding the correct number of clusters from real data remains a classical problem in cluster analysis. It is also an active research topic.

* Corresponding author.
E-mail addresses: masud@szu.edu.cn (M.A. Masud), zx.huang@szu.edu.cn (J.Z. Huang), chenghao.wei@yahoo.com (C. Wei), wjkweb@163.com (J. Wang), imran.khan@sustc.edu.cn (I. Khan), mingz@szu.edu.cn (M. Zhong).

https://doi.org/10.1016/j.ins.2018.07.034
0020-0255/

研究论文基本构成

1. Title
2. Affiliations
3. Abstract
4. Key words
5. Introduction
6. Related work
7. Background and preliminaries
8. Method(s) and analysis
9. Experiment
   - Environment
   - Settings
   - Evaluation Methods
   - Results
   - Discussions
10. Conclusions
11. Acknowledgement
12. References
13. Appendix

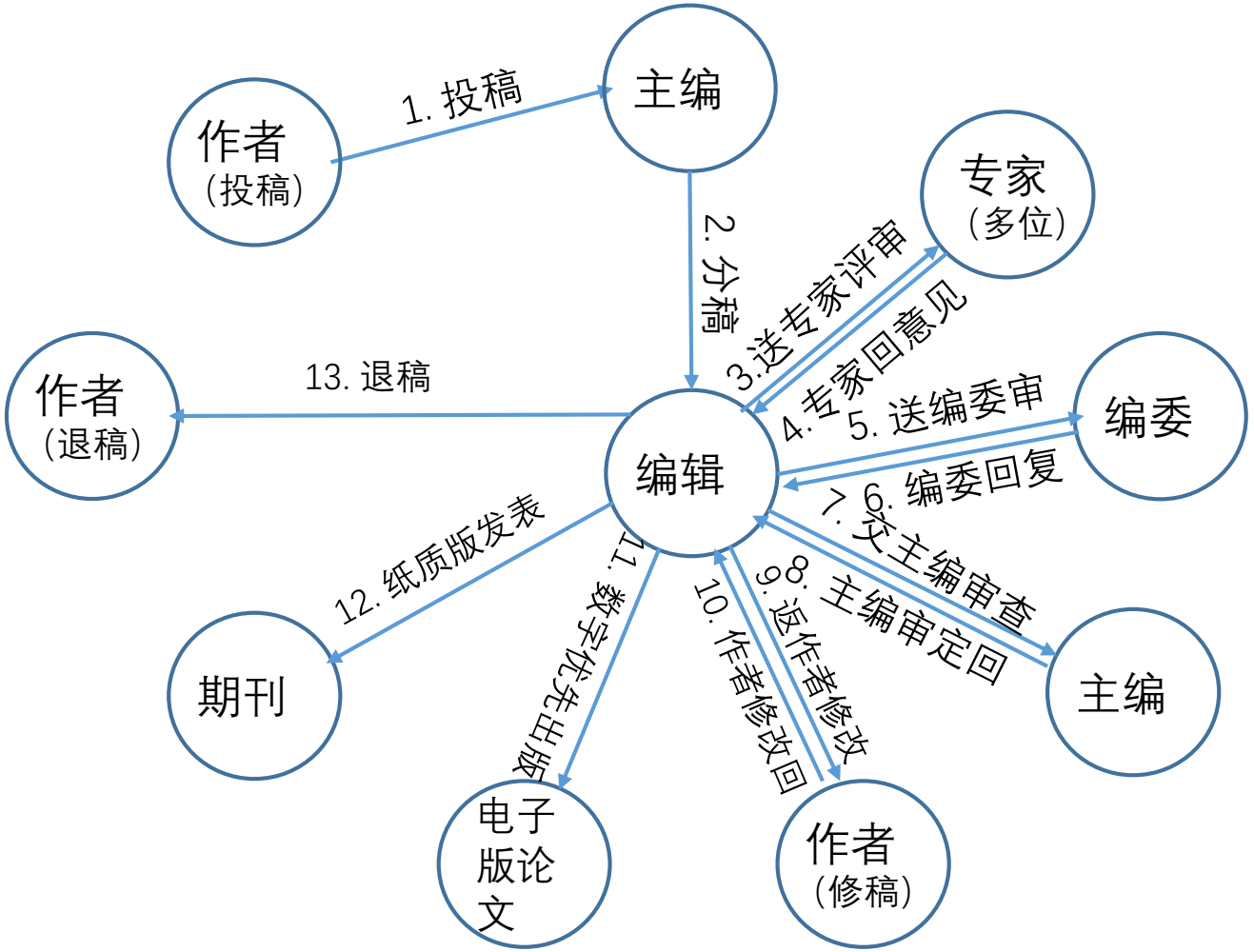
内容提要

1. 为什么要写科技论文
2. 科技论文种类及结构
3. 科技论文出版及传播
4. 科技论文检索
5. 科技论文影响力评价

科技论文出版及传播

- 公开发表
  - 一个新的科研成果公开发表，才能进行广泛传播。
  - 一个新的科研成果只有公开发表，作者才能获得发明权。
  - 新的科技成果通过科技论文出版公开发表，才有被广泛认可和接受的可能和机会。
- 出版及传播
- 学术期刊
- 国际会议论文集
- 技术杂志
- 公开预发表
- 自媒体

科技论文的规范发表


作者
（投稿）
1. 投稿
主编
2. 分稿
专家
（多位）
3.送专家评审
4.专家回意见
编委
5. 送编委审
6. 编委回复
主编
7. 交主编审查
8. 主编审定回
9. 返作者修改
作者
（修稿）
10. 作者修改回
电子版论文
11. 录入生出版
期刊
12. 纸质版发表
作者
（退稿）
13. 退稿
编辑

内容提要

1. 为什么要写科技论文
2. 科技论文种类及结构
3. 科技论文出版及传播
4. 科技论文检索
5. 科技论文影响力评价

科技论文检索

- 公开发表的论文应该能够公开检索到。
- 检索系统
  - 图书馆
  - 各种文献数据库
  - 搜索引擎
  - Google Scholar
  - Web of Sciences
  - 中文学术搜索 （知乎）

内容提要

1. 为什么要写科技论文
2. 科技论文种类及结构
3. 科技论文出版及传播
4. 科技论文检索
5. 科技论文影响力评价

论文学术影响力评价

- 是否成果被学术共同体和社会接受
  - 成果进入教科书、社会采用、进入产品、成果获奖
- 被引用情况
  - 广泛他引
- 同行评价
  - 被同学科知名专家认可、引用、正面评价
- 发表刊物和会议等级
  - 论文引用（citation）及影响因子 （impact factor）
  - 顶级国际会议或学术刊物（高影响因子）
  - 论文评级：CCF A、B、C; 中科院 1-4区；JCR Q1-Q4; CSRangking 论文；CNS 及自然指数 （Nature Index）

Criteria of Assessment for Student Thesis

1. Evidence of adequate and appropriate background reading
2. A clear statement of aims and relevant selection of content
3. Sensible planning and organization
4. Evidence of systematic thought and argument
5. Clarity of expression
6. Careful presentation (e.g. accurate typing and proof-reading, helpful diagrams, etc.)
7. Observation of conventions of academic discourse, including Bibliographic information
8. Observation of length requirements

抄袭的害处

- It is not helpful.
  - If you plagiarize, you suggest that something is your work when it is not. This will not get you good marks. To do well in higher education, you need to be responsible for the ideas and facts that you use in your writing. You need to provide evidence for these ideas and facts. You need to show where they have come from and what they are based on. You do this by acknowledging the sources, by citing. This will support your arguments and help you succeed in your academic writing. It will also show your lecturers that you have read and understood the required texts.
- You need to come to your own conclusions.
  - You need to show that you have understood the material and come to your own conclusions on the basis of what you have read and heard. Copying from textbooks, or pasting text from the Internet into your own writing, is not good enough. Most of what you write will come from the ideas of other people (from the textbooks you read, the lectures and the seminars you attend, and your discussions with other students, etc.). This is what academic study is all about. However, the purpose of an assignment is for you to say something for yourself using the ideas that you have studied, so you can present ideas you have learned in your own way. The emphasis should be on working with other people’s ideas, not simply reproducing their words.
- It is against the regulations.
  - You must not use another person's words or ideas as if they were your own. This is against university regulations and is considered a very serious offence. If you plagiarize, your lecturer cannot understand how well you understand the course and cannot therefore give you useful advice and support. In addition, if you plagiarize, you are not learning. This will become obvious in any written examination you are required to take.

抄袭的形式

1. Changing some of the words and sentences in a text, but keeping the overall structure of the text and the vocabulary the same as in the original text.
2. Taking some short fixed phrases from several different sources and putting them together with some of your own words.
3. Copying a paragraph directly from the source with no changes.
4. Copying a paragraph making only small changes - for example, replacing some words with words with similar meanings.
5. Copying out an article from a journal, website or textbook, and submitting it as your assignment.
6. Cutting and pasting a paragraph: using the sentences of the original, but putting one or two in a different order, and leaving one or two out.
7. Paraphrasing a paragraph: rewriting the paragraph but changing the language, organization and detail, and giving your own examples.
8. Quoting a paragraph by placing it in quotation marks and acknowledging the source.
9. Rewriting a passage from a source and presenting it as your own work.
10. Taking just one word or phrase from a text, because it is very well expressed.
11. Using another author's organization and way of arguing

使用大模型要小心

- 用大模型检查句子的正确性，修饰文字表达是可以的。
- 用大模型来帮助写论文会带来抄袭的风险。
- 用大模型来代替你个人的写作只会是你的写作能力停滞不前。
- 大模型写的东西不能保证正确性，可能产生谬误。

论文写作

- 写一篇论文，概述MapReduce和Non-MapReduce分布式计算框架支持复杂算法大数据计算的步骤、特点和优缺点。
- Non-MapReduce参考文献：
  - Xudong Sun; Yulin He; Dingming Wu; Joshua Zhexue Huang, Survey of Distributed Computing Frameworks for Supporting Big Data Analysis, Big Data Mining and Analytics, 2023,6(2):154-169
  - Xudong Sun; Zhao Lingxiang; Chen Jiaqi; Cai Yongda; Wu Dingming, Joshua Zhexue Huang, Non-MapReduce computing for intelligent big data analysis, ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE, 2024
  - Xudong Sun, Dingming Wu, Yongda Cai, Lingxiang Zhao,Changda Xiao, Joshua Zhexue Huang. MapReduce vs. Non-MapReduce Efficiency and Scalability in Big Data Computing. World Congress: "System Theory, Algebraic Biology, Artificial Intelligence". 2023 from https://publ.icgbio.ru/wp-content/uploads/2023/12/SELECTED-WORKS-Congress26-300623-244-280.pdf
  - Alternate framework for distributed computing tames Big Data's ever growing costs (2023, February 23) retrieved 9 September 2024 from https://techxplore.com/news/2023-02-alternate-framework-big.pdf
- MapReduce 参考文献：
  - 自己查找