Advances in Intelligent Systems and Computing 265

Sirapat Boonkrong
Herwig Unger
Phayung Meesad *Editors*

Recent Advances in Information and Communication Technology

Proceedings of the 10th International Conference on Computing and Information Technology (IC^{2}IT2014)

Springer

Advances in Intelligent Systems and Computing

Volume 265

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

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

Sirapat Boonkrong · Herwig Unger
Phayung Meesad
Editors

Recent Advances in Information and Communication Technology

Proceedings of the 10th International Conference on Computing and Information Technology ($IC^2IT2014$)

Springer

Editors
Sirapat Boonkrong
Faculty of Information Technology
King Mongkut's University of Technology
North Bangkok
Bangkok
Thailand

Phayung Meesad
Faculty of Information Technology
King Mongkut's University of Technology
North Bangkok
Bangkok
Thailand

Herwig Unger
Communication Networks
University of Hagen
Hagen
Germany

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-319-06537-3 ISBN 978-3-319-06538-0 (eBook)
DOI 10.1007/978-3-319-06538-0
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014936765

Preface

This volume contains the papers of the 10^{th} International Conference on Computing and Information Technology (IC^{2}IT2014). IC^{2}IT is an annual conference, which is held in conjunction with the National Conference on Computing and Information Technology (NCCIT), one of the leading Thai national events in the area of Computer Science and Engineering. IC^{2}IT provides a venue for the presentation and discussion of current research in the field of computing and information technology.

IC^{2}IT2014 took place between 8^{th} and 9^{th} May at Angsana Laguna, Phuket, Thailand. This is the first time that the conference is located in the South of Thailand. Following the interests of our participants of the last events, IC^{2}IT2014 proceedings has been structured into five main tracks: Data Mining Algorithms and Methods, Application of Data Mining, Infrastructures and Performance, Text Analysis and Search, and Security. Also, we are delighted to announce that the conference got a large financial support by the Thai government with the aim of encouraging and improving research in the ASEAN countries.

Although the support for the development of ASEAN and AEC (ASEAN Economic Community) are in the focus of the conference, the committee received 96 submissions from authors of 25 countries at 5 continents. The stable number of contributions over the last years is an indicator that our event is well established in the scientific community; with respect to the exploding number of conferences we are proud on that in particular. As usual, each submission was reviewed by at least 2-4 members of the program committee to avoid contradictory results. On these suggestions, the committee decided to accept 32 papers for oral presentation and inclusion in the conference proceedings.

Again, Springer agreed to publish our proceedings in its well-established and worldwide distributed series on Advances in Intelligent Systems and Computing. Last but not least, two internationally well-known scientists, from Germany and Japan, have been invited and accepted to give keynote talks to our participants.

A special thank is given to KMUTNB and its President, Professor Dr. Teeravuti Boonyasopon for his support of our conference from the first year on, and for providing us with a lot of resources from KMUTNB. We hope that IC^{2}IT again provides great opportunities for academic staff, students and researchers to present their work. IC^{2}IT is also a platform for exchange of knowledge in the field of computer and information

technology and shall inspire researchers to generate new ideas and findings and meet partners for future collaboration. We also hope that our participants use this occasion to learn more about Phuket and its beautiful scenery, people, culture and visit its famous beaches before or after the conference.

We would also like to thank all authors for their submissions and the members of the program committee for their great work and valuable time. The staff members of the Faculty of Information Technology at King Mongkut's University of Technology North Bangkok have done a lot of technical and organisational works. A very special and warm thank you is given to our web masters: Mr. Jeerasak Numpradit, and Mr. Armornsak Armornthananun. Without the meticulous work of Ms. Watchareewan Jitsakul the proceedings could not have been completed in the needed form at the right time.

After so much preparation, all of the organisers of course hope and wish that IC^{2}IT2014 will again be a successful event and will be remembered by the participants for a long time.

February 27, 2014
Bangkok

Sirapat Boonkrong
Herwig Unger
Phayung Meesad

Organization

In Cooperation with

King Mongkut's University of Technology North Bangkok (KMUTNB)
FernUniversität in Hagen, Germany (FernUni)
Chemnitz University, Germany (CUT)
Oklahoma State University, USA (OSU)
Edith Cowan University, Western Australia (ECU)
Hanoi National University of Education, Vietnam (HNUE)
Gesellschaft für Informatik (GI)
Mahasarakham University (MSU)
Ubon Ratchathani University (UBU)
Kanchanaburi Rajabhat University (KRU)
Nakhon Pathom Rajabhat University (NPRU)
Mahasarakham Rajabhat University (RMU)
Rajamangala University of Technology Lanna (RMUTL)
Rajamangala University of Technology Krungthep (RMUTK)
Rajamangala University of Technology Thanyaburi (RMUTT)
Prince of Songkla University, Phuket Campus (PSU)
National Institute of Development Administration (NIDA)
Council of IT Deans of Thailand (CITT)

IC²IT 2014 Organizing Committee

Conference Chair	Phayung Meesad, KMUTNB, Thailand
Technical Program Committee Chair	Herwig Unger, FernUni, Germany
Secretary and Publicity Chair	Sirapat Boonkrong, KMUTNB, Thailand



Program Committee

M. Aiello	RUG, The Netherlands
T. Anwar	UTM, Malaysia
T. Bernard	Syscom CReSTIC, France
W. Bodhisuwan	KU, Thailand
A. Bui	Uni Paris 8, France
T. Böhme	TU Ilmenau, Germany
M. Caspar	CUT, Germany
P. Chavan	MMMP, India
T. Chintakovid	KMUTNB, Thailand
H. K. Dai	OSU, USA
D. Delen	OSU, USA
N. Ditcharoen	URU, Thailand
T. Eggendorfer	HdP Hamburg, Germany
R. Gumzej	Uni Maribor, Slovenia
H. C. Ha	HNUE, Vietnam
M. Hagan	OSU, USA
P. Hannay	ECU, Australia
W. Hardt	Chemnitz, Germany
C. Haruechaiyasak	NECTEC, Thaialnd
S. Hengpraprohm	NPRU, Thailand
K. Hengproprohm	NPRU, Thailand
U. Inyaem	RMUTT, Thailand
M. Johnstone	ECU, Australia
T. Joochim	URU, Thailand
J. Kacprzyk	PAS, Poland

A. Kongthon	NECTEC, Thailand
S. Krootjohn	KMUTNB, Thailand
P. Kropf	Uni Neuchatel, Switzerland
M. Kubek	FernUni, Germany
S. Kukanok	MRU, Thailand
G. Kulkarni	MMMP, India
K. Kyamakya	Klagenfurt, Austria
J. Laokietkul	CRU, Thailand
U. Lechner	UniBw, Germany
M. Lohakan	KMUTNB, Thailand
J. Lu	UTS, Australia
A. Mikler	UNT, USA
A. Mingkhwan	KMUTNB, Thailand
C. Namman	URU, Thailand
P. P. NaSakolnakorn	MRU, Thailand
C. Netramai	KMUTNB, Thailand
K. Nimkerdphol	RMUTT, Thailand
S. Nitsuwat	KMUTNB, Thailand
S. Nuanmeesri	RSU, Thailand
N. Porrawatpreyakorn	KMUTNB, Thailand
P. Prathombutr	NECTEC, Thailand
A. Preechayasomboon	NECTEC, Thailand
P. Saengsiri	TISTR, Thailand
P. Sanguansat	PIM, Thailand
R. Shelke	MMMP, India
S. Smanchat	KMUTNB, Thailand
M. Sodanil	KMUTNB, Thailand
S. Sodsee	KMUTNB, Thailand
B. Soiraya	NPRU, Thailand
T. Srikhacha	TOT, Thailand
W. Sriurai	URU, Thailand
P. Sukjit	FernUni, Germany
D.H. Tran	HNUE, Vietnam
K. Treeprapin	UBU, Thailand
D. Tutsch	Uni Wuppertal, Germany
N. Utakrit	KMUTNB, Thailand
C. Valli	ECU, Australia
M. Weiser	OSU, USA
N. Wisitpongphan	KMUTNB, Thailand
A. Woodward	ECU, Australia
K. Woraratpanya	KMITL, Thailand
P. Wuttidittachotti	KMUTNB, Thailand

Contents

Session 2: Application of Data Mining

Session 3: Infrastructures and Performance

Wireless Mesh Networks and Cloud Computing for Real Time Environmental Simulations

Peter Kropf[1], Eryk Schiller[1], Philip Brunner[2], Oliver Schilling[2], Daniel Hunkeler[2], and Andrei Lapin[1]

[1] Université de Neuchâtel, Computer Science department (IIUN), CH-2000 Neuchâtel, Switzerland
{peter.kropf,eryk.schiller,andrei.lapin}@unine.ch
[2] Université de Neuchâtel, Centre for Hydrogeology and Geothermics (CHYN), CH-2000 Neuchâtel, Switzerland
{philip.brunner,oliver.schilling,daniel.hunkeler}@unine.ch

Abstract. Predicting the influence of drinking water pumping on stream and groundwater levels is essential for sustainable water management. Given the highly dynamic nature of such systems any quantitative analysis must be based on robust and reliable modeling and simulation approaches. The paper presents a wireless mesh-network framework for environmental real time monitoring integrated with a cloud computing environment to execute the hydrogeological simulation model. The simulation results can then be used to sustainably control the pumping stations. The use case of the Emmental catchment and pumping location illustrates the feasibility and effectiveness of our approach even in harsh environmental conditions.

Keywords: wireless mesh network, cloud computing, data assimilation, environmental measurements, hydrogeological modelling and simulation, ground water abstraction.

1 Introduction

Climatic or hydrological systems are driven by highly dynamic forcing functions. Quantitative numerical frameworks such as simulation models are powerful tools to understand how these functions control the systems' response. Models are, however, always imperfect descriptions of reality and therefore model calculations increasingly deviate from the "real" physical conditions of the environmental system simulated. We can alleviate these biases by a real time integration of field data into the modeling framework (data assimilation). To accomplish this goal, we have to constantly monitor the environment through dense networks of sensors deployed over the geographical area concerned. The technology should provide high performance even in the case of harsh meteorological conditions (snow, low temperatures, fog, strong winds, etc.) and other location and infrastructure related limitations like high altitude, lack of access to the power grid, and limited accessibility (resulting in long access delays inducing significant installation/maintenance costs).

S. Boonkrong et al. (eds.), *Recent Advances in Information and Communication Technology,* Advances in Intelligent Systems and Computing 265,
DOI: 10.1007/978-3-319-06538-0_1,

1.1 Wireless Infrastructure

In principle, communication networks can be wired or wireless. However, building a vast and complex wired infrastructure is costly and can be technically impossible in remote locations. An alternative are radio-based technologies, which do not require expensive cabled infrastructures. Moreover, this technological choice is extremely portable, because one can easily transfer equipment from one location to another when necessary. The first choice transport technology would be GSM/UMTS, however, this solution suffers from significant shortcomings. On the one hand, it is infeasible to equip every station with a GSM/UMTS connection in the case of vast measuring networks because of the associated cost of this operation, while the provider may charge for every additional SIM card. On the other hand, there exist important locations from an environmental perspective that have poor or non-existent coverage (e.g., highly elevated regions in Swiss Alps). These drawbacks force us to search for another scalable transport technology, which may grow to reach large proportions and provide us with good coverage over remote locations. Because of recent progress in the domain of low power wireless devices we may operate Wireless Mesh Networks that allow us to significantly cut operational expenses.

Wireless Mesh Networking is an interesting communication scheme which can provide cheap Internet connectivity delivered to end users at the last mile, an easily deployable multi-hop wireless bridge between distant bases in no-direct line of sight scenarios, or a wireless back-haul connecting sensors of different purposes such as environmental monitoring or smart-home applications. To properly deploy a wireless network, there are numerous hardware and software challenges. The hardware has to be properly selected to operate under a specific power consumption regime [1], e.g., when a node is solar powered, it has to harvest and store enough energy during the day-light operation to work uninterruptedly at night. Wireless cards and antennas have to provide an acceptable signal strength to allow for high throughput, while the node setup has to provide satisfactory performance such as computational power for ciphering and packet forwarding or other network adapters able to accommodate traffic coming from wireless interfaces. The experience obtained from pilot projects installed in remote and mountainous regions for environmental monitoring [2,3] and backup backbones in urban areas illustrates that mesh networks perfectly integrate into the existing AAA (Authentication, Authorization, Accounting) [4,5], monitoring and cloud infrastructure schemes of Swiss universities. For the purpose of this work, we use Wi-Fi based backhauls to transport information from environmental sensors to Internet storage facilities in real time.

1.2 Data Storage and Monitoring

In addition to provisioning the transmission infrastructure, facilities for data storage and processing have to be developed. Our studies reveal several similarities between environmental monitoring in the wireless mesh setup and network monitoring provided by typical monitoring agents such as Nagios, Zabbix, and

SNMP. In all these cases, current peripherals' status is reported to central storage for future analysis and visual presentation, while the information retrieval is triggered on a time basis. Our experience shows that we can re-use network monitoring for environmental purposes by configuring monitoring agents to constantly read out values provided by a sensor, e.g., every 15 mins. However, this methodology requires specific counter measures against Wi-Fi backhaul failures as we cannot afford the loss of environmental data if the network is inaccessible at a given time.

1.3 Going to the Clouds

The data collected can be processed in numerical models. In our case, we are simulating the use of groundwater water resources in a highly dynamic river-aquifer system—the Emmental[1] in Switzerland. The purpose of the simulation approach is to provide a quantitative basis for sustainable water resource management. Groundwater in the aquifer of the Emmental is pumped to supply the city of Bern with drinking water. However, the abstraction of groundwater[2] causes the water table in the aquifer to drop and can increase the infiltration from the river with adverse impacts on the stream ecosystems, but to what extent and how fast groundwater abstraction influences the flow in the river depends on the system state[3] (i.e., how much water is stored in the aquifer). To optimize the amount of water pumped (pumping scheme), predictions on how groundwater abstraction will affect the system are required. This paper describes recent developments in data acquisition and transmission infrastructures, integrated in the data assimilation system with the goal of generating a real-time pumping scheme for the Emmental. As discussed in the next section, predictions are generated using computationally expensive models that simulate the interactions and feedback mechanisms between the river, the aquifer, the pump, and climatic forcing functions such as precipitation. The models are continually updated with acquired field data. The data assimilation approach implemented for this task requires that multiple models are run in parallel. This allows us to assess the reliability of the proposed pumping rates in a stochastic way. However, computational costs for such an approach are significant. Running a few models in parallel is challenging since even a single one may require a few days of computations on an ordinary desktop machine. We decided to use a recently developed cloud computing paradigm for our computations to speed up the whole process, while the cloud allows us to run a few parallel computing workers. Cloud computing is a growing business with many established players such as Salesforce, Amazon, Akamai, and Google. We integrated our work with the ongoing SwissACC project [6], which aims to establish a Swiss nationwide cloud computing infrastructure. Cloud providers offer several service models,

[1] In German, the Emmental means the valley of the Emme river.
[2] i.e., pumping groundwater from the aquifer.
[3] This system state should be understood in the physical sense, i.e., similar to states in thermodynamics and definitely not as stored information in automata theory.

however, we concentrate on Infrastructure as a Service (IaaS), which provides us with the required number of powerful Virtual Machines (VMs) on-demand and remote control through the Internet. The VMs are extremely useful, because they can accommodate any generic type of computations, while they do not actually require any physical maintenance from the user side. SwissACC builds upon the open-source infrastructure, OpenStack[4]. OpenStack comes along with the Nova controller, which automates pool managing of worker resources. For storing data in the cloud, SwissACC integrates the S3 (Simple Storage Service) driver provided by Amazon[5] with the necessary API.

This work is organized in the following way. Section 2 provides a detailed problem description. In Section 3, we specify our proposed solution and we provide the most important implementation details. The results are gathered in Section 4. Finally, we conclude in Section 5.

2 Use Case: Groundwater Pumping

Drinking water supply in Switzerland is largely based on groundwater (about 80%). Numerous water supply systems abstract groundwater close to rivers. Surface water and groundwater systems interact in highly dynamic and complex ways [7,8], and therefore abstracting groundwater in the vicinity of rivers can substantially influence these dynamics [9]. Environmental laws demanding minimum water levels are in place in Switzerland. This gives rise to a challenging optimization problem. The critical parameters are the discharge in a river[6] (which is minimal due to strict environmental laws), the amount of water stored in the aquifer and the drinking water requirements. Balancing these target functions by adjusting the pumping rate thus requires a solid and quantitative understanding of the dynamics and the interactions of the river-aquifer system.

The Emmental is a perfect example that illustrates the tradeoff between the need for drinking water supply and minimal discharge in the river. The Emmental is a pre-alpine river catchment (about 200 km^2) in central Switzerland (Fig. 1). The catchment features steep hydraulic gradients with rapid groundwater flow rates (up to 100 m/d). The Emme River itself is highly dynamic (discharge between 0 and 300 (m^3/s). The aquifer pumped close to Emme provides roughly 45% of the drinking water for Bern, the Swiss capital. Groundwater abstraction in the Emmental increases the infiltration from the river to the aquifer. In fact, during low flow periods, groundwater abstraction often causes the river to dry up. The stream water levels in the the upper Emmental are strongly affected by seasonality and are highly sensitive to dry periods. In 2003 and 2011 large stretches of the river ran completely dry, as illustrated in Fig. 2. This pronounced seasonality adds an additional level of complexity to the system.

The efficiency and sustainability of water resources management in the Emmental is directly linked to the amount of water pumped from the aquifer. A

[4] There obviously exist other open-source infrastructures such as OpenNebula.
[5] `http://aws.amazon.com/s3/`
[6] The discharge in a river is the volumetric flow rate.

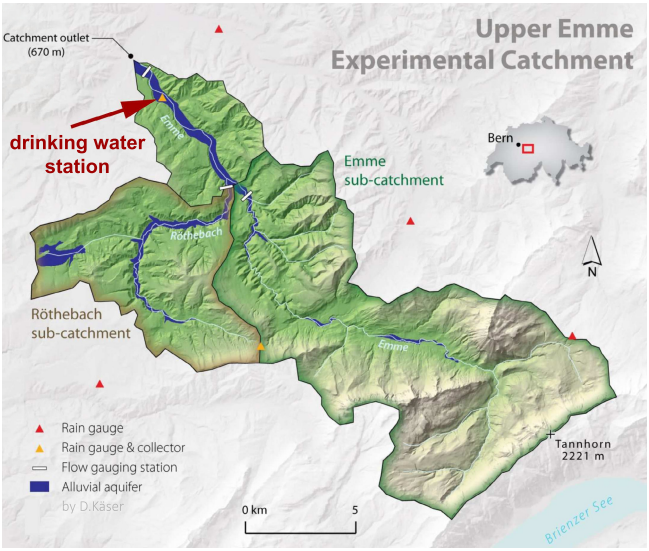

Fig. 1. The upper Emmental catchment is located close to the Swiss capital Bern (shown on the map in the top right corner). The Emme and the Roethebach rivers flow downhill from southeast and continue through the main upper Emmental after their confluence. Figure provided by D. Käser.

Fig. 2. A comparison between a high and a low water level situation in the upper Emmental, right below the pumping station. Photos provided by D. Käser.

quantitative approach to simulate the system is required to optimize pumping rates in this dynamic environment. Numerical models are therefore necessary for this task. By continuously incorporating field observations in the model simulation, any potential model biases can be identified, quantified and corrected. This process is called data assimilation [10]. While data-assimilation approaches are widely used in climatological models, they are rarely applied to hydrogeological

simulations [11]. The "ingredients" for data assimilation systems are a measurement and communication network that provides field observations in real time; a data storage infrastructure; and numerical models that predict for example how different groundwater abstraction schemes affect the flow of the river. Based on these simulations, the pumping rates can be regulated in an optimal way.

3 Implementation

In the Emme river valley, we have established a wireless setup which contains a few stations with environmental sensors attached through USB (such as temperature and pressure meters) and other necessary stations acting as wireless backhaul thus forwarding packets and providing Internet connection (Fig. 3). From the hardware perspective, every node uses the Alix3d2 motherboards with two on-board mini-pci slots. We use the mini-pci bus to install the Winstron DNMA-92 IEEE 802.11abgn interfaces, while our wireless links are provided by directional antennas of high-gain. When the electric power grid is not available, we equip a node with a solar panel and battery to secure a continuous 24 hours operation (normally, the battery charges during the day-light operation). Our nodes are placed in a special-purpose enclosure which protects them against outdoor conditions, e.g., humidity. When a high number of Wi-Fi interfaces are required, we gather a few mother boards together in a single box. The Linux based ADAM system[7] serves as the operating system platform. Due to the installed OLSR and IEEE 802.11s, the network is easily expandable, i.e., the installation of a new node requires little attention from the administration perspective.

3.1 Environmental Monitoring

Zabbix[8] provides a client-server infrastructure which allows us to monitor and control remote machines. There are a large number of predefined parameters, while Zabbix also provides an opportunity to launch user-defined commands to support user-specific peripherals. Due to this feature, we are able to equip Zabbix agents with drivers, i.e., special purpose applications which read out environmental parameters from the sensors through USB and provide the agent with the received data. We deployed one running instance of the Zabbix agent on every node in the mesh and one instance of the Zabbix server at the central storage (online database). To control the Zabbix server (e.g., including another sensor), we are provided with an advanced back-end web interface and rich logging system.

The online database allows the access of the measured environmental system state in real-time, providing the basis for a real-time forecasting system to control the groundwater abstraction rates. Periodically, our software asks sensors about current values of the measurements. Then, again periodically, the Zabbix server

[7] Developed by the University of Bern: `http://cds.unibe.ch`
[8] `http://www.zabbix.org`

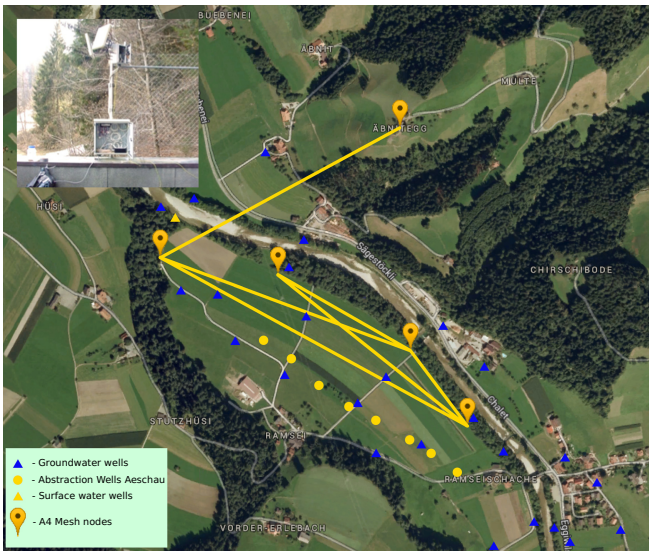

Fig. 3. Environmental setup of the network

executes a remote driver on the nodes and obtains values of the measurements. This data is transferred to the database on the server. Finally, the data can be accessed through the web interface (Fig. 4). The Zabbix infrastructure provides us with wide variety of tools for drawing plots and applying simple formulas to the data; it fits well to the requirements of our application context.

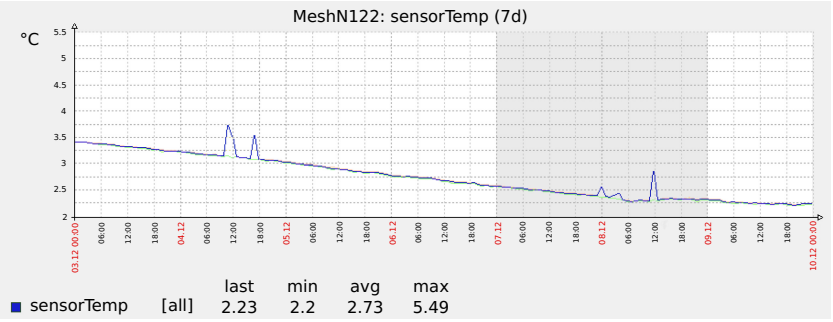

Fig. 4. Online environmental data output (the larger peaks appear to be artefacts)

3.2 Real-Time Modeling

The latest generation of numerical models is now able to simulate the interactions between surface water and groundwater in a fully coupled way [12]. One of

the most advanced codes in this respect is HydroGeoSphere [13]. In addition to simulating surface water and groundwater interactions, the code can also simulate vegetation dynamics as well as the recharge[9] and discharge processes in response to precipitation, evapotranspiration or groundwater abstraction. HydroGeoSphere is therefore used to simulate the Emmental system.

The geometric setup is based on a high resolution digital terrain model. The numerical coupling between the surface and subsurface domain is conceptualized through a dual node approach, as described in [13]. The model requires a very large amount of parameters, such as hydraulic properties of the streambed, the soil or the aquifer. These parameters cannot be measured in the field at the required spatial resolution and therefore have to be estimated. Numerous approaches are available in this regard. A "classic" way is to adjust parameters in order to minimize the mismatch between the available historical measurement data and the corresponding model simulations. Once a model reproduces historic measurement data satisfactorily, it is used to predict future system states under changing forcing functions. However, all numerical models are a simplification of reality, both in terms of the processes considered as well as in their parameterization. Therefore, any calibrated model will sooner or later deviate from the real, physical system state. Clearly, the model state (i.e., the simulated water levels or the actual discharge in the river) has to be as close as possible to the real system in order to provide reliable predictions on how a planned pumping scheme will affect the system in the near future. Therefore, the "classical" calibration approach is not well suited for this application. By using a data assimilation approach, the model is continuously updated in terms of its state and parameters. We implement a data assimilation approach similar to the work of [11]. Currently, the assimilated measurements consist of stream level and water tables along the river.

HydroGeoSphere is a numerically demanding code, and the highly dynamic interactions between surface water and ground water require a fine temporal discretization scheme. Moreover, numerous models are running in parallel to explore the influence of different pumping schemes, as well as different possible model parameterizations. To accomplish this computational burdon, significant computational resources are required (see next section). The multiple simulations of possible abstraction rates with the corresponding predicted impacts on the river-aquifer system allow us to identify the optimal pumping volumes in consideration of the environmental laws and drinking water demands. With the development of this simulation system, the pump can be operated in an optimal way. However, a remote regulation system that transmits the optimal pumping rates to the pump must be implemented.

3.3 Swiss Academic Compute Cloud

Our Cloud based solution allows researchers to perform resource consuming computations with minimal efforts. Firstly, the data collected with the environmental

[9] Water infiltrating the soil reaching the underground water table.

sensors is stored on the pilot-project Swiss Academic Compute Cloud (SACC), a unified cloud service providing storage and computation resources for Swiss academic institutes, which makes use of a specialized S3-based cloud repository—the Object Storage (OS). The user front-end is developed in Django[10], which is a free open source Python based web application framework that provides us with the model-view-controller architectural paradigm. Our currently implemented front-end allows a researcher to visit the web-page, choose required input data, initiate required tasks, and download the results of completed computations. Behind the user front-end, we integrate a Python engine—GC3Pie[11], which is developed by the GC3 group at the University of Zürich[12] and enables all cloud related operations such as starting and stopping new workers. We configured our framework to run several instances of HydroGeoSphere on allocated working VMs. Due to the infrastructure configuration, every instance of the HydroGeoSphere is provided with the input files from the OS, which in turn also acts as a storage facility for models returned by completed instances. This OS-based data organization scheme is important, because it provides portability as there are many different cloud providers supporting this storage manner.

4 Results

Firstly, we deployed a measuring and transporting mesh network in the Emmental which proves its high performance and reliability in harvesting environmental data. Secondly, the first numerical HydroGeoSphere model that is capable of simulating the interactions and feedback mechanisms between the river, the aquifer and the pumps has been set up. It includes the integration of the HydroGeoSphere binary with the cloud computing workers, implementing the web-interface for running tasks, and integrating the OS for maintaining both the input and output of the HydroGeoSphere. All parts of the so far implemented infrastructure fully correspond to our requirements. Due to the integration with cloud infrastructures, simultaneous running of different models showed us significant profit in comparison with the usual sequential running. Also, the web interface for controlling the computations, greatly simplified the whole process of launching models.

The current infrastructure is under ongoing developments. In the future, we plan to strongly integrate all the technological pieces to allow for fully automated model computations thus providing valuable pumping scheme predictions in real time. We also plan to develop our web interface to allow for any generic computational use-case. One of the identified improvements relates to precise definition of input and output to support many different applications (e.g., by employing XML to define program options, input/output files, etc.).

[10] `http://www.djangoproject.com`
[11] `http://code.google.com/p/gc3pie/`
[12] `www.gc3.uzh.ch`

5 Conclusions

The integration of advanced Information Technologies in environmental simulation systems allows for a new dimension of natural resource management. Our proposed solution is especially interesting for remote locations with harsh environmental conditions in which wireless mesh network prove to provide a reliable network infrastructure. When the transporting infrastructure is developed, one can employ cloud computing to solve any computationally expensive problem, while the network monitoring application (e.g., Zabbix) can transport information in different use-cases such as environmental monitoring or smart-home applications.

Acknowledgements. This work is partially funded by the Swiss State Secretariat for Education and Research through SWITCH and the Swiss National Science Foundation through NRP 61 on Sustainable Water Management. We particularly thank Torsten Braun (the leader of the A4Mesh project) and his team as well as Sergio Mafioletti (the leader of the SwissACC project) for their contributions to this work.

References

1. Badawy, G., Sayegh, A., Todd, T.: Solar powered wlan mesh network provisioning for temporary deployments. In: Wireless Communications and Networking Conference, WCNC 2008, pp. 2271–2276. IEEE (March 2008)
2. Wu, D., Mohapatra, P.: Qurinet: A wide-area wireless mesh testbed for research and experimental evaluations. In: 2010 Second International Conference on Communication Systems and Networks (COMSNETS), pp. 1–10 (January 2010)
3. Jamakovic, A., Dimitrova, D.C., Anwander, M., Macicas, T., Braun, T., Schwanbeck, J., Staub, T., Nyffenegger, B.: Real-world energy measurements of a wireless mesh network. In: Pierson, J.-M., Da Costa, G., Dittmann, L. (eds.) EE-LSDS 2013. LNCS, vol. 8046, pp. 218–232. Springer, Heidelberg (2013)
4. Anwander, M., Braun, T., Jamakovic, A., Staub, T.: Authentication and authorisation mechanisms in support of secure access to wmn resources. In: 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1–6 (June 2012)
5. Schiller, E., Monakhov, A., Kropf, P.: Shibboleth based authentication, authorization, accounting and auditing in wireless mesh networks. In: LCN, pp. 918–926 (2011)
6. Kunszt, P., Maffioletti, S., Flanders, D., Eurich, M., Schiller, E., Bohnert, T., Edmonds, A., Stockinger, H., Jamakovic-Kapic, A., Haug, S., Flury, P., Leinen, S.: Towards a swiss national research infrastructure. In: Proceedings of the 1st International Workshop on Federative and Interoperable Cloud Infrastructures 2013, FedICI 2013 Organized in Conjunction with Euro-par (August 2013)
7. Partington, D., Brunner, P., Frei, S., Simmons, C.T., Werner, A.D., Therrien, R., Maier, H.R., Dandy, G.C., Fleckenstein, J.H.: Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. Water Resources Research 49(9), 5501–5519 (2013)

8. Brunner, P., Cook, P.G., Simmons, C.T.: Disconnected surface water and groundwater: from theory to practice. Ground Water 49(4), 460–467 (2011)
9. Winter, T.C., Harvey, J.W., Franke, O.L.: Alley, W.M.: Ground Water and Surface Water A Single Resource. USGS, Circular 1139, Denver, Colorado (1998)
10. Evensen, G.: Data assimilation: the ensemble Kalman filter, 2nd edn. Springer, New York (2009)
11. Hendricks Franssen, H.J., Kinzelbach, W.: Ensemble kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems. Journal of Hydrology 365(3-4), 261–274 (2009)
12. Brunner, P., Simmons, C.T.: Hydrogeosphere: A fully integrated, physically based hydrological model. Ground Water 50(2), 170–176 (2012)
13. Therrien, R., McLaren, R., Sudicky, E., Panday, S.: HydroGeoSphere. Groundwater Simulations Group (2013)

Attribute Reduction Based on Rough Sets and the Discrete Firefly Algorithm

Nguyen Cong Long[1], Phayung Meesad[1], and Herwig Unger[2]

[1] Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Thailand
nclong.c52@moet.edu.vn, pym@kmutnb.ac.th
[2] Faculty of Mathematics and Computer Science, FernUniversitat in Hagen, Germany
herwig.unger@fernuni-hagen.de

Abstract. Attribute reduction is used to allow elimination of redundant attributes while remaining full meaning of the original dataset. Rough sets have been used as attribute reduction techniques with much success. However, rough set applies to attribute reduction are inadequate at finding optimal reductions. This paper proposes an optimal attribute reduction strategy relying on rough sets and discrete firefly algorithm. To demonstrate the applicability and superiority of the proposed model, comparison between the proposed models with existing well-known methods is also investigated. The experiment results illustrate that performances of the proposed model when compared to other attribute reduction can provide comparative solutions efficiently.

Keywords: Attribute Reduction, Reduction, Feature selection, Rough sets, Core, Firefly Algorithm.

1 Introduction

Attribute reduction is a very important issue in many fields such as data mining, machine learning, pattern recognition and signal processing [1-4]. That is a process of choosing a subset of significant attributes (features) and elimination of the irrelevant attributes in a given dataset in order to build a good learning model. The subset is a retaining high accurate representation of original features and sufficient to describe target dataset.

Quite often, abundance of noisy, irrelevant or misleading features is usually presented in real-world problems. The ability to deal with imprecise and inconsistent information has become one of the most important requirements for attribute reduction. Rough sets theory can be utilized as a tool to discover data dependencies and reduce the number of attribute in inconsistent dataset [1]. Rough sets are applied to attribute reduction to remove redundant attributes and select sets of significant attributes which lead to better prediction accuracy and speed than systems using original sets of attributes.

Due to the NP-hard problem, designing an efficient algorithm for minimum attribute reduction is a challenging task [5]. There are many rough sets algorithms that

S. Boonkrong et al. (eds.), *Recent Advances in Information and Communication Technology*, Advances in Intelligent Systems and Computing 265,
DOI: 10.1007/978-3-319-06538-0_2,

have been proposed for attribute reduction in past literature. Generally, there are two categories of rough set methods for attribute reduction, greedy heuristics and meta-heuristic algorithms. The greedy heuristics approaches usually select the significant attributes or eliminate redundant attributes as heuristic knowledge. Algorithms in this category are fast. However, they usually find a reduction than a minimum reduction [6-9]. On the other hand, meta-heuristic algorithms have been applied to find minimal reductions [5][10-12]. In many systems that require minimal subset of attributes, the meta-heuristic algorithms are necessary to use.

There are many meta-heuristic algorithms that have been proposed to find minimal reductions based on rough sets. Wang, X. *et al* [11] proposed feature selection based on rough sets and particle swarm optimization (PSO). In those techniques, PSO is applied to find optimal feature selections. However, the fitness function applied in this algorithm may not correspond to a minimum reduction [5]. Inbaria, H. H. *et al* [12] proposed a hybrid model to combine the strength of rough sets and PSO to find a reduction. In this model, relying on two existing algorithms, quick reduct and relative reduct algorithm, PSO is applied to find optimal reductions. Nevertheless, the fitness function was only considered by correctness of attribute reduction without minimal reduction. As a result, reductions may not be minimal reductions that were found in this research. Ke, L. *et al* [10] investigated a model that combined rough sets and ant colony optimization (ACO) to find minimal attribute reduction. The experiment results shown that ACO applied to rough sets can provide competitive solutions efficiently. Ye, D. *et al* [5] proposed a novel fitness function for meta-heuristic algorithms based on rough sets. Genetic algorithm (GA) and PSO were applied to find the minimal attribute reduction using various fitness functions. The experiment results illustrated that PSO outperformed GA in terms of finding minimal attribute reduction. In addition, the novel fitness function was considered by correctness and minimal attribute reduction.

This paper proposes a new attribute reduction mechanism, which combine rough sets and the firefly algorithm to find minimal attribute reduction. Firefly algorithm (FA) is a new meta-heuristic algorithm that relies on flashing behavior of fireflies in nature to find global optimal solution in search space for special problems [13]. FA has been successfully applied to a large number of difficult combinational optimization problem [14-19]. Preliminary studies suggest that the FA outperforms GAs and PSO [13][19] in terms of accuracy and running time.

The remainder of this paper is organized as follows: Section 2 reviews rough sets preliminaries. The original firefly algorithm is summarized in section 3. In section 4 the attribute reduction using rough sets and the firefly algorithm is presented. Experiment results and comparison of differential models are discussed in section 5. Finally, conclusions are summarized in section 6.

2 Rough Sets Preliminaries

This section reviews some basic notions in the rough sets theory [1][20] which are necessary for this research.

2.1 Decision Table

Let a decision table $DT = (U, A = C \cup D)$, where $U = \{x_1, x_2, \ldots, x_n\}$ is non-empty finite set of objects called the universe of discourse, C is a non-empty finite set of condition attributes, D is a decision attribute.

2.2 Indiscerbility Relation

$\forall a \in A$ determine a function $f_a = U \to V_a$, where V_a is the set of values of a. if $P \subseteq A$, the *P-indiscernibility relation* is denoted by $IND(P)$, is defined as:

$$IND(P) = \{(x, y) \in U | \forall a \in P, f_a(x) = f_a(y)\} \tag{1}$$

The partition of U generated by $IND(P)$ is denoted by U/P. If $(x, y) \in IND(P)$, x and y are said to be indiscernibility with respect to P. The equivalence classes of the *P-indiscernibility relation* are denoted by $[x]_P$. The indiscernibility relation is the mathematical basic notion of rough sets theory.

$$\text{Let } U/C = \{Y_1, Y_2, \ldots, Y_N\}, N \leq n \tag{2}$$

where the equivalence classes Y_i are numbered such that $|Y_1| \leq |Y_2| \leq \cdots \leq |Y_N|$

2.3 Lower and Upper Approximation

Let $X \subseteq U$ and $P \subseteq A$, X could be approximated by the lower and upper approximation. P-lower and P-upper approximation of set X, is denoted by $\underline{P}$ ~~ and $\overline{P}X$, respectively, is defined as:

$$\underline{P}X = \{x \in U : [x]_P \subseteq X\} \tag{3}$$

$$\overline{P}X = \{x \in U : [x]_P \cap X \neq \emptyset\} \tag{4}$$

2.4 Positive, Negative and Boundary Region

Let $P, Q \subseteq A$, be equivalence relations over U, then the positive, negative and boundary regions, denoted $POS_P(Q)$, $NEG_P(Q)$, $BN_P(Q)$, respectively, can be defined as:

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X \tag{5}$$

$$NEG_P(Q) = U - \bigcup_{X \in U/Q} \overline{P}X \tag{6}$$

$$BN_P(X) = \bigcup_{X \in U/Q} \overline{P}X - \bigcup_{X \in U/Q} \underline{P}X \tag{7}$$

A set is said to be rough (imprecise) if its boundary region is non-empty, otherwise the set is crisp.

2.5 Dependency of Attributes

Let $P \subseteq C$, D depends on P in a degree k $(0 \leq k \leq 1)$ denoted by $P \Rightarrow_k D$, is determined by

$$k = \gamma_P(D) = \frac{|POS_P(D)|}{|U|} \tag{8}$$

where $|.|$ denotes the cardinality of a set, $\gamma_P(D)$ is quality of classification. If k=1, D depends totally on P; if $0 < k < 1$, D depends partially on P, if k=0, D is not depends on P. Decision table DT is consistent if $\gamma_C(D) = 1$, otherwise DT is inconsistent.

2.6 Attribute Reduction and Core

Generally, there are often existing redundant condition attributes. So, these redundant attributes can be eliminated without losing essential classificatory information [20]. The goal of attribute reduction is to eliminate redundant attributes leading to the reduced set that provides the same quality of classification as the original.

A given decision table may have many attribute reductions, the set of all reductions are defined as

$$Red(C) = \{R \subseteq C | \gamma_R(D) = \gamma_C(D), \forall \mathrm{B} \subset \mathrm{R}, \gamma_{\mathrm{B}}(D) \neq \gamma_{\mathrm{R}}(D)\} \tag{9}$$

A set of minimal reductions is defined as

$$Red(C)_{min} = \{R \in Red(\mathrm{C}) | \forall R' \in Red(C), |R| \leq |R'|\} \tag{10}$$

Core of condition attributes is an intersection of all reductions, defined as

$$Core(C) = \bigcap Red(æ) \tag{11}$$

3 Original Firefly Algorithm

The Firefly algorithm is a kind of stochastic, meta-heuristic algorithm to find the global optimal solution in search space for special problems. This is inspired by the flashing behavior of fireflies in nature and is originally proposed by Yang, X.S. in 2008 and relies on three key ideas [13]

- All fireflies are unisex and there may be an attractive in any two fireflies
- Their attractiveness is proportional to their light intensity. A firefly with lower light intensity will move toward the fireflies with higher light intensity. If there is not firefly with higher light intensity, the firefly will randomly move in search space.
- The light intensity of a firefly is related to fitness function in genetic algorithms.

FA starts with randomly generated positions of fireflies (population). In each time step t, positions of fireflies with lower light intensity move toward fireflies with higher light intensity by Eqs. (12) and (13). That mean, for any two fireflies, if a

firefly has lower light intensity, it will move toward the other firefly in the search space.

$$X_i(t+1) = X_i(t) + \beta\left(X\ \ (t) - X_j(t)\right) + \alpha\left(rand - \frac{1}{2}\right) \tag{12}$$

$$\beta = \beta_0 e^{-\gamma r_{ij}^2} \tag{13}$$

where $X_i(t)$ and X_j (t) are positions of firefly with lower light intensity and firefly with higher intensity at time t respectively, α is a random parameter which determines randomly behavior of movement, rand is a random number generator uniformly distributed in [0, 1], γ is a light absorption coefficient, β_0 is the attractiveness at $r = 0$, and r_{ij} is Euclidean distance between any two fireflies i and j at X_i and X_j, respectively.

After movements, all fireflies move toward the firefly with the highest light intensity and their light intensity improves. After stopping criteria are satisfied, the firefly with the highest light intensity will be considered as the best solution.

More details of firefly algorithm can see in [13].

4 Firefly Algorithm for Attribute Reduction

The idea of firefly algorithm is used to find minimal attribute reduction problem. The process of firefly algorithm is to find a minimal attribute reduction, is illustrated in Fig. 1.

4.1 Encoding Method

The position of each firefly represents the possible solution of the problem as binary strings of length m ($m = |C|$). Every bit represents an attribute, the value '1' shows that the corresponding attribute is selected while the value '0' illustrates that the corresponding attribute is not selected.

For example, suppose that $C = \{a_1, a_2, \ldots, a_{10}\}$ and a firefly $X = 1010001101$, then an attribute subset is $\{a_1, a_3, a_7, a_8, a_{10}\}$.

4.2 Fitness Function

There are many definitions of fitness function for this problem in past literature [5][11]. There are some drawbacks with these fitness functions. These may considered by correctness of attribute reduction without minimal reductions [5]. The fitness function used in this research is defined as Eq. (14) [5]

$$Fitness(X) = \frac{m-|X|}{m} + \frac{n|R|\,\gamma_X(\mathrm{D})}{m\Gamma} \tag{14}$$

where $m = |C|, n = |U|,\ \gamma_X(\mathrm{D})$ is quality of classification. R is a reduct of condition attribute C, R is computed by an efficient attribute reduction algorithm in [21].

$\Gamma = |Y_1| + |Y_2|$ if the decision table DT is consistent and $\Gamma = |Y_1|$ if not where Y_1 and Y_2 are defined in Eq. (2) .

This fitness function not only considers quality of classification but also considers minimal reductions [5].

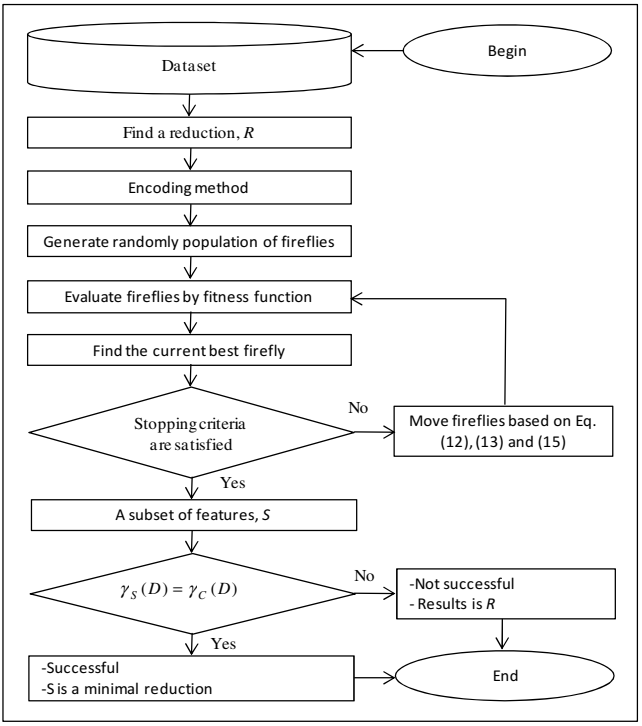

Fig. 1. A process of firefly algorithm applies to find a minimal attribute reduction

4.3 Position Update Strategies

In the original firefly algorithm, position of fireflies with lower light intensity will move towards fireflies with higher light intensity. Those fireflies will change from binary number to a real number when they move in the search spaces. Therefore, this real number must be replaced by a binary number. For this purpose, a sigmoid function can be applied [22]. However, in order to improve the binary firefly algorithm, a tang function is used in this research [23]. The tang function is defined in Eq. (15) as:

$$f\left(X_i^k\right) = \frac{\exp\left(2X_i^k\right)-1}{\exp\left(2X_i^k\right)+1}, i = 1, \ldots, N; k = 1, \ldots, d \tag{15}$$

This function scale the X_i value in the [0, 1] range. The final value of each part of fireflies after movement is determined by: If $f\left(X_i^k\right) \geq rand$ then $X_i^k = 1$ otherwise $X_i^k = 0$. $rand$ is a random number generator uniformly distributed in [0, 1]

5 Simulations

To evaluate the proposed model, attribute reduction based on rough sets and discrete firefly algorithm (ARRSFA), various datasets with different numbers of condition attributes and objects are used to test the model. In addition, in order to demonstrate the superiority of proposed model, the comparison between the model and genetic algorithm for rough set attribute reduction (GenRSAR) [5] and particle swarm optimization for rough set attribute reduction (PSOAR) [5] is also investigated.

5.1 Data Sets

In this paper, 6 well-known datasets are collected from UCI Repository Machine Learning Database those are used to test the models. Most of these datasets are used for evaluating attribute reduction algorithms in the past literature [5][10-11]. Basic information about datasets is shown in Table 1.

5.2 Parameters Setting

The proposed model and other attribute reduction models are implemented in MATLAB. In the experiments, the parameters, except when indicated differently, were set to the following values: Initially, 25 fireflies are randomly generated in a population, $\alpha = 0.2, \gamma = 1, \beta_0 = 0.2$, number of generations is equal to 100. The FA MATLAB code for each dataset ran 100 times with different initial solutions, same as [5].

Table 1. Basic information about datasets

No.	Datasets name	Number of objects (n)	Number of condition attributes (m)	Number of classifications
1	Audiology	200	69	10
2	Bupa	345	6	2
3	Corral	32	6	2
4	Lymphography	148	18	4
5	Soybean-small	47	35	4
6	Vote	300	16	2

5.3 Results and Discussion

A number of results from the experiments are recorded, consisting of minimal (Min) and average (AVG) length of output attribute reduction during 100 runs of the algorithms. The fitness function uses in this research not only considers quality of

classification but also considers minimal reductions [5]. Therefore, results of minimal and average length of attribute reduction during 100 runs are adequate for simulation. The experiment results are illustrated in Table 2 and Fig. 2.

Table 2. Length of minimal and average attribute reduction of differential models

Datasets name	GenRSAR		PSOAR		ARRSFA	
	Min	AVG	Min	AVG	Min	AVG
Audiology	12	14.35	12	14.32	11	11.5
Bupa	3	3	3	3	3	3
Corral	4	4.04	4	4.02	4	4.02
Lymphography	5	5.66	5	5.6	6	6.6
Soybean-small	2	2.9	2	2.24	2	2.1
Vote	8	8.28	8	8.2	5	5.63

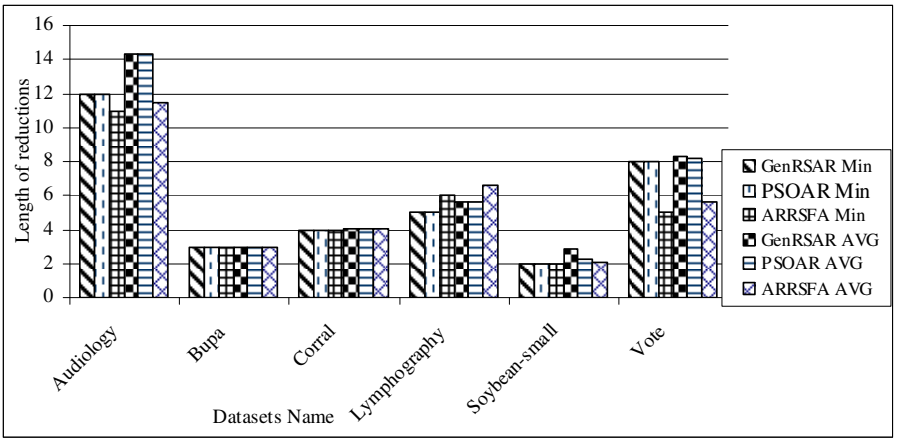

Fig. 2. Comparison results of different models

Looking at the Table 2 and Fig 2, it can be seen that all of the algorithms could find minimal reductions of the Bupa dataset with the same length. ARRSFA and PSOAR had the same results in terms of min and average of attribute reduction length and they outperform GenRSAR in terms of average of attribute reduction length of the Corral dataset. In addition, ARRSFA could find the best minimal reductions for the all of other tested datasets except for Lymphography. ARRSFA outperforms both GenRSAR and PSOAR in 3 datasets namely Audiology, Soybean-small and Vote. However, ARRSFA sometimes could not obtain the best solutions same as the other models. It is not better when compared to the other methods for the Lymphography dataset. There is no single model that always find the best solution for all data sets, but ARRSFA outperforms in terms of obtaining better solutions for such datasets as Vote and Audiology.

6 Conclusion and Future Work

This paper proposed a model to find minimal reductions based on rough sets and the discrete firefly algorithm. In this model, rough sets are used to build fitness function of the firefly algorithm as well as verifying correctness of reductions. The discrete firefly algorithm is applied to find minimal reductions. To demonstrate the superiority of the proposed model, numerical experiments have been conducted on 6 well-known datasets. Comparisons of performance with the proposed model and two meta-heuristic algorithms have revealed that the proposed model has a superior performance.

The proposed model is only tested on 6 differential datasets. Further investigation will concentrate on two other aspects, namely running time and classification accuracy. Furthermore, there are existing several extended types of attribute reduction in the concepts of rough sets such as entrope-based reducts [24], distribution reducts [25]. These extension may assist to improve the performance of the proposed model. In conclusion, all the future work will contribute to further improve the proposed model, making it a more robust technique for attribute reduction.

References

1. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Springer (1991)
2. Cheng, C.H., Chen, T.L., Wei, L.Y.: A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. Inf. Sci. 180(9), 1610–1629 (2010)
3. Chen, H.L., Yang, B., Liu, J., Liu, D.Y.: A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Syst. Appl. 38(7), 9014–9022 (2011)
4. Zhao, S., Tsang, E.C.C., Chen, D., Wang, X.: Building a Rule-Based Classifier, A Fuzzy-Rough Set Approach. IEEE Trans. Knowl. Data Eng. 22(5), 624–638 (2010)
5. Ye, D., Chen, Z., Ma, S.: A novel and better fitness evaluation for rough set based minimum attribute reduction problem. Inf. Sci. 222, 413–423 (2013)
6. Hoa, N.S.: Some Efficient Algorithms For Rough Set Methods. In: Proceedings IPMU 1996 Granada, Spain, pp. 1541–1457 (1996)
7. Degang, C., Changzhong, W., Qinghua, H.: A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets. Inf. Sci. 177(17), 3500–3518 (2007)
8. Wang, C., He, Q., Chen, D., Hu, Q.: A novel method for attribute reduction of covering decision systems. Inf. Sci. 254, 181–196 (2014)
9. Meng, Z., Shi, Z.: A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets. Inf. Sci. 179(16), 2774–2793 (2009)
10. Ke, L., Feng, Z., Ren, Z.: An efficient ant colony optimization approach to attribute reduction in rough set theory. Pattern Recognit. Lett. 29(9), 1351–1357 (2008)
11. Wang, X., Yang, J., Teng, X., Xia, W., Jensen, R.: Feature selection based on rough sets and particle swarm optimization. Pattern Recognit. Lett. 28(4), 459–471 (2007)
12. Inbarani, H.H., Azar, A.T., Jothi, G.: Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. Comput. Methods Programs Biomed. 113(1), 175–185 (2014)

13. Yang, X.-S.: Firefly algorithms for multimodal optimization. In: Watanabe, O., Zeugmann, T. (eds.) SAGA 2009. LNCS, vol. 5792, pp. 169–178. Springer, Heidelberg (2009)
14. Fister, I., Fister Jr., I., Yang, X.S., Brest, J.: A comprehensive review of firefly algorithms. Swarm Evol. Comput., 34–46 (2013)
15. Luthra, J., Pal, S.K.: A hybrid Firefly Algorithm using genetic operators for the cryptanalysis of a monoalphabetic substitution cipher. In: 2011 World Congress on Information and Communication Technologies (WICT), pp. 202–206 (2011)
16. Mohammadi, S., Mozafari, B., Solimani, S., Niknam, T.: An Adaptive Modified Firefly Optimisation Algorithm based on Hong's Point Estimate Method to optimal operation management in a microgrid with consideration of uncertainties. Energy 51, 339–348 (2013)
17. Dos, L., Coelho, S., Mariani, V.C.: Firefly algorithm approach based on chaotic Tinkerbell map applied to multivariable PID controller tuning. Comput. Math. Appl. 64(8), 2371–2382 (2012)
18. Kazem, A., Sharifi, E., Hussain, F.K., Saberi, M., Hussain, O.K.: Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Appl. Soft Comput. 13(2), 947–958 (2013)
19. Long, N.C., Meesad, P.: Meta-heuristic algorithms applied to the optimization of type-1 and type 2 TSK fuzzy logic systems for sea water level prediction. In: 2013 IEEE Sixth International Workshop on Computational Intelligence Applications (IWCIA), Hiroshima, Japan, pp. 69–74 (2013)
20. Pawlak, Z.: Rough set approach to knowledge-based decision support. Eur. J. Oper. Res. 99(1), 48–57 (1997)
21. Shi, Z., Liu, S., Zheng, Z.: Efficient Attribute Reduction Algorithm. In: Bramer, M., Devedzic, V. (eds.) Artificial Intelligence Applications and Innovations. IFIP AICT, vol. 154, pp. 211–222. Springer, Heidelberg (2004)
22. Sayadi, M.K., Hafezalkotob, A., Naini, S.G.J.: Firefly-inspired algorithm for discrete optimization problems: An application to manufacturing cell formation. J. Manuf. Syst. 32(1), 78–84 (2013)
23. Chandrasekaran, K., Simon, S.P., Padhy, N.P.: Binary real coded firefly algorithm for solving unit commitment problem. Inf. Sci. 249, 67–84 (2013)
24. Kryszkiewicz, M.: Comparative studies of alternative type of knowledge reduction in inconsistent systems. Int. J. Intell. Syst. 16, 105–120 (2001)
25. Wang, G.: Rough reduction in algebra view and information view. Int. J. Intell. Syst. 18(6), 679–688 (2003)

A New Clustering Algorithm Based on Chameleon Army Strategy

Nadjet Kamel[1,2] and Rafik Boucheta[2]

[1] Univ-Setif, Fac-Sciences, Depart. Computer Science, Setif, Algeria
[2] USTHB, LRIA, Algiers, Algeria
nkamel@usthb.dz, rafik911@yahoo.fr

Abstract. In this paper we present a new clustering algorithm based on a new heuristic we call Chameleon Army. This heuristic simulates a Army stratagem and Chameleon behavior. The proposed algorithm is implemented and tested on well known dataset. The obtained results are compared to those of the algorithms K-means, PSO, and PSO-kmeans. The results show that the proposed algorithm gives better clusters.

Keywords: Clustering algorithm, K-means, PSO, metaheuristic.

1 Introduction

Clustering is grouping objects such that similar objects are within a same group, and dissimilar objects are in different groups. The main problem of clustering is to obtain optimal grouping. This issue arises in many scientific applications, such as biology, education, genetics, criminology, etc… Several approaches [1], [2], [3], [4], [5], [6], [7] have been developed in this regard.

Many clustering methods have been proposed, and they are classified into major algorithms classes: hierarchical clustering, partitioning clustering, density based clustering and graph based clustering.

In this paper we propose a new clustering algorithm based on a new heuristic we call Chameleon Army. This heuristic simulates an Army stratagem and some Chameleon behavior.

The efficiency of the proposed algorithm is tested on different datasets issued from literature [8]. The obtained results are compared with those of the algorithms kmeans [9], PSO [10], and PSO-kmeans.

The remaining of the paper is organized as follows: the next section presents the related works. The section 3 presents the new algorithm and its implementation. The results of our algorithm are presented in section 4. Finally, section 5 presents our conclusion and future works.

2 Related Works

Many clustering algorithms are defined in the literature. In general, they are classed into two classes: partitioned algorithms, and hierarchical algorithms. The kmeans

S. Boonkrong et al. (eds.), *Recent Advances in Information and Communication Technology*, Advances in Intelligent Systems and Computing 265,
DOI: 10.1007/978-3-319-06538-0_3,