[[1]](#footnote-1)

PSOkmeans算法

First A. Author, *Fellow, IEEE*, Second B. Author, and Third C. Author, Jr., Member, IEEE

*Abstract*—These instructions give you guidelines for preparing papers for IEEE Transactions and Journals*.* Use this document as a template if you are using Microsoft *Word* 6.0 or later. Otherwise, use this document as an instruction set. The electronic file of your paper will be formatted further at IEEE. Paper titles should be written in uppercase and lowercase letters, not all uppercase. Avoid writing long formulas with subscripts in the title; short formulas that identify the elements are fine (e.g., "Nd–Fe–B"). Do not write “(Invited)” in the title. Full names of authors are preferred in the author field, but are not required. Put a space between authors’ initials. Define all symbols used in the abstract. Do not cite references in the abstract. Do not delete the blank line immediately above the abstract; it sets the footnote at the bottom of this column.

*Index Terms*—Enter key words or phrases in alphabetical order, separated by commas.

# 引言

随着生命科学前所未有的高速发展，产生的生物信息也急剧增加，为了处理这些庞大的数据集，人们巧妙的将计算机科学等分析工具和方法引入到生命科学研究中来，形成了生物信息学基于PSO的基因表达数据聚类研究

癌症是一种通过多种途径演变的异质性疾病，涉及多种癌基因和肿瘤抑制基因活性的变化。这些变化的基础是产生复杂分子和细胞现象类型的大量和多样的体细胞替代物，影响每个个体肿瘤的行为和反应治疗。由于突变和分子机制的多样性，结果差异很大。因此，将机器学习的方法对癌症病人的基因表达情况进行研究，对于识别病症亚型，癌症基因表达的内在联系识别有着重要的意义Multi-omic tumor data reveal diversity of molecularmechanisms that correlate with survival

对于基因数据常常是高维数据这一特点，国内外有学者采用了各种不同降维算法，相似度学习方法，来降低待测样本的维度例如相似度网络融合SNF，tsne，icluster+等，也有学者根据多核学习理论提出了一种优化求解相似度矩阵的方法Multi-omic tumor data reveal diversity of molecularmechanisms that correlate with survival。

以上这些方法，对于高维数据集都有着不错效果，但是优化过程比较复杂，涉及到很多数学理论知识。因此，我们想要采用启发式算法，同时结合较为简单降维算法，对高维数据集进行聚类分析。

# MATH

## K-均值算法

在所有的聚类算法之中，K-means聚类算法可以说是众所周知的简单实用的经典聚类算法，数学定义如下：

给定样本集 ,K-means算法的目的就是对于根据算法划分个簇有最小化的平方误差。



其中的 是每个簇自身的质心，也就是这整个簇的均值向量。

算法的大体步骤如下：首先，它根据某种算法先选取个样本，把这个样本作为这个簇的样本均值向量，也就是这一个簇的质心。对于其余每个样本，计算出样本与簇质心之间的距离，将这个样本分配给最相近的簇。在这之后下一步计算每个簇的新的心。一直不断重复这个过程，直到这个过程迭代到目标函数收敛[15]。

## 粒子群算法

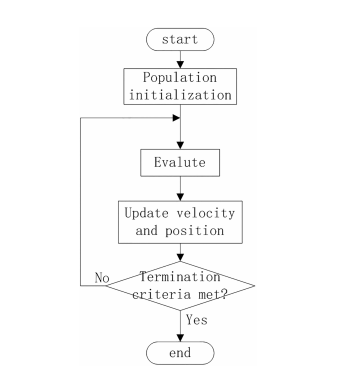
粒子群优化算法(PSO)是一种进化计算技术，它模拟鸟群的运动行为建模

假设这里有个粒子，所以种群即为，每一个粒子的位置为，速度是，每一个粒子的个体最优值是，全局最优值是，速度和位置的更新公式如下：





公式中代表迭代次数，是惯性权重，是学习率，是两个独立的随机参数，取值在0到1之间，算法的整体流程如下图所示，首先随机初始化种群，然后计算每个粒子的适应度，得到个体最优值和全局最优值，接着按照公式更新每个粒子的位置和速度，最后看能否达到迭代终止条件，若不能则继续从第二步开始迭代。



## 粒子群K-均值算法

我们在进行优化过程中，把一个粒子视为一组质心坐标。选取一定数量的粒子，进行随机初始化，采用的评价标准是Calinski-Harabasz分数。聚类效果的目标函数如下



其中的是训练集样本数，是类别数。是类别之间协方差矩阵，是类别内部数据的协方差矩阵。为矩阵的迹。

根据粒子群算法，我们初始化一定数量的质心粒子，根据目标函数值及更新公式不断优化更新粒子，找到最好的质心位置。

# 实验

实验数据集是来自TCGA的低级胶质瘤数据集一共282个病例样本，相关基因有28920个，每个基因有基因表达量，拷贝数替代，甲基化，点变异四种测量值。

对相关的四个数据集进行清洗融合

# Some Common Mistakes

The word “data” is plural, not singular. The subscript for the permeability of vacuum µ0 is zero, not a lowercase letter “o.” The term for residual magnetization is “remanence”; the adjective is “remanent”; do not write “remnance” or “remnant.” Use the word “micrometer” instead of “micron.” A graph within a graph is an “inset,” not an “insert.” The word “alternatively” is preferred to the word “alternately” (unless you really mean something that alternates). Use the word “whereas” instead of “while” (unless you are referring to simultaneous events). Do not use the word “essentially” to mean “approximately” or “effectively.” Do not use the word “issue” as a euphemism for “problem.” When compositions are not specified, separate chemical symbols by en-dashes; for example, “NiMn” indicates the intermetallic compound Ni0.5Mn0.5 whereas “Ni–Mn” indicates an alloy of some composition NixMn1-x.

Be aware of the different meanings of the homophones “affect” (usually a verb) and “effect” (usually a noun), “complement” and “compliment,” “discreet” and “discrete,” “principal” (e.g., “principal investigator”) and “principle” (e.g., “principle of measurement”). Do not confuse “imply” and “infer.”

Prefixes such as “non,” “sub,” “micro,” “multi,” and “ultra” are not independent words; they should be joined to the words they modify, usually without a hyphen. There is no period after the “et” in the Latin abbreviation “*et al.*” (it is also italicized). The abbreviation “i.e.,” means “that is,” and the abbreviation “e.g.,” means “for example” (these abbreviations are not italicized).

# Guidelines for Graphics Preparation and Submission

## Types of Graphics

The following list outlines the different types of graphics published in IEEE journals. They are categorized based on their construction, and use of color / shades of gray:

### *Color/Grayscale figures*

### Figures that are meant to appear in color, or shades of black/gray. Such figures may include photographs, illustrations, multicolor graphs, and flowcharts.

### *Lineart figures*

### Figures that are composed of only black lines and shapes. These figures should have no shades or half-tones of gray. Only black and white.

### *Author photos*

### Head and shoulders shots of authors which appear at the end of our papers.

### *Tables* Data charts which are typically black and white, but sometimes include color.

## Multipart figures

Figures compiled of more than one sub-figure presented side-by-side, or stacked. If a multipart figure is made up of multiple figure types (one part is lineart, and another is grayscale or color) the figure should meet the stricter guidelines.

## File Formats For Graphics

Format and save your graphics using a suitable graphics processing program that will allow you to create the images as PostScript (PS), Encapsulated PostScript (.EPS), Tagged Image File Format (.TIFF), Portable Document Format (.PDF), or Portable Network Graphics (.PNG) sizes them, and adjusts the resolution settings. If you created your source files in one of the following programs you will be able to submit the graphics without converting to a PS, EPS, TIFF, PDF, or PNG file: Microsoft Word, Microsoft PowerPoint, or Microsoft Excel. Though it is not required, it is recommended that these files be saved in PDF format rather than DOC, XLS, or PPT. Doing so will protect your figures from common font and arrow stroke issues that occur when working on the files across multiple platforms. When submitting your final paper, your graphics should all be submitted individually in one of these formats along with the manuscript.

## Sizing of Graphics

Most charts, graphs, and tables are one column wide (3.5 inches / 88 millimeters / 21 picas) or page wide (7.16 inches / 181 millimeters / 43 picas). The maximum depth a graphic can be is 8.5 inches (216 millimeters / 54 picas). When choosing the depth of a graphic, please allow space for a caption. Figures can be sized between column and page widths if the author chooses, however it is recommended that figures are not sized less than column width unless when necessary.

There is currently one publication with column measurements that don’t coincide with those listed above. Proceedings of the IEEE has a column measurement of 3.25 inches (82.5 millimeters / 19.5 picas).

The final printed size of author photographs is exactly   
1 inch wide by 1.25 inches tall (25.4 millimeters x 31.75 millimeters / 6 picas x 7.5 picas). Author photos printed in editorials measure 1.59 inches wide by 2 inches tall (40 millimeters x 50 millimeters / 9.5 picas x 12 picas).

## Resolution

The proper resolution of your figures will depend on the type of figure it is as defined in the “Types of Figures” section. Author photographs, color, and grayscale figures should be at least 300dpi. Lineart, including tables should be a minimum of 600dpi.

## Vector Art

While IEEE does accept, and even recommends that authors submit artwork in vector format, it is our policy is to rasterize all figures for publication. This is done in order to preserve the figures’ integrity across multiple computer platforms.

# Conclusion

## A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

References

*Basic format for books:*

1. J. K. Author, “Title of chapter in the book,” in *Title of His Published Book, x*th ed. City of Publisher, Country if not
2. USA: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx.*

*Examples:*

1. G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics,* 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, *Linear Networks and Systems.* Belmont, CA: Wadsworth, 1993, pp. 123–135.

*Basic format for periodicals:*

1. J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. *x,* no. *x,* pp*. xxx-xxx,* Abbrev. Month, year.

*Examples:*

1. J. U. Duncombe, “Infrared navigation—Part I: An assessment   
   of feasibility,” *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959.
2. E. P. Wigner, “Theory of traveling-wave optical laser,” *Phys. Rev*.,   
   vol. 134, pp. A635–A646, Dec. 1965.
3. E. H. Miller, “A note on reflector arrays,” *IEEE Trans. Antennas Propagat*., to be published.

*Basic format for reports:*

1. J. K. Author, “Title of report,” Abbrev. Name of Co., City of Co., Abbrev. State, Rep. *xxx*, year.

*Examples:*

1. E. E. Reber, R. L. Michell, and C. J. Carter, “Oxygen absorption in the earth’s atmosphere,” Aerospace Corp., Los Angeles, CA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.
2. J. H. Davis and J. R. Cogdell, “Calibration program for the 16-foot antenna,” Elect. Eng. Res. Lab., Univ. Texas, Austin, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.

*Basic format for handbooks:*

1. *Name of Manual/Handbook*, *x* ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. *xxx-xxx.*

*Examples:*

1. *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44–60.
2. *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

*Basic format for books (when available online):*

1. Author. (year, month day). *Title.* (edition) [Type of medium]. *volume (issue).* Available: site/path/file

*Example:*

1. J. Jones. (1991, May 10). *Networks.* (2nd ed.) [Online]. Available: [http://www.atm.com](http://www.atm.com/)

*Basic format for journals (when available online):*

1. Author. (year, month). Title. *Journal.* [Type of medium]. *volume (issue),* pages. Available: site/path/file

*Example:*

1. R. J. Vidmar. (1992, Aug.). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. *21(3),* pp. 876–880. Available:<http://www.halcyon.com/pub/journals/21ps03-vidmar>

*Basic format for papers presented at conferences (when available online):*

1. Author. (year, month). Title. Presented at Conference title. [Type of Medium]. Available: site/path/file

*Example:*

1. PROCESS Corp., MA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at   
   INET96 Annual Meeting. [Online]. Available: <http://home.process.com/Intranets/wp2.htp>

*Basic format for reports and handbooks (when available online):*

1. Author. (year, month). Title. Comp an y . C ity, State or Country. [Type of Medium]. Available: site/path/file

*Example:*

1. S. L. Tall een. (1996 , Apr . ). The In t r an et Archi -tecture: M a nagi ng i n f o rm at i on i n t h e ne w paradigm. Amdahl Corp., CA. [Online]. Available:<http://www.amdahl.com/doc/products/bsg/intra/infra/html>

*Basic format for computer programs and electronic documents (when available online):* ISO recommends that capitalization follow the accepted practice for the language or script in which the information is given.

*Example:*

1. A. Harriman. (1993, June). Compendium of genealogical software. *Humanist.* [Online]. Available e-mail: [HUMANIST@NYVM.ORG](mailto:HUMANIST@NYVM.ORG) Message: get GENEALOGY REPORT

*Basic format for patents (when available online):*

1. Name of the invention, by inventor’s name. (year, month day). *Patent Number* [Type of medium]. Available: site/path/file

*Example:*

1. Musical toothbrush with adjustable neck and mirror, by L.M.R. Brooks. (1992, May 19). *Patent D 326 189*

[Online]. Available: NEXIS Library: LEXPAT File: DESIGN

*Basic format for conference proceedings (published):*

1. J. K. Author, “Title of paper,” in *Abbreviated Name of Conf.*, City of Conf., Abbrev. State (if given), year, pp. *xxxxxx.*

*Example:*

1. D. B. Payne and J. R. Stern, “Wavelength-switched pas- sively coupled single-mode optical network,” in *Proc. IOOC-ECOC,* 1985,   
   pp. 585–590.

*Example for papers presented at conferences (unpublished):*

1. D. Ebehard and E. Voges, “Digital single sideband detection for interferometric sensors,” presented at the 2nd Int. Conf. Optical Fiber Sensors, Stuttgart, Germany, Jan. 2-5, 1984.

*Basic format for patents:*

1. J. K. Author, “Title of patent,” U.S. Patent *x xxx xxx*, Abbrev. Month, day, year.

*Example:*

1. G. Brandli and M. Dick, “Alternating current fed power supply,”   
   U.S. Patent 4 084 217, Nov. 4, 1978.

*Basic format**for theses (M.S.) and dissertations (Ph.D.):*

1. J. K. Author, “Title of thesis,” M.S. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.
2. J. K. Author, “Title of dissertation,” Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

*Examples:*

1. J. O. Williams, “Narrow-band analyzer,” Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993.
2. N. Kawasaki, “Parametric study of thermal and chemical nonequilibrium nozzle flow,” M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

*Basic format for the most common types of unpublished references:*

1. J. K. Author, private communication, Abbrev. Month, year.
2. J. K. Author, “Title of paper,” unpublished.
3. J. K. Author, “Title of paper,” to be published.

*Examples:*

1. A. Harrison, private communication, May 1995.
2. B. Smith, “An approach to graphs of linear forms,” unpublished.
3. A. Brahms, “Representation error for real numbers in binary computer arithmetic,” IEEE Computer Group Repository, Paper R-67-85.

*Basic format for standards:*

1. *Title of Standard*, Standard number, date.

*Examples:*

1. IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.
2. Letter Symbols for Quantities, ANSI Standard Y10.5-1968.

1. [↑](#footnote-ref-1)