

Srikanta Patnaik  
Xiaolong Li *Editors*

# Proceedings of International Conference on Soft Computing Techniques and Engineering Application

ICSCTEA 2013, September 25–27,  
2013, Kunming, China

# **Advances in Intelligent Systems and Computing**

Volume 250

*Series editor*

Janusz Kacprzyk, Warsaw, Poland

For further volumes:

<http://www.springer.com/series/11156>

### *About this Series*

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

### *Advisory Board*

#### Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India  
e-mail: [nikhil@isical.ac.in](mailto:nikhil@isical.ac.in)

#### Members

Emilio S. Corchado, University of Salamanca, Salamanca, Spain  
e-mail: [escorchado@usal.es](mailto:escorchado@usal.es)

Hani Hagrass, University of Essex, Colchester, UK  
e-mail: [hani@essex.ac.uk](mailto:hani@essex.ac.uk)

László T. Kóczy, Széchenyi István University, Győr, Hungary  
e-mail: [koczy@sze.hu](mailto:koczy@sze.hu)

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA  
e-mail: [vladik@utep.edu](mailto:vladik@utep.edu)

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan  
e-mail: [ctlin@mail.nctu.edu.tw](mailto:ctlin@mail.nctu.edu.tw)

Jie Lu, University of Technology, Sydney, Australia  
e-mail: [Jie.Lu@uts.edu.au](mailto:Jie.Lu@uts.edu.au)

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico  
e-mail: [epmelin@hafsamx.org](mailto:epmelin@hafsamx.org)

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil  
e-mail: [nadia@eng.uerj.br](mailto:nadia@eng.uerj.br)

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland  
e-mail: [Ngoc-Thanh.Nguyen@pwr.edu.pl](mailto:Ngoc-Thanh.Nguyen@pwr.edu.pl)

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong  
e-mail: [jwang@mae.cuhk.edu.hk](mailto:jwang@mae.cuhk.edu.hk)

Srikanta Patnaik · Xiaolong Li  
Editors

# Proceedings of International Conference on Soft Computing Techniques and Engineering Application

ICSCTEA 2013, September 25–27, 2013,  
Kunming, China

 Springer

*Editors*  
Srikanta Patnaik  
Department of Computer Science  
and Engineering  
SOA University  
Bhubaneswar, Orissa  
India

Xiaolong Li  
Electronics and Computer  
Engineering Technology  
Indiana State University  
Indiana, IN  
USA

ISSN 2194-5357  
ISBN 978-81-322-1694-0  
DOI 10.1007/978-81-322-1695-7  
Springer New Delhi Heidelberg New York Dordrecht London

ISSN 2194-5365 (electronic)  
ISBN 978-81-322-1695-7 (eBook)

Library of Congress Control Number: 2013956601

© Springer India 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

On behalf of the Programme Committee, we welcome you to *International Conference on Soft Computing Techniques and Engineering Application (SCTEA 2013)* held during September 25–27, 2013 in Kunming China. The main objective of SCTEA 2013 is to provide a platform for researchers, engineers, and academicians from all over the world to present their research results and recent developments in soft computing techniques and engineering application. This conference provides opportunities for them to exchange new ideas and application experiences face to face, to establish business or research relations, and to find global partners for future collaboration.

During the last decade, soft computing techniques have been recognized as one and most attractive alternatives to the standard, well-established “hard computing” paradigms. Traditional computing methods are often too cumbersome for the present day multiobjective problems, which requires a precisely stated analytical model and often consumes large computational time. Soft computing techniques, on the other hand, emphasize gains in understanding system behavior in exchange for unnecessary precision have proved to be important practical tools for many contemporary problems. There are various soft computing techniques available and popular for the engineering and other applications. They are artificial neural networks (NN), genetic algorithms (GA), fuzzy logic models (FLM), and particle swarm techniques. NNs and FLMs are universal approximators of any multivariate function because they can be used for modeling highly nonlinear, unknown, or partially known complex systems, plants, or processes. GA and particle swarm optimization techniques have emerged as potential and robust optimization tools in recent years.

*The International Conference on Soft Computing Techniques and Engineering Application (SCTEA 2013)* covers all areas of soft computing and its applications to engineering domain. We have received 232 numbers of papers through “Call for Paper,” out of which 67 numbers of papers were accepted for publication in the conference proceedings through double blind review process. The broad areas, which are covered in this Conference, are: theoretical computer science, computer systems organization, soft computing, computing methodologies, computer applications, and software engineering application.

We are sure this proceeding will contribute to the knowledge of soft computing and particularly its application to various engineering domains. We are sure the participants will share their innovations in this conference.

Srikanta Patnaik  
Xiaolong Li

# Organizing Committee

## General Chairs

Prof. Srikanta Patnaik, Computer Science and Engineering in SOA University, India

## General Co-Chair

Dr. Xiaolong Li, Indiana State University, USA

## Technical Program Committee

Prof. Ilona Weinreich, Koblenz University of Applied Sciences, Germany

Prof. Flaminia L. Luccio, Università Ca' Foscari Venezia, Italy

Dr. Rajarshi Gupta, University of Calcutta, India

Prof. Yi-Hsing Chang, Southern Taiwan University of Science and Technology, Taiwan

Prof. Min-Shiang Hwang, Asia University, Taiwan

Prof. Chien-Lung Hsu Chang-Gung University, Taiwan

Prof. Chien-Jen Wang, National University of Tainan, Taiwan

Prof. W. David Pan, University of Alabama in Huntsville, USA

Prof. Sotirios G. Ziavras, New Jersey Institute of Technology, USA

Dr. Afolabi Babajide Samuel, The Computer Professionals Registration Council of Nigeria, Nigeria

Prof. Mauricio Papa, The University of Tulsa, USA

Prof. Tianshi Lu, Wichita State University, USA

Prof. Rita Moura Fortes, Mackenzie Presbyterian University, Brazil



Dr. Cettina Santagati, University of Catania, Italy  
Prof. Michael McAleer, Erasmus University Rotterdam, Netherland  
Prof. Bo Yang, Shanghai Jiao Tong University, China  
Dr. Vimal S. Joshi, Gujarat University, India  
Dr. Michael Joseph, St. Joseph's College of Engineering and Technology, India  
Prof. Roberto Caldelli, Universita' degli Studi Firenze, Italy  
Dr. Akshi Kumar, Delhi Technological University, India  
Prof. Chi-Lun Liu, Kainan University, Taiwan  
Dr. Isabel Barbancho, Universidad de Malaga, Spain  
Dr. Sunil R. Das FIEEE, University of Ottawa, Canada  
Dr. Valentin Todorov, United Nations Industrial Development Organization, Austria  
Dr. Qi Wang, University of the West of Scotland, United Kingdom  
Dr. Izzat Al-Smadi, Yarmouk University, Jordan  
Dr. Muhammad Rais, King Saud University, Saudi Arabia  
Dr. Michael Joseph, Mother Teresa College of Engineering and Technology, India  
Dr. Shunan Wu, Dalian University of Technology, China  
Dr. Shan Chu, Stony Brook University, USA  
Dr. B. K. Tripathy, VIT University, India  
Dr. Carlo Trigona, University of Catania, Italy  
Dr. R. Shashi Kumar, Bangalore University, India  
Dr. Ki Young Kim, Samsung Advanced Institute of Technology, Korea

# Contents

<b>Research on Hierarchical Clustering Algorithm Based on Cluster Outline. . . . .</b>	<b>1</b>
Hai-Dong Meng, Jing-Pei Ren and Yu-Chen Song	
<b>Exploration of Fault Diagnosis Technology for Air Compressor Based on Internet of Things. . . . .</b>	<b>11</b>
Zheng Yue-zhai and Chen Xiao-ying	
<b>An Enhancement Algorithm Based on Fuzzy Sets Algorithm Using Computer Vision System for Chip Image Processing. . . . .</b>	<b>17</b>
Chengxiang Tan, Lina Yang and Xichun Li	
<b>Development of DDoS Attack Defense System Based on IKEv2 Protocol . . . . .</b>	<b>25</b>
Qing Tan and Xiaojing Yue	
<b>Power Data Network Dynamic Simulation Platform . . . . .</b>	<b>33</b>
Qian Guo, Xingchuan Bao and Gu Feng	
<b>An Emotional Model Based on Multiple Factors. . . . .</b>	<b>43</b>
Qiong Xiao, Gangyi Ding and Yongkang Liu	
<b>Multi-Source and Heterogeneous Knowledge Organization and Representation for Knowledge Fusion in Cloud Manufacturing . . . . .</b>	<b>55</b>
Jihong Liu, Wenting Xu and Hongfei Zhan	
<b>Research of Dynamics and Deploying Control Method on Tethered Satellite . . . . .</b>	<b>63</b>
Lei Gang, Xian Yong, Feng Jie and Wang Kui	

**Using Auto-Associative Neural Networks for Signal Recognition Technology on Sky Screen . . . . .** 71  
 Yan Lou, Zhipeng Ren, Yiwu Zhao and Yugui Song

**Research on the Hotspot Information Push System for the Online Journal Based on Open-Source Framework . . . . .** 81  
 Jiya Jiang, Tong Liu, Yanqing Shi and Changhua Lu

**Development of Control System of Wheel Type Backhoe Loader . . . .** 87  
 Luwei Yang

**Some Results on Fuzzy Weak Boolean Filters of Non-commutative Residuated Lattice. . . . .** 97  
 Wei Wang, Yang Xu, Dan Tong, Xiao-yan Cheng and Yong-fei Li

**CGPS: A Collaborative Game in Parking-Lot Search. . . . .** 105  
 Peng Li, Demin Li and Xiaolu Zhang

**A Fuzzy-Based Context-Aware Privacy Preserving Scheme for Mobile Computing Services . . . . .** 115  
 Eric Ke Wang and Yunming Ye

**Research and Application of Trust Management System. . . . .** 123  
 Fengyin Li and Peiyu Liu

**Ranaad-Xek: A Prototype Design of Traditional Thai Musical Instrument Application for Android Tablet PC . . . . .** 131  
 Kasikrit Damkliang, Chawee Kaeoaiad and Sulkiplee Chehmasong

**Identifying Accurate Refactoring Opportunities Using Metrics . . . . .** 141  
 Yixin Bian, Xiaohong Su and Peijun Ma

**Research on Neural Network Predictive Control of Induction Motor Servo System for Robot . . . . .** 147  
 Chaofa Yu, Zelong Zhou, Zhiyong Chen and Xiangyong Su

**Research on Scale-Out Workloads and Optimal Design of Multicore Processors . . . . .** 157  
 Qiong Wang, Li Shen and Zhiying Wang

**Study of Modified Montgomery’s Algorithm and Its Application to 1,024-bit RSA . . . . .** 167  
 Yulin Zhang and Xinggang Wang

<b>A MVS-Based Object Relational Model of the Internet of Things . . . .</b>	177
Huijuan Zhang and Ran Xu	
<b>Rateless Code-Based Unequal Loss Protection for Layer-Coded Media Delivery . . . . .</b>	185
Xuan Dong, Shaohu Lv, Hu Shen, Junquan Deng, Xiaodong Wang and Xingming Zhou	
<b>Construction of the Grade-3 System for GJB5000A-2008 . . . . .</b>	195
Yonggang Li, Jinbiao Zhou, Jianwei He, Xiangming Li and Libing Guo	
<b>Virtual Training System of Assembly and Disassembly Based on Petri Net . . . . .</b>	205
Xiaoqiang Yang, Jinhua Han and Yi Pan	
<b>The Design of Visual RBAC Model Based on UML and XACML Integrating . . . . .</b>	213
Baode Fan and Mengmeng Li	
<b>A New Approach to Reproduce Traffic Accident Based on the Data of Vehicle Video Recorders . . . . .</b>	223
Hong Li, Qing Kang and Jing He	
<b>Improved RNS Montgomery Modular Multiplication with Residue Recovery . . . . .</b>	233
Tao Wu, Shuguo Li and Litian Liu	
<b>Functionally Equivalent C Code Clone Refactoring by Combining Static Analysis with Dynamic Testing. . . . .</b>	247
Xiaohong Su, Fanlong Zhang, Xia Li, Peijun Ma and Tiantian Wang	
<b>Architecture Designing of Astronaut Onboard Training System Based on AR Technology. . . . .</b>	257
He Ning, Hou Quanchao and Hu Fuchao	
<b>Design and Implementation of Bibliometrics System Based on RIA. . .</b>	263
Geyang Han and Bing Sun	
<b>SAR Image Filtering Based on Quantum-Inspired Estimation of Speckle Variance. . . . .</b>	273
Xiaowei Fu, Li Chen, Jing Tian, Xin Xu and Yi Wang	

**Image Semantic Annotation Approach Based on the Feature Matching** . . . . . 281  
 Cong Jin and Jinglei Guo

**Research on Transmission and Transformation Land Reclamation Based on BP Neural Network** . . . . . 289  
 Xi Wu, Hai-Ting Ming, Xue-Huan Qin and Wen-Jing Zhu

**TTP-ACE: A Trusted Third Party for Auditing in Cloud Environment.** . . . . . 299  
 Songzhu Mei, Haihe Ba, Fang Tu, Jiangchun Ren and Zhiying Wang

**Pattern Recognition Based on the Nonparametric Kernel Regression Method in A-share Market.** . . . . . 309  
 Huaiyu Sun, Mi Zhu and Feng He

**The Research on the Detection and Defense Method of the Smurf-Type DDoS Attack.** . . . . . 315  
 Wantian Cao and Xingchuan Bao

**A Preliminary Analysis of Web Usage Behaviors from Web Access Log Files** . . . . . 325  
 Thakerng Wongsirichot, Sukgamon Sukpisit and Warakorn Hanghu

**Assessment of BER Performance of a Power Line Communication System in the Presence of Transformer and Performance Improvement Using Diversity Reception.** . . . . . 333  
 Munshi Mahbubur Rahman and S. P. Majumder

**A Multi-Constraint Anonymous Parameter Design Method Based on the Attribute Significance of Rough Set.** . . . . . 345  
 Taorong Qiu, Lu Liu, Wenying Duan, Xiaoming Bai and Zhongda Lin

**Design and Implementation of a Middleware for Service-Oriented Distributed Systems.** . . . . . 351  
 Hong Xie, Donglin Su, Yijia Pan and Zhongfu Xu

**Refactoring Structure Semantics Similar Clones Combining Standardization with Metrics.** . . . . . 361  
 Xia Li, Xiaohong Su, Peijun Ma and Tiantian Wang

**A Retail Outlet Classification Model Based on AdaBoost** . . . . . 369  
 Kai Liu, Bing Wang, Xinshi Lin, Yeyun Ma and Jianqiang Xing

**Extensions of Statecharts with Time of Transition, Time Delay of Message Transmitting, and Arrival Probability of Message. . . . .** 381  
 Junqiao Li, Jun Tang and Shuang Wan

**The Optimization of Hadoop Scheduling Algorithms on Distributed System for Processing Traffic Information . . . . .** 389  
 Weizhen Sun and Xiujin Wang

**Understanding the Capacity Scaling of Personal Communications Services . . . . .** 397  
 Zheng Wang

**VHDL Implementation of Complex Number Multiplier Using Vedic Mathematics . . . . .** 403  
 Laxman P. Thakare, A. Y. Deshmukh and Gopichand D. Khandale

**Key Security Technologies of Cloud Computing Platforms . . . . .** 411  
 Liang Junjie

**The Measurement and Analysis of Software Engineering Risk Based on Information Entropy. . . . .** 419  
 Ming Yang, Hongzhi Liao, Rong Jiang and Junhui Liu

**GRACE: A Gradient Distance-Based Peer-to-Peer Network Supporting Efficient Content-Based Retrieval. . . . .** 427  
 Jianming Lv, Can Yang and Kaidong Liang

**Image Denoising Using Discrete Orthonormal S-Transform . . . . .** 435  
 Feng-rong Sun, Paul Babyn, Yu-huan Luan, Shang-ling Song and Gui-hua Yao

**Analyzing Services Composition Using Petri Nets . . . . .** 443  
 Jiajun Xu and Shuzhen Yao

**FPGA-Based Image Processing for Seamless Tiled Display System . . .** 451  
 Mingyu Wang, Yan Han, Rui Wang, Xiaopeng Liu and Yuji Qian

**2D Simulation of Static Interface States in GaN HEMT with AlN/GaN Super-Lattice as Barrier Layer . . . . .** 457  
 Imtiaz Alamgir and Aminur Rahman

**Study on Model and Platform Architecture of Cloud Manufacturing for Aerospace Conglomerate . . . . .** 467  
 Jihong Liu, Hongfei Zhan and Wenting Xu

**Service Composition Algorithm for Vehicle Network Based on Multiple Ontology . . . . .** 475  
Yamei Xia and Chen Liu

**Design and Application of Virtual Laboratory for Photography . . . . .** 483  
Xuefei Shi

**The Research of Travel-Time Tomography Based on Forward Calculation and Inversion . . . . .** 491  
Yaping Li and Suping Yu

**Protein Secondary Structure Prediction Based on Improved C-SVM for Unbalanced Datasets . . . . .** 499  
Ao Pei

**An Algorithm for Speckle Noise Based on SVD and QSF . . . . .** 507  
Weizhou Zhao, Hui Zhang, Baozhen Yang and Huili Jing

**A Generation Model of Function Call Based on the Control Flow Graph . . . . .** 513  
Weizhen Sun and Xiangyan Du

**A Novel Community-Based Trust Model for P2P Networks . . . . .** 521  
Songxin Wang

**The Algorithm of Mining Frequent Itemsets Based on MapReduce . . . . .** 529  
Bo He

**Multi-Feature Metric-Guided Mesh Simplification . . . . .** 535  
Hailing Wang, Fu Qiao and Bo Zhou

**Research on Medical Image Fusion Algorithms Based on Nonsampled Contourlet . . . . .** 543  
Junyu Long, Hong Yu and Aiming Yu

**Influence of Previous Cueing Validity on Gaze-Evoked Attention Orienting . . . . .** 553  
Qian Qian, Yong Feng, Lin Shi and Feng Wang

**Application of RBF Neural Network in Intelligent Fault Diagnosis System . . . . .** 561  
Yingying Wang, Ming Chang, Hongwei Chen and Ming Qian Wang

<b>An Analysis of the Keys to the Executable Domain-Specific Model . . .</b>	<b>567</b>
Qing Duan, Junhui Liu, Zhihong Liang, Hongwei Kang and Xingping Sun	
<b>About the Editors . . . . .</b>	<b>575</b>
<b>Author Index . . . . .</b>	<b>577</b>



# Research on Hierarchical Clustering Algorithm Based on Cluster Outline

Hai-Dong Meng, Jing-Pei Ren and Yu-Chen Song

**Abstract** The traditional hierarchical methods always fail to take both the features of connectivity and proximity of clusters into consideration at the same time. This paper presents a hierarchical clustering algorithm based on cluster outline, which effectively addresses clusters of arbitrary shapes and sizes, and is relatively resistant to noise and easily detects outliers. The definition of the boundary point and cluster outline is firstly given, and the standard and approach of measuring similarity between clusters is then taken with the feature of connectivity and proximity of clusters. The experiments on the Iris and image data sets confirm the feasibility and validity of the algorithm.

**Keywords** Connectivity · Proximity · Cluster outline · Hierarchical clustering

## 1 Introduction

Hierarchical algorithm in cluster analysis as a main method in data mining, no matter what the field of research and application is, has been developed largely and promoted greatly. In recent years, there are various hierarchical clustering algorithm put forward, which are partitioned into two categories in terms of the similarity standard: one for connectivity-based similarity and the other for proximity-based similarity. The first one is free to find clusters of arbitrary shapes, but

---

H.-D. Meng (✉) · J.-P. Ren · Y.-C. Song  
Inner Mongolia University of Science and Technology, Baotou 014010, China  
e-mail: haidongm@imust.edu.cn

J.-P. Ren  
e-mail: fm190135@126.com

Y.-C. Song  
e-mail: songyuchen@imust.edu.cn

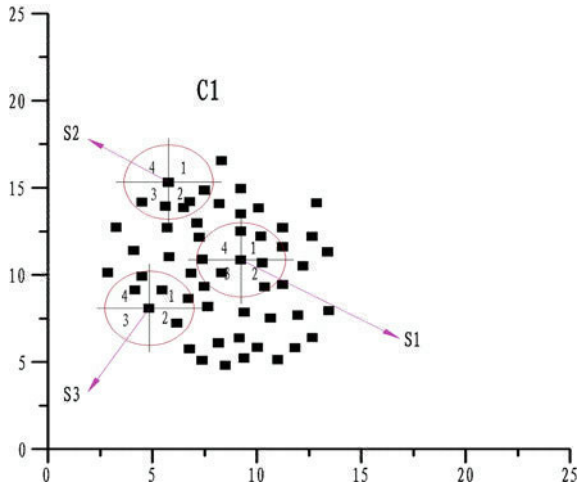
lack of the consideration for density of certain cluster; the second is only limited in discovering spherical-shaped clusters, but the best consideration for density is given in clustering. At present, there are some representative hierarchical clustering algorithms such as BIRCH, CURE, ROCK, and Chameleon algorithm, and all of them have both of merits and inadequacies. The BIRCH [1] algorithm cannot effectively discover non-spherical-shaped clusters because of the method of dealing with borders of clusters with the radius or diameter of clusters, and it is not feasible in addressing the high-dimensional data sets because the exact number of clusters and diameters can be hardly obtained. The CURE [2] algorithm can handle various shaped clusters, but the efficiency of the algorithm is low, and it is subjective to choosing the number of typical points, regardless of consideration for the connectivity between clusters. In 1999, Sudipno Guha put forward the ROCK [3] algorithm that adopts a approach about the global information based on the number of public linking points to measure the correlation of data points, this means the feature of connectivity is considered, but the proximity is ignored. In 1999, Karypis proposes the Chameleon [4] hierarchical clustering algorithm based on the dynamic modeling technology, though the Chameleon algorithm takes both connectivity and proximity into consideration, but there are still some shortcomings as follows: (1) The value of  $K$  in the  $K$ -nearest figure must be given before the experiment; (2) there is difficulty in the Min-Bisection. (3) the threshold of similarity must be given before the experiment [5], and in the literature [6], the probability of the limitation of dealing with high-dimensional data sets is pointed out in the Chameleon algorithm; In the stage of partition, an object may be assigned to a cluster, which doesn't actually belong to the cluster.

Considering the above clustering methods, the hierarchical clustering algorithm based on cluster outline is proposed in this paper, which reduces the influences of the existence of outliers, considers reasonably both the connectivity and proximity between clusters with an automatic choice of cluster centers, and optimizes the clustering result effectively.

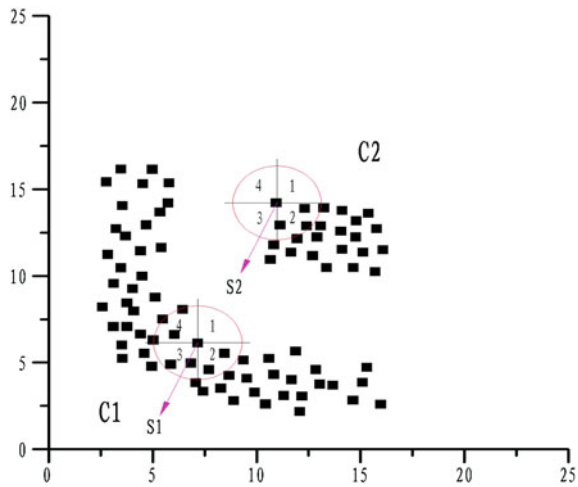
## 2 Definition and Solution

### 2.1 The Definition of Boundary Point and Cluster Outline

Cluster boundary points are defined as follows: In a cluster with  $n$ -dimensional data set, all data objects are regarded as spatial points distributed in  $n$ -dimensional data space. If a spatial point is discussed and considered in certain scope (such as the similarity-measuring standards based on distance and density), and a spatial point does not belong to all of  $2^n$  quadrants in  $n$ -dimensional data space at the same time, the point can be defined as a boundary point (as Figs. 1 and 2 shown in two-dimensional data space). The outline of a cluster consists of all boundary points belonging to the cluster (as Figs. 3 and 4 shown in two-dimensional data space).



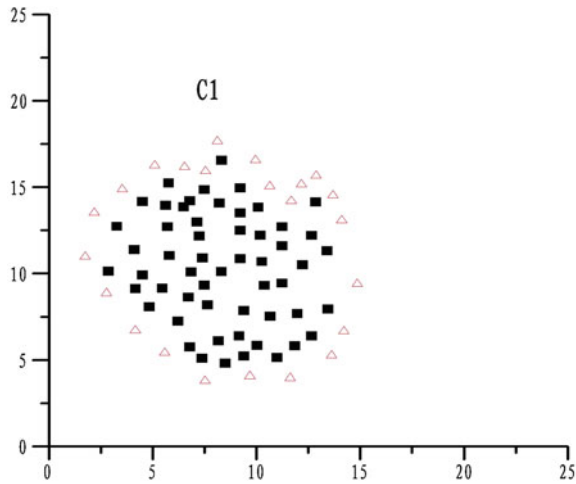
**Fig. 1** Boundary points of cluster  $C_1$



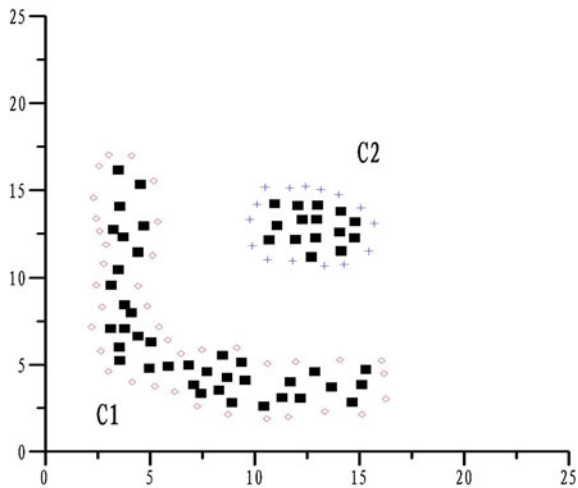
**Fig. 2** Boundary points of cluster  $C_2$

In Fig. 1, the point  $S_1$  of the cluster  $C_1$  is not an boundary point, because in its scope based on distance or density, there are always one or more points that can be found in any of its own quadrants of 1, 2, 3, and 4; the scope of the point  $S_2$  does not touch any point in its own quadrants of 1 and 4, so the point  $S_2$  is an boundary point; the third quadrant of the  $S_3$  can cover no point, so the  $S_3$  is an boundary point.

In Fig. 2, the points  $S_1$  and  $S_2$ , respectively, from  $C_1$  and  $C_2$  are boundary points with the same as shown in Fig. 1.



**Fig. 3** Cluster outline  $C_1$



**Fig. 4** Outlines of clusters  $C_1$  and  $C_2$

Cluster outline is defined as follows: All boundary points consist of the cluster outline; that is also the cluster outline is a collection of all boundary points.

In Fig. 3, the cluster outline  $C_1$  consists of all red triangle symbols.

In Fig. 4, the cluster outline  $C_1$  consists of all red diamond symbols, and  $C_2$  all blue cross symbols.

## 2.2 The Solution of Cluster Outline

The solution of finding cluster outline aims at searching for all boundary points of cluster. Boundary points are the space within the cluster of a relative positional relationship of the points, so the cluster constructs the relative position of the point and point matrix. A complete solution of cluster outline includes three steps.

### (1) Construction of the relative matrix $P$

Suppose a cluster  $C$ ,  $C = \{S_1, S_2, S_3, \dots S_n\}$ , and the cluster  $C$  is discussed and considered in two-dimensional data space consisting of the  $x$  axis and the  $y$  axis. The symbol  $D_0$  represents the threshold of distance or density. The expression  $D(S_i, S_j)$  represents the relative distance or density between the points  $S_i$  and  $S_j$ . The relative distance or density between the points  $S_i$  and  $S_j$  is defined as follows:

$$D(S_i, S_j) \leq D_0 \quad (1)$$

If the formula  $D(S_i, S_j) \leq D_0$  is true, and the expression  $S_{jX} > S_{iX}$  &  $S_{jY} > S_{iY}$  is true, this means that  $S_j$  is in the first quadrant of  $S_i$  and  $P_r = 1$ . So we can get the following expressions:

$$\begin{aligned} S_{jX} > S_{iX} \ \& \ S_{jY} > S_{iY} \Rightarrow P_r = 1 \\ S_{jX} > S_{iX} \ \& \ S_{jY} < S_{iY} \Rightarrow P_r = 2 \\ S_{jX} < S_{iX} \ \& \ S_{jY} < S_{iY} \Rightarrow P_r = 3 \\ S_{jX} < S_{iX} \ \& \ S_{jY} > S_{iY} \Rightarrow P_r = 4 \end{aligned} \quad (2)$$

Note: for the point  $S_i$ , the value of  $X$  is the value of  $S_{iX}$ , and the value of  $Y$  is the value of  $S_{iY}$ .

The relative matrix  $P$  is constructed as follows:

$$P = \begin{matrix} & S_1 & S_2 & S_i & S_n \\ S_1 & 0 & P_r & P_r & P_r \\ S_2 & P_r & 0 & P_r & P_r \\ S_i & P_r & P_r & 0 & P_r \\ S_n & P_r & P_r & P_r & 0 \end{matrix} \quad (i < n; \text{ if } D(S_i, S_j) > D_0, \text{ then } P_r = 0) \quad (3)$$

### (2) Search for boundary point

To figure out whether the point  $S_i$  is an boundary point or not, we can check the  $i$  column of  $S_i$  to find all different values of  $P_r$ . If the set composed of all values of  $P_r$  in the  $i$  column of  $S_i$  is not  $\{1, 2, 3, 4\}$ , the point  $S_i$  is an boundary point.

### (3) Search for cluster outline

The cluster outline is a definite set of all boundary points belonging to cluster in the step (2).

### 3 The Design of Hierarchical Clustering Algorithm Based on Cluster Outline

#### 3.1 The Clustering Between Clusters

##### (1) Standard of connectivity

Suppose that the outline of the cluster  $C_1$  is  $P_{r1}$  and  $C_2$  is  $P_{r2}$ . The symbol  $N$  shows the number of all shared boundary points between clusters. The calculation of  $N$  is as follows:

$$N_1 = \text{count}(P_{r1} \cap P_{r2}) \quad (4)$$

$$N_2 = \text{count}(P_{r1} \cap C_2) \quad (5)$$

$$N_3 = \text{count}(P_{r2} \cap C_1) \quad (6)$$

$$N_4 = N_2 + N_3 - N_1 \quad (7)$$

Note:  $N_1$  represents the number of the same points of  $P_{r1}$  and  $P_{r2}$ ;  $N_2$  represents the number of the same points of  $P_{r1}$  and  $C_2$ ;  $N_3$  represents the number of the same points of  $P_{r2}$  and  $C_1$ . Figure 5 shows the connectivity of the cluster outlines.

In Fig. 5, the clusters  $C_1$  and  $C_2$  share with the points of the same shape and color. The number of all boundary points of  $C_1$  and  $C_2$  is 17, which contains 10 yellow diamonds, 4 dark red stars, and 3 light red sunshine-shaped points. That also means the connectivity of the outline of the clusters  $C_1$  and  $C_2$  is 17.

##### (2) Standard of proximity

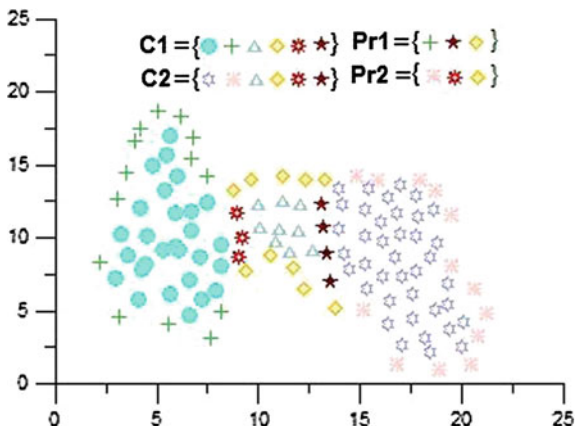
Suppose that the number of clusters is  $N$  with its corresponding outline of each cluster:  $S_1, S_2, \dots, S_n, S_i = \{B_{i1}, B_{i2}, \dots, B_{ij}\}$ , where  $S_i$  represents the outline of the cluster  $i$  and  $B_{ij}$  represents the boundary point  $j$  of the cluster  $i$ , then the average distance as the measurement of proximity of clusters is defined. The following formula shows the average distance between the cluster outlines:

$$\bar{D}(S_i, S_j) = \frac{1}{n_{S_i} n_{S_j}} \sum_{B_i \in S_i} \sum_{B_j \in S_j} \text{dist} \|B_i - B_j\|. \quad (8)$$

#### 3.2 The Clustering Between Clusters and Outliers

Considering a ratio  $r$ , the symbol  $n_r$  represents the number of all boundary points of one cluster, and the symbol  $n$  represents all points of the cluster. The ratio  $r$  can be set up as follows:

**Fig. 5** Connectivity of the outlines of clusters  $C_1$  and  $C_2$



$$r = \frac{n_r}{n} \quad (0 < r \leq 1) \tag{9}$$

The smaller the value of  $r$ , the denser the cluster, and vice versa.

When an outlier is assigned to a cluster to form a new cluster, the ratio of the new cluster is  $r'$ . If  $r'$  is less than or equal to  $r$  (the ratio of the original cluster), the outlier belongs to the new cluster. If the condition just mentioned is not true, we should keep the original situation not to be changed. The snake-shaped and linear clusters can be easily detected in this way to some degree.

### 3.3 The Process of the Algorithm

For a data set  $D_0$ , the hierarchical clustering algorithm based on cluster outline is as follows:

1. Input the value of threshold of distance or density ( $d_{\min} < r_0 < d_{\max}$ ), the number of data points:  $n$ , and the number of clusters:  $K$ .
2. Construct distance or density matrix  $d_{\text{matrix}}$ .
3. Take a data point  $X$  from the data set  $D_0$ , search out all points based on the threshold of distance of  $X$  in  $d_{\text{matrix}}$ , and combine  $X$  and all the points into a cluster in the condition of the number of all data points more than  $n$ , or else set these points as temporary outliers.
4. Construct the relative matrix  $P$  and form outlines of all clusters with step 3.
5. Calculate the number of the shared boundary points between clusters (the connectivity matrix), begin to cluster from the biggest number, and readjust the cluster outline until there is  $K$  clusters or the number of the shared boundary points is up to  $n$ .
6. Repeat the step 4 for the new outline.

**Fig. 6** Original image  
(500 × 375)



7. If the number of clusters is still more than  $K$ , then the proximity matrix can be calculated and the min distance can be combined into a cluster. Renew the cluster outlines until there are  $K$  clusters.
8. Handle the outliers according to equation (9).

## 4 The Experimental Analysis

Experiment 1: In order to facilitate the realization of the data sets and visualization of the experiment results, the experiment takes an image data as an experimental object, which consists of 500 pixels of width and 375 pixels of height. Each data point is represented by all RGB color values of the pixel. The following figures show the result of the experiment by contrast between the original image in Fig. 6 and the processed image in Fig. 7.

In Fig. 7, there are 10 clusters and an outlier data set. The experiment shows the advantage effectiveness in clustering with this algorithm.

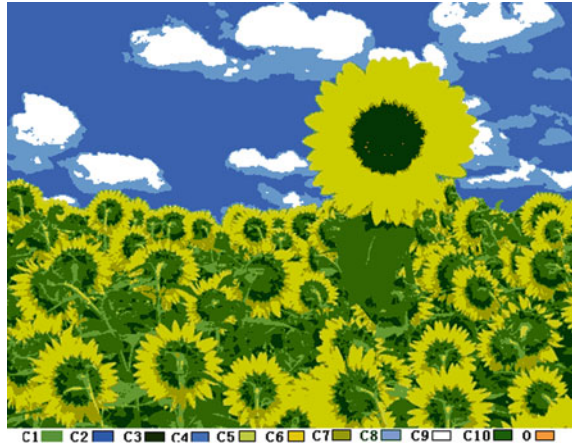
Experiment 2: In order to validate further the feasibility and effectiveness of the hierarchical clustering algorithm based on cluster outline, the experiment 2 is given with the Iris data set from the UCI. The Iris data set is a kind of biometric data set of three various flowers, which includes 150 records divided into three clusters, each cluster has 50 records, and each record consists of four attributes.

The results of the clustering in the Iris data set are shown in Fig. 8 by contrast between the clustering algorithm based on the cluster outline, the CURE algorithm, and the various algorithms provided by SPSS statistical analysis software.

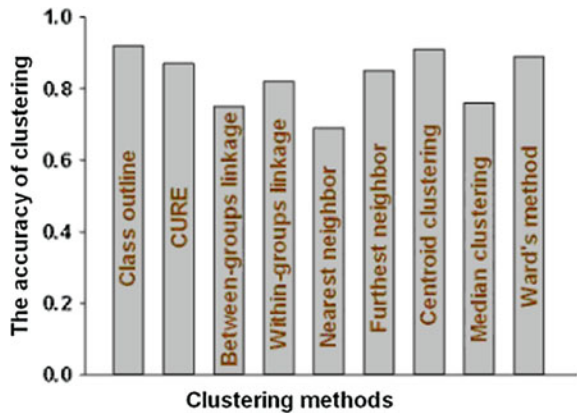
As shown in Fig. 8, the algorithm accuracy of the cluster outline and the centroid clustering in Iris data set obtained is more than 90 % (there are 14 records in error). They show a relatively good clustering effect.



**Fig. 7** Simulated image of experimental result



**Fig. 8** Accuracy comparison between clustering algorithms



The values of threshold of the initial distance and number of points can be hardly determined in clustering. In order to address this, the initial value of threshold can be small, and then, the processed points are compared with the rest of the outliers; if the rest of the outliers are more than 10 % of the number of the total data points, the value of threshold of the outliers should be set bigger to cluster again. It can guarantee that the number of new clusters combining the subclusters in terms of connectivity is more than  $K$ , and then, subclusters are combined again in terms of proximity; if the value of threshold of the number of the initial data points is given too small, the number of clusters can be less than  $K$ , and it can also give rise to the worst situation that there is only one cluster including all data points.

There is almost no change in the time complexity of the cluster-outline-based algorithm, in contrast to other traditional clustering algorithms, and in the worst situation, the time complexity is  $O(n^2)$ , where the time is mainly consumed in

dealing with distance matrixes and relative matrixes. However, the points of outline are processed instead of all points of whole clusters, so the space complexity is relatively low, especially for densely based clusters.

## 5 Conclusions

The feasibility and effectiveness of the hierarchical clustering algorithm based on cluster outline are proven in the experiments. Because of the standard of measuring the similarity between clusters based on both connectivity and proximity, the clusters with arbitrary shape, noise, or outliers can be easily discovered and detected. And the cluster outline is chosen to replace the whole cluster, so the task of dealing with huge data points can be avoided in the algorithm.

The concept of cluster outline has much practical utility in application, which can also be applied to other traditional clustering methods, such as discovering shapes of clusters by the visualization of cluster outline; measuring density of clusters by the proportion between the number of boundary points and all points of one cluster; and detecting noise or outliers.

**Acknowledgments** The research is funded in part by The Natural Science Foundation of Inner Mongolia under Grant No. 2012MS0611, Chunhui Project of Ministry of Education under Grant No. Z2009-1-01041, and Higher School Science Research Project of Inner Mongolia under Grant No. NJZZ11140.

## References

1. Zhang, T., Ramakrishnan, R., Livny, M.: BIRCH: An efficient data clustering method for very large databases. Technical Report, Computer Sciences Department, University of Wisconsin-Madison (1995)
2. Guha, S., Rastogi, R., Shim, K.: Cure: An efficient clustering algorithm for large database. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, pp. 73–84, Seattle, Washington (1998)
3. Guha, S., Rastogi, R., Shim, K.: ROCK: A robust clustering algorithm for categorical attributes. In: Proceedings of the 15th ICDE, pp. 512–521, Sydney (1999)
4. Karypis, G., Han, E.-H., Kumar, V.: Chameleon: hierarchical clustering using dynamic modeling. *Computer* **32**(8), 68–75 (1999)
5. Long, Zhen-Zhen, Zhang, C., Liu, F.-Y., Zhang, Z.-W.: An improved chameleon algorithm. *Comput. Eng.* **20**(35), 189–191 (2009)
6. Ma, X.-Y., Tang, Y.: Research on hierarchical clustering algorithm. *Comput. Sci.* **34**(7), 34–36 (2008)

# Exploration of Fault Diagnosis Technology for Air Compressor Based on Internet of Things

Zheng Yue-zhai and Chen Xiao-ying

**Abstract** With the development of network and communication technology, this article puts forward to new methods and ideas on the fault diagnosis technology of air compressor based on Internet of Things, ensure the safe, stable, and reliable running of air compressors, and carry out some beneficial exploration for remote diagnosis technology of mechanical equipment.

**Keywords** Internet of Things · Air compressor · Fault diagnosis technology · Remote monitoring

## 1 Introduction

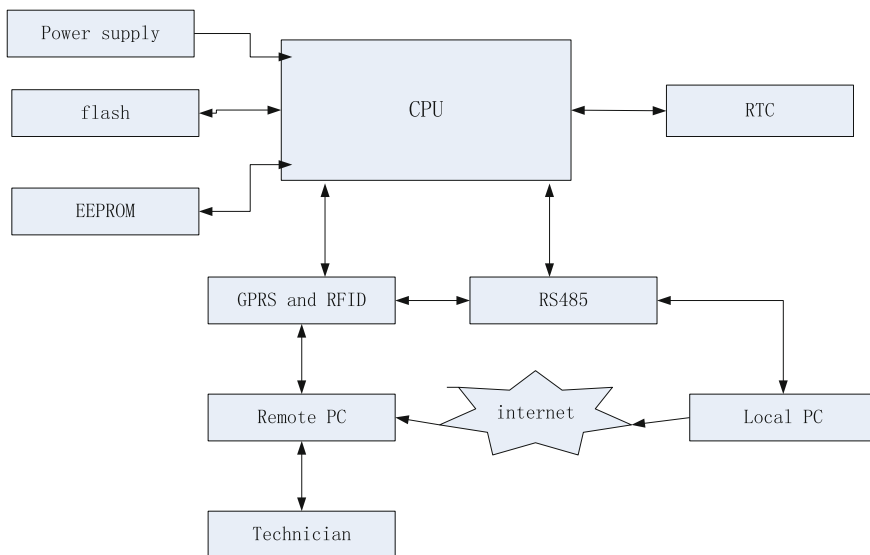
Compressed air is the second largest power after the electric power energy, air compressor (hereinafter referred to as the compressor) is the main part of manufacturing the compressed air, and it is the original motivation (typically a motor) that convert the mechanical energy into a air pressure energy. Air compressors are widely used in industrial production, and most of the work environments are harsh [1]. When a machine fails, it cannot be diagnosed and be treated immediately, this must cause the increase of the loss. So the establishment of remote intelligent control for monitoring the response of the system has become a pressing problem of air compressor technology for the early prediction of failure.

In this paper, the use of Internet of Things technology carries out a useful exploration of the air compressor fault diagnosis technology.

---

Z. Yue-zhai (✉) · C. Xiao-ying  
Quzhou College, Quzhou, Zhejiang 324000, People's Republic of China  
e-mail: zyz8602@126.com

C. Xiao-ying  
e-mail: chxy1@163.com



**Fig. 1** Basic structural diagram of the system

## 2 The Remote Diagnosis System Designment of the Air Compressor Using Internet of Things

With the attention of the state to the Internet of Things technology, the Internet of Things technology has been more widely used in China. The application of the Internet of Things technology in air compressor for remote monitoring and diagnosis function conforms to the needs of society, and it will show considerable application prospects for the future development.

Basic structural diagram of the system, see Fig. 1

This system includes CPU with external GPRS system, flash memory, RS485 communication port; by the RS485 port, the real-time state of the compressor is transferred to a host computer and realizes the remote communication of the screw air compressor. The operator can start and stop the screw air compressor through the computer. The GPRS module and radio frequency identification (RFID) of the screw air compressor control system can fix the position of the air compressor, get to know where the parameters come, and ensure to control the machine correctly. The flash memory is used to store the historical running state, and when a mistake happens, the parameters can be drawn to help diagnose the problem and solve it. In addition, the acquisition of parameters can be transferred to the server through the Internet, there are many data from the working compressor in the server, and there is an expert diagnostic system according to the experience, the server can deal with collected data and make the plan of the equipment maintenance [2].

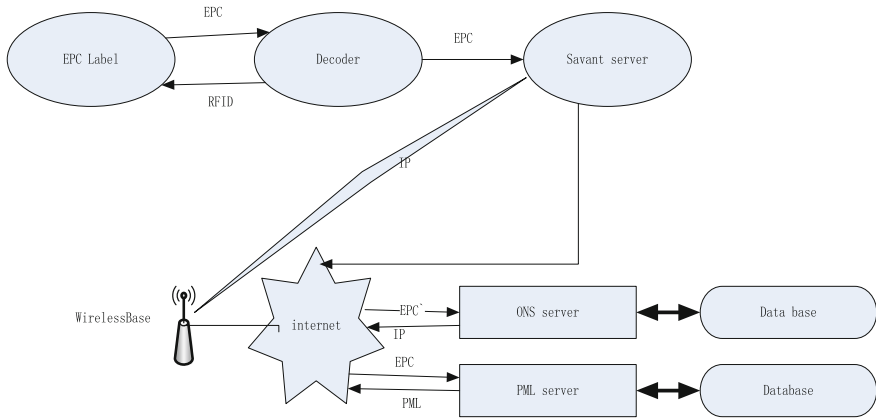


Fig. 2 Technical principle logic diagram

### 3 Field Information Collection with the Internet of Things Technology

In this paper, the technology of the Internet of Things is mainly used to locate the compressor, as well as its relevant technical indicators, operation parameters, etc.

Internet of Things is a kind of network that according to the agreed protocol, through the RFID, infrared sensors, global positioning system, laser scanners, and other information sensing devices, connects any objects with the Internet, and exchanges information and communication, and realizes the object's intelligent identification, location, tracking and monitoring and management. With the Use of the Internet of Things technology, people can implement intelligent and accurate regulation and operation of the devices in the network [3].

Internet of Things' key technology includes five segments: electronic product code (EPC), the information identification system (ID system), the EPC middle-ware (implementation information filtering and sampling), discovery service (information discovery service), the EPC information service (EPCIS). Its technical principle logic diagram is shown in Fig. 2:

There is a quite good coverage of wireless communication network in China, and it is a good infrastructure of the Internet of Things. The M2M business support platform launched by China mobile can provide the M2M business users of the data acquisition, transmission, processing, and business management, and other functions.

## 4 To Realize Remote Monitoring Using Communication Technology

This article adopts the type of site monitoring and remote monitoring, and through the GPRS network or the Internet, information is transmitted to the specified server, carries on the analysis processing, and achieves remote monitoring of equipment through the communication interface link.

Remote monitoring refers to the use of a computer to monitor and control of the remote industrial production process control system via a wired or wireless network system. A remote monitoring system is a computer hardware and software system that can achieve remote monitoring.

There are two types of remote monitoring system, one is no on-site monitoring system, but collects the data directly to a remote computer for processing, this remote monitoring has no difference with the general site monitoring except there is a long distance to transmit; the other is site monitoring and remote monitoring coexist. Generally, the remote monitoring adopts fieldbus technology connecting distributed sensors of monitoring equipment and develops the discrete cell to integrate unit, and at last, Intranet is built.

Remote monitoring function to achieve

1. Acquisition and process function: detect sample and pretreat all kinds of analog or digital signals, and output them in a certain form, such as printing form, display, or database server to help technician to analysis and know the situation of the compressor.
2. Supervision function: analyze, summarize, finish, and compute the detected real-time data, inputted parameters by the operators and so on, store them for the real-time data and historical data.
3. Management function: with the help of the existent valid data, images and reports related with the working condition, analysis, fault diagnosis, risk predict the working state, alarm with noise, light and electrical form.
4. Control function: Based on the detected processing information and pre-determined control strategy, control output is formed and works on the compressor directly.

Because the remote monitoring system can realize real-time data collection of the site operation data and rapid concentration, the base for remote fault diagnosis technology is formed. The remote monitoring system connects with the enterprise Intranet; this makes it possible for manager to grasp the working states, and it works with the management strategy system, a much more advanced applying system will be built. It provides the possibilities of no-one on-duty at the site and achieves more profit ultimately.