

摘要：考虑应急决策问题，基于 Jaccard、Dice 和余弦相似度，提出了一种新的相似度定义——变异系数相似度。介绍了四种相似度的特点，然后将变异系数相似性应用于应急群决策问题。通过每一个方案和目标方案的相似度，可以得出所有方案的一个排序，然后找出最优方案。最后，举例评价中国湖南省重大冰雪灾害急救管理能力。作为比较，也给出了不同的相似性度量的决策结果。从结果看出，新方法具有更好的相似性识别，适用于紧急群决策应用。

关键词：相似性；应急决策；变异系数；多准则群决策

中图分类号：O231.1

文献标识码：A

文章编号：1006-883X(2018)07-0026-05

收稿日期：2018-05-21

变异系数相似度在紧急群决策中的应用

冯长敏

北京信息科技大学 理学院，北京 100192

一、引言

自然灾害事件，如地震和飓风等，常常会造成严重的后果，为了解决这种情况，管理部门需要采取应急计划和管理。也需要平时加强自然环境的治理，提高自然稳定性。随着经济、社会、自然条件的变化，管理决策的环境也变得日益复杂。

在决策过程中，需要考虑多方面的因素。王坚强等^[1]提出了多属性群决策方法，基于算例分析验证所提多准则群决策方法的有效性与合理性。然而，相似性度量能够对两个事物之间相近程度进行综合评定。在实体人类学、数值分类学、生态学、信息检索、心理学、引文分析和自动分类等方面都使用了相似性度量^[2]。其中，研究对象之间的相似性程度起到重要作用。在多准则群决策环境中，按各自的偏好对目标进行评价，从中寻求群满意目标或者排序。

Jaccard、Dice 和余弦相似度经常被用来达到这个目的。Ye^[3]用梯形模糊数的向量相似度，提出了一种多准则群决策方法，他延伸了 Jaccard、Dice 和余弦相似度。通过每一个方案和理想方案的权重相似度，可以得出所有方案的一个排序，然后找出最优方案。

本文以上述三种相似度为基础，提出一种新的变异系数相似度，并应用在紧急群决策中，研究结果表明，所提出的方法具有更好的相似性识别，适用于事件紧急群决策应用。

二、向量相似度

1、四种向量相似度的定义

定义 1 对于 $X=(x_1, x_2, \dots, x_n)$ 和 $Y=(y_1, y_2, \dots, y_n)$ 两个长度为 n 的向量（其中每一项都是正数），两个向量的 Jaccard 相似度定义如下^[4]：

$$J(X, Y) = \frac{XY}{\|X\|_2^2 + \|Y\|_2^2 - XY} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i y_i} \quad (1)$$

其中， $XY = \sum_{i=1}^n x_i y_i$ —向量 X 和向量 Y 的内积；

$$\|X\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \text{—} X \text{ 的二范数；}$$

$$\|Y\|_2 = \sqrt{\sum_{i=1}^n y_i^2} \text{—} Y \text{ 的二范数。}$$

定义2 对于 $X=(x_1, x_2, \dots, x_n)$ 和 $Y=(y_1, y_2, \dots, y_n)$ 两个长度为 n 的向量（其中每一项都是正数），两个相量的 Dice 相似度定义如下^[5]：

$$E(X, Y) = \frac{2XY}{\|X\|_2^2 + \|Y\|_2^2} = \frac{2 \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2} \quad (2)$$

定义3 对于 $X=(x_1, x_2, \dots, x_n)$ 和 $Y=(y_1, y_2, \dots, y_n)$ 两个长度为 n 的向量（其中每一项都是正数），两个相量的余弦相似度定义如下^[6]：

$$C(X, Y) = \frac{XY}{\|X\|_2 \cdot \|Y\|_2} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (3)$$

这三个公式是类似的，他们都在区间 $[0, 1]$ 之间取值。但是上面三个定义都有一些缺点。对于所有的 $i(i=1, 2, \dots, n)$ ， $x_i=y_i=0$ ，Jaccard 和 Dice 公式没有定义，对于所有的 $i(i=1, 2, \dots, n)$ ， $x_i=0$ 或 $y_i=0$ ，余弦相似度公式没有定义。

根据这三个相似度的缺点，我们在定义4中定义了一种新的相似度——变异系数相似度。

定义4 对于 $X=(x_1, x_2, \dots, x_n)$ 和 $Y=(y_1, y_2, \dots, y_n)$ 两个长度为 n 的向量（其中每一项都是正数），变异系数相似度定义如下^[7]：

$$\begin{aligned} V(X, Y) &= \alpha \frac{2XY}{\|X\|_2^2 + \|Y\|_2^2} + (1-\alpha) \frac{XY}{\|X\|_2 \|Y\|_2} \\ &= \alpha \frac{2 \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2} + (1-\alpha) \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \end{aligned} \quad (4)$$

其中， α —变化的系数，且 $0 \leq \alpha \leq 1$ 。

在定理1中，我们需要证明向量 X 和向量 Y 之间的变异系数相似度满足下面一些特点。

定理1 令 $X=(x_1, x_2, \dots, x_n)$ 和 $Y=(y_1, y_2, \dots, y_n)$ 是两个长度为 n 的向量，变异系数相似度满足下面一些特点：

$$(P_1) 0 \leq V(X, Y) \leq 1 \quad (5)$$

$$(P_2) V(X, Y) = V(Y, X) \quad (6)$$

$$(P_3) \text{ 对于 } i=1, 2, \dots, n, \text{ 如果 } X=Y, \text{ 即 } x_i=y_i, \text{ 则}$$

$$V(X, Y)=1 \quad (7)$$

(P₁) 证明：很明显 $V(X, Y) \geq 0$ ，因此我们仅需要证明 $V(X, Y) \leq 1$

$$\begin{aligned} \text{因为: } \|X\|_2^2 + \|Y\|_2^2 &= x_1^2 + x_2^2 + \dots + x_n^2 + y_1^2 + y_2^2 + \dots + y_n^2 \\ 2XY &= 2(x_1 y_1 + x_2 y_2 + \dots + x_n y_n) \end{aligned}$$

$$\begin{aligned} \text{所以有: } \|X\|_2^2 + \|Y\|_2^2 - 2XY &= (x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots \\ &+ (x_n - y_n)^2 \geq 0 \end{aligned}$$

又因为 X 和 Y 中每一项都是正数，所以 $XY \geq 0$ ，

$$\text{且 } \frac{XY}{\|X\|_2 \|Y\|_2} = \cos \theta \leq 1 \quad (\theta \text{ 为向量 } X \text{ 与 } Y \text{ 的夹角})，\text{ 有：}$$

$$\begin{cases} 0 \leq \frac{2XY}{\|X\|_2^2 + \|Y\|_2^2} \leq 1 \\ 0 \leq \frac{XY}{\|X\|_2 \|Y\|_2} \leq 1 \end{cases} \quad (8)$$

将(8)式代入(5)式得：

$$V(X, Y) \leq \alpha + (1-\alpha) = 1 \quad (9)$$

(P₂) 证明：由式(5)有：

$$\begin{aligned} V(X, Y) &= \alpha \frac{2XY}{\|X\|_2^2 + \|Y\|_2^2} + (1-\alpha) \frac{XY}{\|X\|_2 \|Y\|_2} \\ &= \alpha \frac{2YX}{\|Y\|_2^2 + \|X\|_2^2} + (1-\alpha) \frac{YX}{\|Y\|_2 \|X\|_2} = V(Y, X) \end{aligned} \quad (10)$$

(P₃) 证明：当 $X=Y$ ，也即对于所有的 $i=1, 2, \dots, n$ ，都有 $x_i=y_i$ 成立，则由式(5)有：

$$\begin{aligned} V(X, Y) &= \alpha \frac{2XY}{\|X\|_2^2 + \|Y\|_2^2} + (1-\alpha) \frac{XY}{\|X\|_2 \|Y\|_2} \\ &= \alpha \frac{2 \sum_{i=1}^n x_i x_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i^2} + (1-\alpha) \frac{\sum_{i=1}^n x_i x_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n x_i^2}} \\ &= \alpha + (1-\alpha) = 1 \end{aligned} \quad (11)$$

由此，证明完毕。

2、不同相似度的比较

表1中列出了18个向量，并且偏序向量为 $a=[0.8542, 0.3460, 0.7000, 0.789, 0.1234, 0.6534]$ ，令 $\alpha=0.1$ ， $\alpha=0.75$ ， $\alpha=0.5$ 和 $\alpha=0.9$ ，计算出这18个向量和 a 的相似度^[8]。

表 1 群成员的偏序向量

城市	G_M	A_1	A_2	A_3	A_4	A_5	A_6
湘潭	V_1	0.000	0.248	0.076	0.561	0.208	0.248
	V_2	0.529	0.445	0.072	0.404	0.538	0.664
	V_3	0.666	0.623	0.911	0.910	0.931	0.844
	V_4	0.402	0.000	0.326	0.167	0.136	0.664
	V_5	1.000	0.898	0.496	0.850	0.733	0.665
	V_6	0.463	0.898	0.094	0.404	0.538	0.664
娄底	V_7	0.378	0.571	0.000	0.705	0.638	0.532
	V_8	0.485	0.445	0.826	0.889	0.282	0.278
	V_9	0.206	0.500	0.669	0.650	0.834	0.480
	V_{10}	0.456	0.316	0.816	0.924	0.282	0.471
	V_{11}	0.079	0.016	0.326	0.116	0.000	0.138
	V_{12}	0.258	1.000	0.326	1.000	0.899	0.862
郴州	V_{13}	0.425	0.180	0.594	0.000	0.367	0.000
	V_{14}	0.552	0.398	1.000	1.000	1.000	1.000
	V_{15}	0.344	0.180	0.679	0.690	0.282	0.187
	V_{16}	0.398	0.398	0.499	0.817	0.733	0.741
	V_{17}	0.715	0.698	0.924	0.017	0.598	0.478
	V_{18}	0.689	0.291	0.326	0.954	0.282	0.664

表工程收益、 A_3 是维护运行费、 A_4 代表治理效果、 A_5 是风险性、 A_6 是对环境的影响。

提出的这种方法的决策过程可以概括如下：

步骤 1：把文献 [10] 中 4 个方案的偏向量用下面这种方法进行标准化：

$$Z_{ij} = \frac{y_{ij} - y_j^{\min}}{y_j^{\max} - y_j^{\min}} \tag{12}$$

得到三个城市的偏序向量：湘潭（ $V_1 \sim V_6$ ）、娄底（ $V_7 \sim V_{12}$ ）和郴州（ $V_{13} \sim V_{18}$ ），并在表 1 中列出；

步骤 2：从六个专家可以看出，目标偏序向量是 $a=[0.8542, 0.346, 0.7, 0.789, 0.1234, 0.6534]$ 。令 $\alpha=0.75$ ，由公式（4），得到 18 个向量 $V_i(i=1, 2, \dots, 18)$ 和目标偏序向量的相似度 [11]；

步骤 3：用 18 个向量和目标偏序向量的相似度，给出此方法的决策矩阵 $D=(d_{ij})_{n \times n}$ 和权重值：

$$D_{Xt} = \begin{bmatrix} 0.3626 & 0.5323 & 0.7112 & 0.5770 \\ 0.6399 & 0.7804 & 0.8109 & 0.7880 \\ 0.7655 & 0.8672 & 0.8973 & 0.8747 \\ 0.5755 & 0.7306 & 0.8565 & 0.7620 \\ 0.7847 & 0.8793 & 0.9013 & 0.8848 \\ 0.5815 & 0.7353 & 0.7399 & 0.7365 \end{bmatrix}$$

图 1 对比了四种相似度图像。由图 1(a)(b)(c) 很容易看出，随着系数 α 变化，变异系数相似度（V-similarity）比 Jcaard 相似度（J-similarity）和 Dice 相似度（D-similarity）要好；由图 1(c)(d) 看出，余弦相似度（C-similarity）比其他三种相似度更好，但是在紧急群决策中它也存在一些问题，在下一部分将要介绍 [9]。

三、紧急决策的应用

我们利用文献 [10] 中的例子，评估了湖南省部分县市的紧急管理能力。邀请了六个专家评估湘潭、娄底和郴州三个城市大雪灾的紧急管理能力，考虑了多个因素，汇总后得到六个一级指标： A_1 为投资总额、 A_2 代

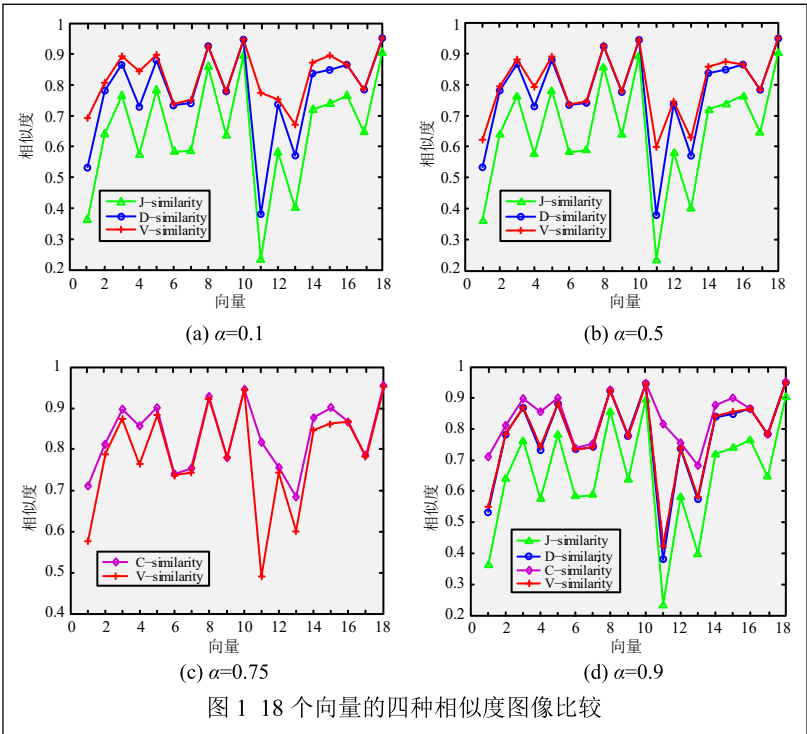


图 1 18 个向量的四种相似度图像比较

$$D_{Ld} = \begin{bmatrix} 0.5878 & 0.7404 & 0.7533 & 0.7436 \\ 0.8593 & 0.9243 & 0.9270 & 0.9250 \\ 0.6361 & 0.7776 & 0.7795 & 0.7781 \\ 0.8956 & 0.9449 & 0.9467 & 0.9454 \\ 0.2337 & 0.3789 & 0.8172 & 0.4884 \\ 0.5833 & 0.7368 & 0.7546 & 0.7413 \end{bmatrix}$$

$$D_{Cz} = \begin{bmatrix} 0.3996 & 0.5710 & 0.6830 & 0.5990 \\ 0.7202 & 0.8373 & 0.8777 & 0.8474 \\ 0.7392 & 0.8500 & 0.9016 & 0.8629 \\ 0.7642 & 0.8664 & 0.8665 & 0.8664 \\ 0.6449 & 0.7841 & 0.7841 & 0.7841 \\ 0.9057 & 0.9505 & 0.9527 & 0.9511 \end{bmatrix}$$

$$W=(0.162, 0.198, 0.201, 0.157, 0.174, 0.108)$$

计算权重相似度, 得到:

$$W_s=WD \quad (13)$$

步骤 4: 用式 (13) 得到 4 种相似度的决策数值, 列在表 2 中。为了作比较, 系数为 $\alpha=0.25$ 时 4 种相似度列在表 3 中。

表 2 当 $\alpha=0.75$ 时 4 种相似度的决策数值

	W_{J-s}	W_{C-s}	W_{D-s}	W_{I-s}
湘潭 Xt	0.6290	0.7622	0.8273	0.7785
娄底 Ld	0.6375	0.7531	0.8346	0.7795
郴州 Cz	0.6859	0.8043	0.8410	0.8135
排序	Cz>Ld>Xt	Cz>Xt>Ld	Cz>Ld>Xt	Cz>Ld>Xt

表 3 当 $\alpha=0.25$ 时, 4 个相似度的决策数值

	W_{J-s}	W_{C-s}	W_{D-s}	W_{I-s}
湘潭 Xt	0.6290	0.7622	0.8273	0.8111
娄底 Ld	0.6375	0.7531	0.8346	0.8142
郴州 Cz	0.6859	0.8043	0.8410	0.8318
排序	Cz>Ld>Xt	Cz>Xt>Ld	Cz>Ld>Xt	Cz>Ld>Xt

从表 2 和表 3 中, 可以看出来四个权重相似度 W_{J-s} 、 W_{C-s} 、 W_{D-s} 和 W_{I-s} 的排列顺序, 紧急管理能力最佳的地方是郴州^[16]。

四、结束语

本文提出了一种变异系数相似度并且证明了它的

一些特点, 并将变异系数相似度和现存的相似度做了比较, 显示了此方法具有更好的数值。这四个权重相似度被用来解决紧急群决策问题。最终, 给出一个紧急能力评估的实际例子评估中国湖南三个城市的相似度能力。不同的相似度决策数值说明了可以得到三个城市的排序, 并且紧急管理能力最佳的城市可以很容易得到, 显示了本文提出的方法是合理和有效的。

参考文献

- [1] 王坚强, 聂荣荣. 基于直觉梯形模糊信息的多准则群决策方法 [J]. 系统工程理论与实践, 2012, 32(08): 1747-1753.
- [2] 李海林. 时间序列数据挖掘中的特征表示与相似性度量方法研究 [D]. 大连: 大连理工大学, 2012.
- [3] J. Ye. Cosine Similarity Measures for Intuitionistic Fuzzy Sets and their Applications[J]. Mathematical and Computer Modeling, 2011, 53(1-2): 91-97.
- [4] Soojung Lee. Improving Jaccard Index for Measuring Similarity in Collaborative Filtering[M]. Springer Singapore: 2017-06-15.
- [5] Dice, Lee R. (1945). "Measures of the Amount of Ecologic Association Between Species". Ecology. 26 (3): 297-302.
- [6] 陈大力, 沈岩涛, 谢槟竹, 马颖异. 基于余弦相似度模型的最佳教练遴选算法 [J]. 东北大学学报 (自然科学版), 2014, 35(12): 1697-1700.
- [7] Y. Lean, K. K. Lai. A Distance-based Group Decision-making Methodology for Multi-person Multi-criteria Emergency Decision Support[J]. Decision Support Systems 2011, 51(2): 307-315.
- [8] 谢建明, 刘三阳. 区间直觉模糊集的相似性测度及其应用 [J]. 数学的实践与认识, 2018, 48(4): 249-255.
- [9] 李宝萍, 陈华友. 直觉模糊面集的相似度及其在模式识别中的应用 [J]. 模糊系统与数学, 2015, 29(1): 134-141.
- [10] 徐选华, 李芳. 重大雪灾应急管理能力评价——以湖南省为例 [J]. 灾害学, 2011, 26(2): 130-137.
- [11] 贾凡. 不确定信息环境下多准则群决策方法研究 [D]. 济南: 山东大学, 2017.
- [12] 吴胜, 李延来, 陈振颂. 群决策方法及其在供应商选择中的应用 [J]. 计算机仿真, 2018, 35(3): 184-189.
- [13] 王瑞, 贾瑞玉. 基于形态模式的时间序列相似性度量算法 [J]. 计算机应用与软件, 2017, 34(9): 253-256+305.
- [14] 黄冬梅, 郑霞, 赵丹枫, 王丽琳. 一种多元台风时间序

列的相似性度量方法[J/OL]. 计算机应用研究, 2019(03): 1-9[2018-07-05].<http://kns.cnki.net/kcms/detail/51.1196.TP.20180209.1114.014.html>.

[15] 徐小连. 地质灾害的政府应急管理研究[D]. 昆明: 云南财经大学, 2017.

[16] 孙宁宁. 城市突发公共事件应急能力评价指标体系研究[D]. 大连: 大连理工大学, 2009.

Application of Variation Coefficient Similarity in Emergency Group Decision Making

FENG Chang-min

(School of Applied Science, Beijing Information Science and Technology University, Beijing 100192, China)

Abstract: Considering the problem of emergency group decision making, a new similarity- variation coefficient similarity is defined based on Jaccard, Dice and cosine similarity. The features of four kinds of similarity are introduced, and then the variation coefficient similarity is applied to the emergency group decision making.

According to the similarity of each scheme, a ranking of all the schemes can be obtained, and then the optimal scheme can be found. Finally, an example is given to evaluate the emergency management ability of major snow and ice disasters in Hunan province of China. As a comparison, the decision results of different similarity measures are given. It is shown that the similarity defined in this paper is better in recognition and applicable.

Key words: similarity; emergency decision making; variation coefficient; multi-criteria group decision-making

作者简介

冯长敏: 北京信息科技大学理学院, 硕士研究生, 研究方向为优化与控制决策。

通讯地址: 北京市海淀区西三旗街道北京信息科技大学小营校区(小营东路12号)学一公寓

邮编: 100192

邮箱: changmin114@163.com