Active constraints selection based on density peak

Viet-Vu Vu*[1], Byeongnam Yoon**, Hong-Quan Do*, Hai-Minh Nguyen***, Tran-Chung Dao***, Cong-Mau Tran****, Doan-Vinh Tran*****, Thi-Nhuong Phi******, Viet-Thang Vu*******, Tien-Dung Duong********

*VNU Information Technology Institute, Vietnam National University, Hanoi, Vietnam
**Global IT Research Institute GIRI
***ICTU – University of Information and Communication Technology, Thai Nguyen University, Vietnam
****Ha Tinh University, Ha Tinh City, Vietnam
*****University of Education, Vietnam National University, Hanoi, Vietnam
****** People's Security Academy, Hanoi, Vietnam
*******Hanoi University of Industry, Hanoi, Vietnam
********Vietnam Academy of science and technology, Hanoi, Vietnam
[1]Corresponding Author: vuvietvu@vnu.edu.vn

Abstract—**Semi-supervised clustering which aims to integrate side information to improve the performance of clustering process, has received a lot of attentions in research community. Generally, there are two kinds of side information called seed (labelled data) and constraint (must-link, cannot-link). By integrating information provided by the user or domain expert, the semi-supervised clustering can produce expected results of users. In fact, clustering results usually depend on side information provided, so different side information will produce different results. In some cases, the performance of clustering may decrease if the side information is not carefully chosen. This paper addresses the problem of selecting good constraints for semi-supervised clustering algorithms. For this purpose, we propose an active learning algorithm for the constraints collection task, which relies on the min-max algorithm and peaks estimation based on density score. Experiments conducted on some real data sets from UCI show the effectiveness of our approach.**

Keyword—**Clustering, Semi-supervised clustering, Constraint, Density peak, Active Learning.**

Assoc. Prof. Dr. Viet-Vu Vu received the B.S. degree in Computer Science from Hanoi University of Education in 2000, a M.S. degree in Computer Science from Hanoi University of Technology in 2004, and a Doctor Degree in Computer Science from Paris 6 University in 2011. He is a researcher at Information Technology Institute, Vietnam National University, Hanoi. His research interests include clustering, active learning, semi-supervised clustering, and E-government applications

Byeongnam Yoon (M'97) He became a Member of IEEE in 1997. He was born in Seoul Korea 1949. He got the PhD in computer science, Chungnam National University, Korea, 1997. He worked for the Sperryrand UNIVAC as a Computer Specialist 1974 -1978, Samsung as a Manager of Telecommunications Section 1978 - 1982, Electronics & Telecommunications Research Institute (ETRI) as a Principal Researcher 1982 – 1999, National Information society Agency (NIA) as a Senior Executive Director General 1999 – 2010, Kyonggi University (KGU) as an Associate Professor, Faculty of Computer Science 2010 - 2016. He is an invited professor at Information Technology Institute, Vietnam National University, Hanoi Vietnam since 2017, and Global IT Research Institute GIRI) as a President 1999 – current. His research area includes Telecommunications, Internet, Software, Web programming & security, e-Government, Enterprise Architecture, Workflow, BPM. System Work Method. SPiCE, CMMI, etc.

MSc. Hong-Quan Do received a double M.S. degree in Information and Communication Technology from University of Science and Technology of Hanoi, Vietnam and The University of Rennes 1, France in 2015. He is a researcher at Information Technology Institute, Vietnam National University, Hanoi. His research concentrates primarily on Clustering, Semi-supervised Clustering, Image processing and Recommender Systems. At the present, he has also been involved in some E-Government projects, and E-Commerce Recommendation applications.

Dr. Nguyen Hai Minh received a PhD degree in Computer science and Engineering from Kyungpook National University, Korea in 2014. He is a dean of Faculty of Information Technology - ICTU. His study focuses on Health Informatics, Health IT Standards and Medical Image Processing.

M.IT Tran-Chung Dao received a M.IT degree in Information Technology from Manuel S. Enverga University Foundation, Philippines 2013. He is a lecturer at Thai Nguyen University Information and communication technology, Thai Nguyen, Viet Nam. His research concentrates primarily on Clustering and Image processing.

M.S Cong-Mau Tran received the B.S. degree in Computer Science from Hanoi University of Education in 2000, a M.S. degree in Computer Science from University of Technology – Vietnam National University, Hanoi in 2011. Now, he is a lecturer at the Faculty of Engineering – Technology, Ha Tinh University, Vietnam. His research interested include data mining, clustering, image processing and its applications.

Assoc. Prof. Dr. Doan-Vinh Tran received a PhD degree in Informatics Education from the Russian Academy of Educational Sciences, Russia in 1997. Now, he is a lecturer at the Faculty of Educational Technology, University of Education, Vietnam National University, Hanoi. His research concentrates primarily on Informatics Education, Digital Education, applications VR/AR/MR/XR in Digital Education, Clustering and Image processing.

Mrs. Thi-Nhuong Phi received the B.S. degree in Electronics and Telecommunications Engineering from Thai Nguyen University in 2007, a M.S. degree in Electronics - Telecommunications Technology from College of Technology Confers – Vietnam national university, Hanoi in 2010. She is a teacher at People's Security Academy. Her research interests include technology, engineering, and information security.

Dr. Viet-Thang Vu received the B.S. degree in Computer Science from Mendeleev University of Chemical Technology of Russia in 2008, a M.S. degree in Computer Science from Le Quy Don Technical University in 2012, and a Doctor Degree in Computer Science from Moscow Institute of Physics and Technology in 2019. He is a teacher at Information Technology faculty, Hanoi university of industry. His research interest is in the area of: clustering, computer vision, network security.

MSc. Dzung-Tien Duong received a M.S. degree in Information and Computing Technology from Peter the Great St.Petersburg Polytechnic University, Saint Peterburg, Russia in 2013. Now, he is a researcher at VietNam Academy of science and technology, Hanoi, Vietnam. His research concentrates primarily on Clustering, Semi-supervised Clustering, semi-super fuzzy clustering.