

Patient Profiling Using Evolutionary Clustering

Veselka Boeva
veselka.boeva@bth.se

Blekinge Institute of Technology
Karlskrona, Sweden

Vishnu Manasa Devagiri
vishnu.manasa.devagiri@bth.se
Blekinge Institute of Technology
Karlskrona, Sweden

Milena Angelova
mangelova@tu-plovdiv.bg
Technical University of Sofia
Plovdiv, Bulgaria

Elena Tsiporkova
elena.tsiporkova@sirris.be
EluciDATA Lab, Sirris
Brussels, Belgium

ABSTRACT

The selection of a personalized treatment plan for a patient can be of significant importance for his/her health and the treatment outcome. For example, a decision support system that can associate patient-specific disease characteristics with the patient treatment response can provide the physician with a better understanding of the personalized conditions of every single patient. In this study, we propose a split-merge evolutionary clustering algorithm that can be used for efficient maintenance of patient profiles in healthcare domain. For example, profiling of individuals with similar anthropometric features with a purpose to identify the person's increased risk for cardiovascular disease. The patients who share the same profile should probably have a similar predisposition and should be provided similar healthcare recommendations. As more individuals get involved one needs to re-cluster the initial clusters (profiles) and also assign new incoming individuals to the existing clusters. However, the existing original clusters can become outdated caused by appearing of patients with new disease characteristics due to different factors. This outdated of models is in fact, a concept drift and requires that the clustering techniques, used for deriving the original patient profiles, can deal with such a concept drift and enable reliable and scalable model update. The proposed split-merge evolutionary clustering technique is designed to be robust to such concept drift scenarios by providing the flexibility to compute clusters on a new portion of data and to update the existing clustering solution by the computed new one.

CCS CONCEPTS

• **Computing methodologies** → **Cluster analysis**; • **Applied computing** → **Health care information systems**.

KEYWORDS

evolutionary clustering, unsupervised learning, healthcare, patient profiling

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

womENCourage '19, September 16–18, 2019, Rome, Italy

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

ACM Reference Format:

Veselka Boeva, Milena Angelova, Vishnu Manasa Devagiri, and Elena Tsiporkova. 2019. Patient Profiling Using Evolutionary Clustering. In *womENCourage '19: ACM Celebration of Women in Computing, September 16–18, 2019, Rome, Italy*. 1 page. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 ALGORITHM OVERVIEW

The proposed split-merge evolutionary clustering algorithm updates the existing clustering solution in order to make it adaptable to newly arriving data [2]. It is based on a bipartite graph that consists of existing clustering solution on one side and clustering of newly arriving data on the other side. The edges of the graph represent the correlation between the two clustering solutions and are used to decide whether a cluster node needs to be split or merged. If a node from the existing clustering is connected to more than one node of the new clusters, then it is split among these new clusters. Similarly, if two or more nodes from the existing clusters are connected to a single node from the new clustering, all these nodes are merged together. The algorithm has been evaluated on patient anthropometric measurements profiling for studying hypertension.

2 STATE OF THE ART ALGORITHMS

Evolutionary clustering solutions have been proposed by many authors. Two interesting works are PivotBiCluster algorithm [1], and dynamic split-merge approach [3]. Our initial experimental results show that the proposed split-merge algorithm has fared well against PivotBiCluster in most studied cases. We plan to further study and compare the proposed split-merge algorithm with the dynamic split-merge approach proposed by Lughofer in [3].

ACKNOWLEDGMENTS

This work is part of the research project "Scalable resource efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.

REFERENCES

- [1] N. Ailon, N. Avidor-Elgrabli, E. Liberty, and A. van Zuylen. 2011. Improved Approximation Algorithms for Bipartite Correlation Clustering. In *Algorithms - ESA 2011 - 19th Annual European Symposium, Saarbrücken, Germany, September 5-9, 2011. Proceedings*. 25–36.
- [2] Veselka Boeva, Milena Angelova, and Elena Tsiporkova. 2019. A Split-Merge Evolutionary Clustering Algorithm. 337–346. <https://doi.org/10.5220/0007573103370346>
- [3] Edwin Lughofer. 2012. A dynamic split-and-merge approach for evolving cluster models. *Evolving Systems* 3 (2012), 135–151. Issue 3.