

# Research on Data Fusion Algorithm Based on Fuzzy Mathematics and Principal Component Analysis

Wanxia Xu, Bo Wang, He Zhang

School of Mathematics and Physics, China University of Geosciences, Wuhan Hubei  
Email: 1911680348@qq.com

Received: May 2<sup>nd</sup>, 2019; accepted: May 17<sup>th</sup>, 2019; published: May 24<sup>th</sup>, 2019

---

## Abstract

In this paper, we use the principal component analysis in the multivariate statistical method, count the correlation coefficient between the measured value and the principal component, the comprehensive support degree is obtained, the lower comprehensive support level data are excluded. We acquired valid data. For valid data, using fuzzy correlation characteristics of multi-sensor data, a method of data fusion based on fuzzy closeness is proposed. The calculation process of the data fusion method is relatively fixed and the calculation amount is small, which is easy to be realized by computer in real time.

## Keywords

Principal Component Analysis, Fuzzy Similarity, Data Fusion

---

# 基于模糊数学和主成分分析的数据融合算法研究

徐晚霞, 王 博, 张 贺

中国地质大学(武汉)数学与物理学院, 湖北 武汉  
Email: 1911680348@qq.com

收稿日期: 2019年5月2日; 录用日期: 2019年5月17日; 发布日期: 2019年5月24日

---

## 摘 要

本文利用多元统计方法中的主成分分析法计算测量值与主成分的相关系数, 得出各传感器的综合支持程

度,剔除综合支持程度较低的数据,取得有效数据。对有效数据,利用多传感器数据模糊相关特性,提出了一种基于模糊贴近度的数据融合方法。该数据融合方法计算过程相对固定,计算量小,便于计算机实时实现。

## 关键词

主成分分析法,模糊贴近度,数据融合

Copyright © 2019 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

数据融合[1]技术是一种多传感器信息数据处理的技术,它利用计算机技术,对来自各个传感器的信息数据在一定的的方法下,加以分析、提取、综合,得到被测对象更加全面、可靠、准确的结果。

多传感器数据融合可以增加测量的可信度,改善系统的可靠度,在工业和军事等方面得到了广泛的应用。本文利用多元统计方法中的主成分分析法计算测量值与主成分的相关系数,得出各传感器的综合支持程度,剔除综合支持程度较低的数据,对于剔除后的有效数据,首先将测量与估计值都进行模糊化,然后计算测量值与估计值之间的模糊贴近度,利用贴近度来描述各个传感器的重要度,确定各个传感器在融合算法中的权重。最后利用融合公式求出融合结果。

## 2. 主成分分析法

主成分分析法[2]是将研究对象的多个相关变量(指标)化为少数几个不相关的变量的一种多元统计方法。它作为一种数据分析技术,能对原有数据集进行简化处理,有效找到其中最主要的元素和组成结构,同时剔除信息中的噪音,降低数据冗余性,还能将原有的复杂数据进行降维处理,使得复杂数据结构能够简单化。主成分分析[3]的主要内容是把多个变量综合成一个或者少数几个综合指标,通过使用少数的综合指标来代表尽可能多的变量所包含的信息。

定义1: 设总体  $X = (x_1, x_2, \dots, x_n)$  的协方差阵为  $\Sigma$ , 其特征根为  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , 对应的单位化正交特征向量为  $e_1, e_2, \dots, e_n$ , 其中  $e_i = (e_{i1}, e_{i2}, \dots, e_{in})$ , 则第  $i$  个主成分为:

$$y_i = e_i X = e_{i1}x_1 + e_{i2}x_2 + \dots + e_{in}x_n \quad (1)$$

定义1中,当总体的协方差未知时,可以用样本的协方差矩阵或相关矩阵代替。

定义2: 在主成分分析中,定义主成分  $y_i$  的方差贡献率

$$a_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad i=1, 2, \dots, n \quad (2)$$

来说明各主成分  $y_i$  概括原变量  $X$  信息的大小。称  $\sum_{j=1}^k a_i$  为前  $k$  个主成分的累计方差贡献率。一般情况

下取  $k$  使得累计方差贡献率  $\sum_{j=1}^k a_i$  达到 85% 以上。

定理 1: 总体  $X$  的原变量  $x_i$  与主成分  $y_i$  的相关系数为

$$\rho(x_i, y_i) = \frac{e_{ij} \sqrt{\lambda_j}}{\sigma_i} \quad (3)$$

其中,  $\sigma_i$  为  $x_i$  的标准差,  $e_{ij}$  为特征向量  $e_i$  的第  $j$  个分量。

定理 2: 设  $(y_1, y_2, \dots, y_k)$  为总体  $X$  的前  $k$  个主成分, 则  $x_j$  与  $(y_1, y_2, \dots, y_k)$  的复相关系数的平方:

$$\tau_j = \rho^2[x_j, (y_1, y_2, \dots, y_k)] = \sum_{i=1}^k \rho^2(y_i, x_j) \quad (4)$$

因为  $\tau_j$  表示  $k$  个主成分  $(y_1, y_2, \dots, y_k)$  对原变量  $x_j$  的贡献率, 而这  $k$  个主成分  $(y_1, y_2, \dots, y_k)$  概括了原变量  $X$  的 85% 以上的信息, 所以在某种程度上,  $\tau_j$  可以看成是在总体  $X$  信息中所在的比重, 可以利用  $\tau_j$  来定义各传感器的综合支持程度。

定义 3: 定义第  $i$  个传感器的综合支持程度:

$$z_i = \frac{\tau_i}{\sum_{i=1}^n \tau_i} \quad (5)$$

累计综合支持程度越高, 则所包含的信息比重越大, 一般情况下, 取累计综合支持程度达 85% 以上的传感器。

### 3. 模糊数学

基于模糊贴近度的数据融合算法[4]先将传感器量测值与目标估计值进行模糊化, 再定义与计算每个传感器的量测值模糊量与目标估计值模糊量之间的贴近度。由贴近度的大小来表示传感器的重要性, 贴近度越大表示该传感器对数据融合结果的影响越重要, 在融合过程中其权重也越大。最后提出数据融合公式实现多传感器数据融合。

对于第  $i$  个传感器, 设其对真实值  $A$  的  $m$  次测量后的测量均值为  $x_i$ , 测量方差为  $\sigma_i$ 。则量测值的模糊量表示为

$$\tilde{A}_i = (a_{i1}, a_{i2}, a_{i3}) = (x_i - 2\sigma_i, x_i, x_i + 2\sigma_i) \quad (6)$$

定义  $\tilde{A}_i$  与  $\tilde{A}_j$  的贴近度为

$$S(\tilde{A}_i, \tilde{A}_j) = \frac{1}{1 + d(\tilde{A}_i, \tilde{A}_j)} \quad (7)$$

其中

$$d(\tilde{A}_i, \tilde{A}_j) = \left| \frac{a_{i1} + 4a_{i2} + a_{i3} - a_{j1} - 4a_{j2} - a_{j3}}{6} \right|$$

第  $i$  个传感器权重  $r_i$  为:

$$r_i = S(\tilde{A}_i, \tilde{A}_0) \quad (8)$$

将它们归一化后可以得到各个传感器之间的相对权重为:

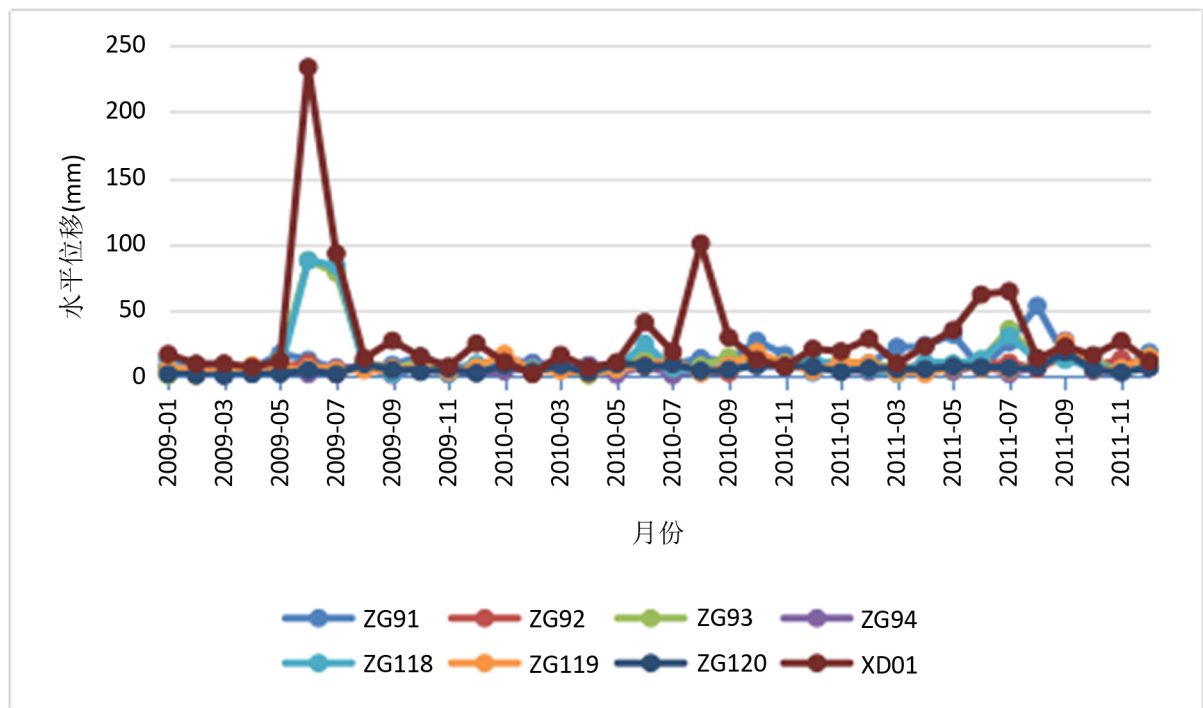
$$w_i = \frac{r_i}{\sum_{i=1}^n r_i} \quad (9)$$

则融合结果为:

$$\bar{x} = \sum_{i=1}^n w_i x_i$$

#### 4. 应用实例

秭归县白水河滑坡 8 个传感器在 2009 年~2011 年期间, 水平位移监测值见图 1。ZG91, ZG92, ZG93, ZG94, ZG118, ZG119, ZG120, XD01, 8 个传感器在秭归县白水河测得 2009 年 1 月到 2011 年 12 月的水平滑坡位移监测值[5], 通过数据融合的到最后融合数据, 以此判断秭归县白水河的滑坡情况。



**Figure 1.** Displacement monitoring curve of baishuihe landslide  
**图 1.** 白水河滑坡水平位移监测曲线

对于秭归县白水河滑坡 8 个传感器在 2009 年~2011 年每月的水平位移监测值, 运用主成分分析法可知, 上面 8 个传感器的综合支持度分别为 0.024, 0.022, 0.047, 0.058, 0.064, 0.118, 0.331, 0.358, 则根据累计综合支持度大于 85%, ZG91、ZG92、ZG93、ZG94 的综合支持度不够, 融合时删其读数, 取其 余传感器的测量数据用模糊贴近度的数据融合算法进行融合。传感器个数  $n = 4$ , 测量次数  $m = 36$ 。模糊数学和主成分与模糊数学融合所得的结果见表 1。

**Table 1.** Compared the results of data fusion  
**表 1.** 数据融合结果比较

月份	模糊数学	主成分与模糊数学	月份	模糊数学	主成分与模糊数学	月份	模糊数学	主成分与模糊数学
2009-01	7.2254	6.9157	2010-01	9.8751	9.0554	2011-01	9.7454	9.2193
2009-02	4.0090	4.3692	2010-02	7.2483	7.7973	2011-02	7.8579	7.9652
2009-03	3.2495	3.0972	2010-03	7.7743	7.2560	2011-03	8.1523	9.8117

## Continued

2009-04	5.5209	4.1866	2010-04	4.3222	6.3234	2011-04	12.2241	12.1453
2009-05	9.5001	10.2954	2010-05	6.7060	5.2967	2011-05	12.9376	11.8854
2009-06	55.7504	28.2653	2010-06	16.4983	10.6761	2011-06	11.2149	9.4339
2009-07	48.0232	22.5999	2010-07	7.3084	5.4740	2011-07	25.9504	18.9167
2009-08	8.6906	9.2618	2010-08	8.5564	7.0791	2011-08	20.3569	19.8556
2009-09	5.5021	5.9267	2010-09	10.2850	8.6360	2011-09	15.7700	17.4668
2009-10	9.9850	10.0486	2010-10	13.8203	14.8225	2011-10	7.7023	7.2570
2009-11	3.7874	3.5872	2010-11	10.4971	10.8352	2011-11	6.1310	7.9647
2009-12	8.1161	6.7388	2010-12	8.9062	8.0677	2011-12	12.1885	11.8192

由表 1 可以看到, 本文提出的主成分与模糊数学方法与模糊数学融合结果相比, 在计算过程中, 由于用主成分减少了 4 个综合支持度不高的数据, 在后面用模糊数学融合时, 计算量少了很多。

## 5. 总结

多传感器数据融合综合利用多种传感器的信息以获得有关目标的属性特征, 可以避免单一传感器的局限性限制, 减少各传感器不确定误差的影响。本文提出了一种基于模糊理论与主成分分析方法集成的多传感器数据融合算法, 该方法计算过程相对固定, 便于计算机实时实现。

## 参考文献

- [1] 严怀成, 黄心汉, 王敏. 多传感器数据融合技术及其应用[J]. 传感器技术, 2005, 24(10): 1-4.
- [2] 董九英. 多传感器数据融合的主成分方法研究[J]. 计算机工程与应用, 2009, 45(33): 111-113.
- [3] 向东进, 李宏伟, 刘小雅. 实用多元统计分析[M]. 武汉: 中国地质大学出版社, 2005.
- [4] 向新建. 基于模糊数学与统计理论集成的多传感器数据融合方法[J]. 传感技术学报, 2004, 17(2): 197-199.
- [5] 张俊, 殷坤龙, 王佳佳, 等. 基于时间序列与 PSO-SVR 耦合模型的白水河滑坡位移预测研究[J]. 岩石力学与工程学报, 2015, 34(2): 382-391.

### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7991, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [aam@hanspub.org](mailto:aam@hanspub.org)