

文章编号: 2096-1618(2020)04-0424-06

基于粒子群 K 均值聚类算法的多时相去云处理

许晨¹, 康雪¹, 杨玲²

(1. 成都市气象局, 四川 成都 610000; 2. 成都信息工程大学, 四川 成都 610225)

摘要: 如何有效地减少或去除云的影响, 不仅可以提高遥感数据的利用率, 同时也是遥感数据进行准确解译的重要途径。将粒子群算法和 K 均值算法有效结合, 提出一种改进的多时相去云方法。该算法有效克服了 K 均值算法容易陷入局部极值的缺点, 结合粒子群算法反复迭代寻求全局最优解。实验结果表明, 该算法收敛速度快, 效果具有全局优化的特征, 为研究去云算法提供一种简单有效的方法。

关键词: 去云; 粒子群; K 均值聚类

中图分类号: TP311

文献标志码: A

doi: 10. 16836/j. cnki. jcuit. 2020. 04. 010

0 引言

去云处理是遥感数据融合及图像改善技术中的一个重要方面。遥感数据里常出现范围较大的薄云, 所以通常说的去云是指去除薄云。薄云遮挡的地方不仅包含云信息, 还包含下垫面信息^[1]。根据薄云的成像机理, 云信息主要处于数据的低频部分, 可以通过处理数据的低频达到去云目的。通过结合粒子群优化算法和 K 均值聚类算法, 将两种算法优势互补结合起来检测云区, 并通过多时相融合进行云区替换, 实现去云。

1 资料来源与方法

1.1 资料来源

选取 NOAA-17 卫星的 1、4、5 通道, 如表 1 所示。其中, 通道 1 获得白天云图及地表数据; 通道 4、5 获得昼夜云图、海表、地表温度和土壤湿度^[2]。

表 1 NOAA-17 卫星 1、4、5 通道的波长范围

通道	波长/ μm	波段
Band 1	0.58 ~ 0.68	可见光
Band 4	10.3 ~ 11.3	远红外
Band 5	11.5 ~ 12.5	远红外

1.2 粒子群和 K 均值聚类算法

1.2.1 粒子群算法

粒子群算法从随机解出发, 用迭代法求最优解, 并通过适应度评价解的品质^[2]。其优点在于: 同时具有全局寻优能力和局部寻优能力, 没有个体“交叉”、“变异”

等复杂运算操作, 参数调整较为简单, 且比遗传算法更快收敛于最优解, 可避免完全随机寻优退化现象^[3]。

算法首先对粒子群进行初始化, 随机设定各粒子的初始位置、初始速度。然后计算各粒子的适应度值, 通过迭代优化搜索当前空间中的最优解^[4]。在每一次迭代过程中, 粒子先找到两个极值^[5]: 一是粒子自身搜索到的最优解, 称之为个体极值 (Pbest); 另一个是在整个种群找到的最优解, 称之为全局极值 (Gbest)。然后根据式 (1) 和 (2) 调整自己的速度和位置。

式中, 用维向量表示粒子信息, $v[]$ 表示粒子速度, 粒子具有全局搜索能力; $present[]$ 表示当前位置, 具有平衡全局搜索和局部搜索的作用; $rand()$ 是介于 (0, 1) 的随机数, 用于增加搜索的随机性; $c1$ 、 $c2$ 是加速因子, 通常 $c1=c2=2$; w 为惯性因子, 取值为非负数, 用于调节对解空间的搜索范围, 当 w 取值较大时, 适于对解空间进行大范围搜索, 反之亦然。

$$v[] = w \cdot v[] + c1 \cdot rand() \cdot (Pbest[] - present[]) + c2 \cdot rand() \cdot (Gbest[] - present[]) \quad (1)$$

$$present[] = present[] + v[] \quad (2)$$

其中, V_{\max} 表示每一维粒子的最大速度, 初始值由用户设定, 随着某一维的变化随时更新为维度最大速度。

1.2.2 K 均值聚类算法

K 均值聚类是解决聚类分析问题的一种经典算法, 归类原则是使相似的样本尽可能划分为一类, 不相似的样本尽可能归为不同类^[6]。该算法是一种动态聚类算法, 具有简单、快速并且能有效处理大数据库的优点。

算法首先取 n 个初始聚类中心, 计算每个样品到初始聚类中心的距离, 依据最近距离原则, 将样品归入不同初始聚类中心的类中, 并将聚类中心调整为该类所有样品的均值, 再次计算各样品到 n 个中心的距离, 依据同样原则重新归类、修改聚类中心点, 重复此过程

直至聚类中心位置不再发生变化时结束,使聚类域中所有样品到聚类中心距离的平方和最小^[6]。

该算法受到聚类中心个数 n 和初始聚类中心位置的选择,也会受到样品排列次序和几何性质的影响,若样品几何特性表明它们能形成几个相距较远的小块孤立区域,则算法多能收敛。

1.3 算法实现

以上两种算法各有优势,但也存在缺点,K 均值聚类容易陷入局部极值,即找到的局部极小值和全局最优值之间偏差较大。而粒子群优化算法则可以克服这一缺陷,通过反复迭代寻求全局最优解。所以将两种算法互补结合检测云区,并通过多时相融合进行云区替换,实现去云。

1.3.1 算法参数

一是反射率阈值。相对不同下垫面而言,不同的云在可见光和近红外波段均具有较高反射率,这有利于判云^[7]。分析 NOAA 可见光通道数据,结果显示:因下垫面介质不同,云的反射率不同;相同下垫面情况下,受高云或对流云顶影响,云的反射率也不相同。据相关资料分析^[8-9],可找出 NOAA 卫星在可见光通道绿色植被的最小反射率和云区最高反射率之间的相对关系和特征,以此确定云区反射率阈值,如表 2 所示。

表 2 NOAA 卫星在可见光波段的反射率判云阈值表 单位: %		
绿色植被最小反射率	云区最高反射率	云区反射率阈值
min<3	max<65	15
	65<max ≤90	16
	max>90	18
3 ≤min<6	max ≤45	15
	45<max ≤90	16
	max>90	20
min ≥6	max ≤70	23
	max>70	24

二是亮温阈值。在热红外波段,云的亮温灵敏度很高。据相关资料:对 NOAA-11、NOAA-12 卫星在不同季节的 12 个时相数据资料进行分析^[10-11],发现以下三点结论:其一,云在通道 4、5 具有相近的规律;再者,热红外温度阈值随着季节的变化而变化。在中国南方地区,春秋季节时,云的热红外温度阈值偏低,一般 5℃ 以下为云;夏季温度阈值较高,一般 10℃ 以下为云;冬季温度阈值较低,一般 0℃ 以下为云。最后,在相同季节时,不同地区的判云温度阈值也有所不同,如中国北方地区的温度阈值低于南方地区。综上,对多个云区时相资料的检测结果显示,一般选取比当地相应季节的地面或海面最低气温低 1℃~2℃ 作为经验阈值,可取得较好的去云效果。

1.3.2 算法实现

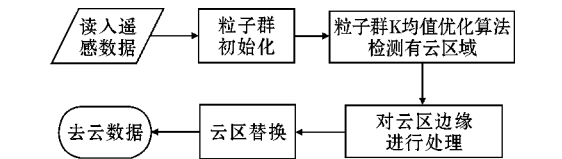


图 1 算法实现框图

初始化粒子群,通过 ENVI 软件读取 HDF 格式的 NOAA-17 遥感数据,显示通道 1、4、5。初始化聚类,设定聚类中心的灰度值在反射率灰度阈值和亮温阈值范围内,此时要针对不同季节和不同地区的地形特征进行具体分析,这样能使聚类效果更加准确,聚类速度更快。并将设定的聚类中心灰度值作为初始粒子的位置编码,计算粒子适应度、初始化粒子速度。如此重复 N 次,生成 N 个初始粒子群^[4],如图 2 所示。

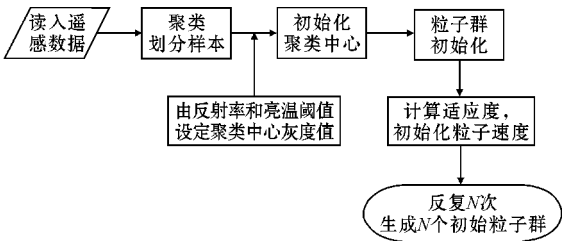


图 2 初始化粒子群实现框图

- 粒子群 K 均值算法检测云区算法步骤如下:
- A. 对单个粒子,比较其适应度值和它经历过最优位置的适应度值,取二者间的较小值更新粒子最优位置。
 - B. 对所有粒子,比较其最优位置的适应度值,取最小值更新群体所经历的最优位置。
 - C. 根据式(1)和式(2)调整粒子速度和位置。
 - D. 将更新后的粒子进行 K 均值优化:首先依据最近邻原则,参考粒子的聚类中心编码,确定对应粒子的聚类划分。其次按照聚类划分计算新的聚类中心,更新粒子的适应度值,取代原编码值。若达到结束条件,则结束;否则转步骤 A 继续迭代。

设定两个结束条件:满足迭代次数和找到最优位置,同时满足这两个条件时迭代结束。这样可避免出现以下情况:(1). 如果设定的固定阈值仅经过几次迭代就达到最小解,但此最小解仍然大于初始阈值,也必须重复初始给定的迭代次数才能结束,重复运算降低了算法效率;(2) 如果设定重复迭代次数,在已经重复了初始给定的迭代次数情况下,还未找到最优解,不能使算法最优^[4]。

其中,步骤 D 在对更新后粒子重新划分类别时可能会出现空聚类。这时,随机地从某非空聚类中取距离聚类中心最远的模式向量,将该向量放入空聚类。重复此过程,直到划分中没有空聚类为止。

由于 K 均值具有较强的局部搜索能力,引入 K 均值优化后的粒子群优化算法大大提高了收敛速度。

图 3 介绍粒子群更新过程中的迭代优化。

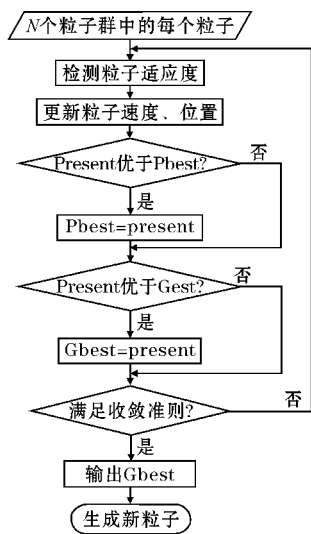


图 3 更新粒子群流程图

互补替换融合技术去云,互补替换融合技术是将多光谱、多空间的遥感数据进行整合,通过替代融合得到一个多层面的、在时间和空间上连续的遥感数据。互补替换去云通过对比不同时间获得的同一地区的遥感数据,对云雾覆盖地区进行替代填充,实现遥感数据去云,从而提高其分辨率和利用率。该方法主要分两个步骤:首先对不同时间获得的云区数据进行配准,然后对云区数据互补替换,进行辐射校正,消除替换过程中出现的接缝,保证数据的连续性^[12]。

选取 NOAA-17 卫星于相邻两天在相同时间对同一地区获取的数据进行处理,选取云区面积相对较小的遥感数据作为基准数据,另一组作辅助数据,这样可使去云后的图像减小信息损耗和去云痕迹,流程如图 4 所示。

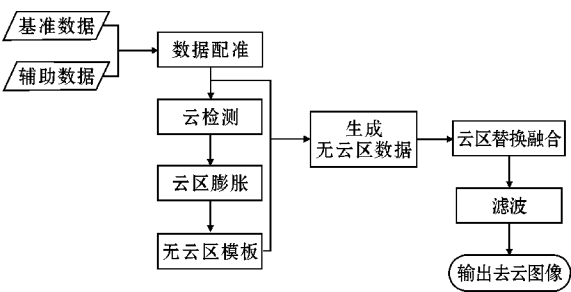
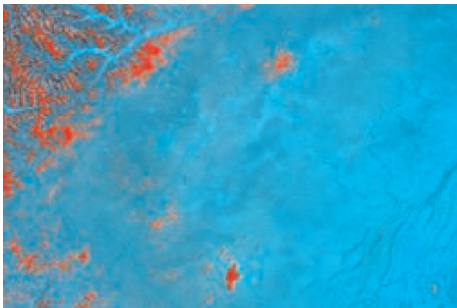


图 4 多时相融合去云流程图

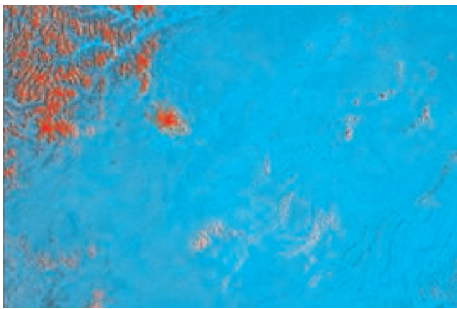
2 结果与分析

2.1 读取数据

通过 ENVI 软件读取 NOAA-17 卫星 8 月 17 日、18 日在四川盆地的两组数据,并合成 1、4、5 通道数据,如图 5 所示。



(a) 8 月 17 日数据



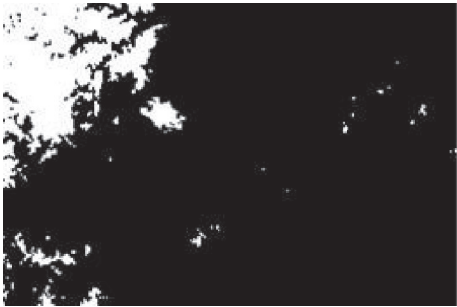
(a) 8 月 18 日数据

图 5 NOAA-17 的 1、4、5 通道合成图

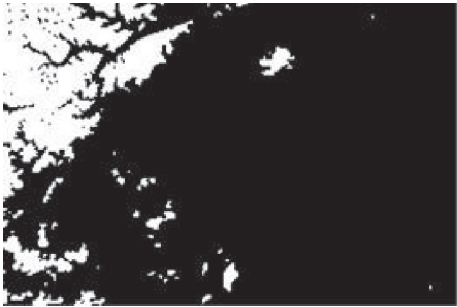
由图 5 分析,亮色区域可能为云与雪山。因云反射可见光,反照率较强,且温度较低,与其他地区有明显差异;而雪山地区温度较低,其亮温层信息较明显。

2.2 数据聚类

采用 K 均值聚类 and 粒子群 K 均值聚类,比较二者,如图 6、图 7 所示。



(a) 8 月 17 日数据

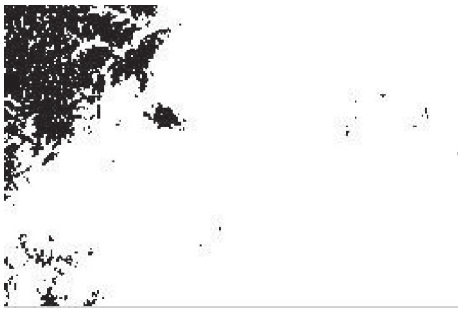


(b) 8 月 18 日数据

图 6 K 均值聚类结果



(a) 8 日 17 日数据



(b) 8 月 18 日数据

图 7 粒子群 K 均值聚类结果

对比两图,发现通过粒子群 K 均值聚类后的图 7 结果更好,其细节更多,聚类划分更准确。

选取图 7(b)作为基准数据,通过另外一个辅助数据 7(a)的替换融合实现去云。

2.3 云区过滤

如图 8,通过聚类算法选取两个数据中的有云区域,但选取的区域不仅是云区,可能还包含雪山等其他信息。分析地域条件,川西高原属高原气候,海拔 3000 ~ 4000 m 是寒温带地区,4000 ~ 5000 m 是亚寒带,5000 m 以上是永久冰雪带。由于冰雪的温度较低、反照率较强,其遥感数据特征和云的信息相似,所以图 7 选取的云区可能包含寒带地区中的雪山区域。为避免非云区域混合在选取的信息中,需对选取后的

信息过滤。
对辅助数据 7(a)进行膨胀,如图 8(a),其目的是为了强化雪山区域,可以发现:在膨胀后,雪山信息得到加强。将 8(a)与基准数据 7(b)相减,去除两组数据的公共区域,得到过滤后的云区。



(a) 膨胀后的图像



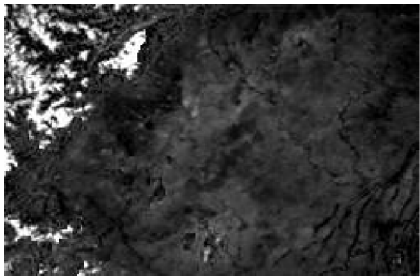
(b) 过滤后的云区

图 8 云区过滤

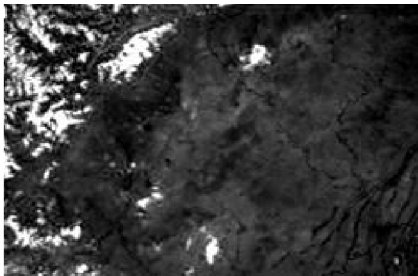
图 8(b)为过滤后的云区,在此区域中,只有少部分山区等其他信息混合于其中。因为大部分噪声信息已去除,剩下的残留噪声肉眼不容易发现,所以在此忽略。

2.4 替换融合

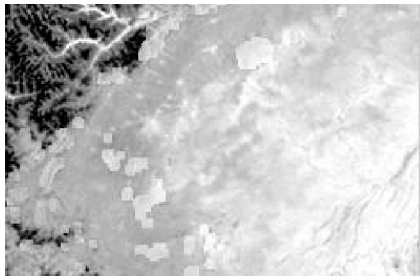
替换融合是遥感数据去云的最后阶段,传统的替换融合方法多对合成通道数据进行替换,但这样容易留下替换痕迹和修复边缘。所以分别对 3 个通道的灰度数据进行替换融合,如图 9 所示。



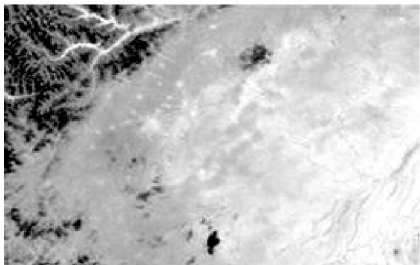
(a)通道1原始灰度图



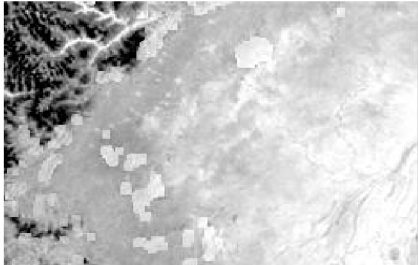
(b)通道1替换后图像



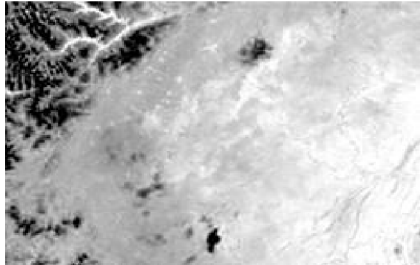
(c)通道4原始灰度图



(d)通道4替换后图像

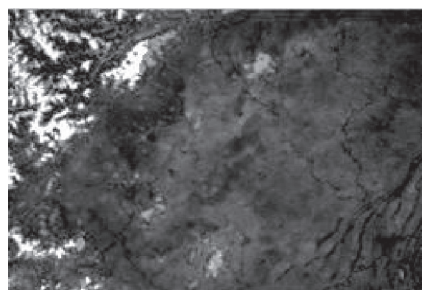


(e)通道5原始灰度图

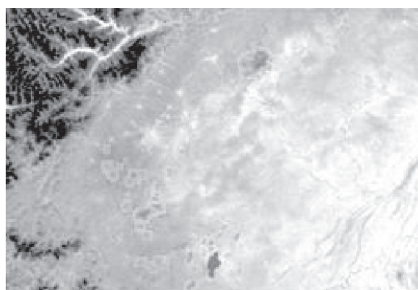


(f)通道5替换后图像

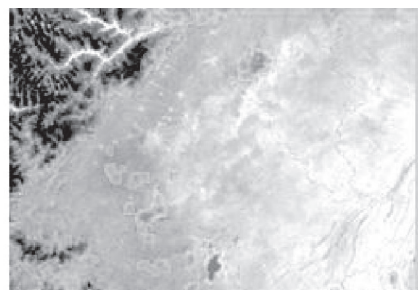
图 9 替换融合



(a)通道1替换后图像



(b)通道4替换后图像



(c)通道5替换后图像

图10 对云区周围区域进行50%的替换融合

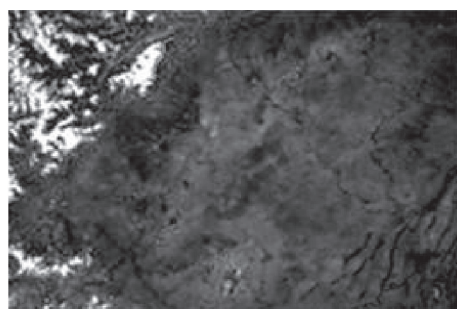
分析图9,替换后效果不太理想,仍包含部分云区。所以对数据进行二次融合,基于逐步替换的方法对原图像进行处理:即先进行一次过渡处理,对云区周围的区域进行50%的替换融合;再对所有云区进行替换融合。

如图10,对云区周围做50%替换融合的过渡效果较自然,较小的云区得到修复,较大的云区得到衰减。

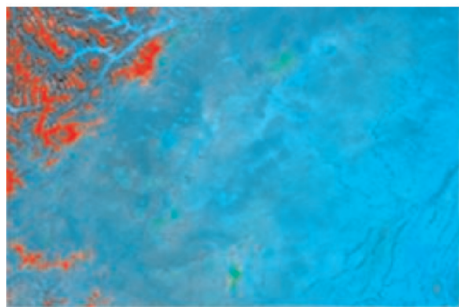
图11是在图10的基础上,对所有云区进行替换融合的结果。



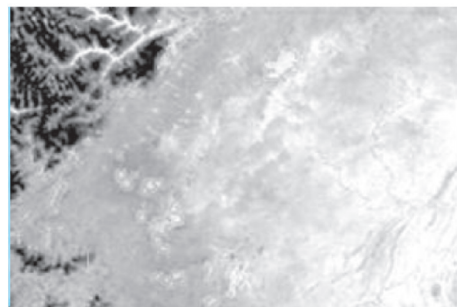
(a)通道1去云结果



(b)通道4去云结果



(c)通道5去云结果



(d)三通道合成的彩色去云结果

图11 二次融合结果

从图11中可以看到,原来存在的云区得到了较自然且准确的替换。

3 结论

提出了一种简单有效的多时相去云方法,结合粒子群和K均值聚类算法检测云区,该算法发挥了粒子群算法较强的全局寻优能力,有效克服了K均值算法陷入局部极值的缺点,取得较好的效果。算法主要针对薄云或薄雾影响下、云区小于50%的数据,对于云区较大数据,效果不太理想,对这类数据的处理必须结合其他的辅助信息进行。

参考文献:

- [1] 刘洋,白俊武. 遥感影像中薄云的去除方法研究[J]. 测绘与空间地理信息,2008,31(3):120-122.
- [2] H Li, Manjunath B S, Mitra S K. Multi-sensor image fusion using the wavelet transform[J]. Graphical Models and Image Processing, 1995, 57(3): 235-245.
- [3] 许永峰,姜振益. 一种基于粒子群优化的K-均值彩色图像量化算法[J]. 西北大学学报(自然科学版),2012,42(3):351-354.

- [4] 刘靖明、韩丽川、侯立文. 一种新的聚类算法——粒子群聚类算法[J]. 计算机工程与应用, 2005, 20(3): 183–185.
- [5] 王克华、牛慧、张亚南, 等. 一种参数自适应调整和边界约束的粒子群算法[J]. 电子设计工程, 2011, 19(21): 46–49.
- [6] Pohl C. Tools and method used in data fusion[J]. Future Trends in Remote Sensing, 1998, 21(5): 69–72.
- [7] Serra J. Image analysis and mathematical morphology[J]. Theoretical advances, 1988, 21(5): 69–72.
- [8] Nguyen Thanh Hoan, Ryutaro Tateishi. Cloud Removal of Optical Image Using Sar Data for Alos Applications[J]. Centre for Environmental Remote Sensing, 2008, 1(31): 263–265.
- [9] Tapasmini Sahoo, Suprava Patnaik. Cloud Removal from Satellite Images using Auto Associative Neural Network and Stationary Wavelet Transform [G]. First International Conference on Emerging Trends in Engineering and Technology, 2008: 125–129.
- [10] W G. Rees. Physical Principles of Remote Sensing [D]. UK: Cambridge University, 2001: 273–279.
- [11] Grodecki J, Dial G. Block adjustment of high-resolution satellite images described by rational polynomials. PE&RS, 2003, 3(1): 265–268.
- [12] 孙立明. 卫星及航拍图像的云雾噪声去除研究 [D]. 哈尔滨: 哈尔滨工程大学. 2006.

Multi-temporal Cloud Removal based on Particle Swarm K-means Clustering Algorithm

XU Chen¹, KANG Xue¹, YANG Ling²

(1. Chengdu Meteorological Office, Chengdu 610000 China; 2. Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: How to reduce or remove the influence of clouds effectively can not only improve the utilization of remote sensing data, but also an important way for interpretation of remote sensing data accurately. The paper considers the problem of clouds covering area acquired at different dates, by using an improved K-means clustering based on particle swarm optimization algorithm (K-M PSO) to remove or reduce cloud disturbance. This algorithm overcomes the shortcoming of traditional K-means clustering algorithm easily involving into the local optima, searches for optimization solutions based on the iterative particle swarm optimization method. The results shows that K-M PSO algorithm not only improves the convergence rate, but also optimized all area, it provides a simple and effective method to study the cloud removal algorithm.

Keywords: cloud removal algorithm; particle swarm; K-means clustering