

Accepted Manuscript

Efficient In-Database Patient Similarity Analysis for Personalized Medical Decision Support Systems

Araek Tashkandi, Ingmar Wiese, Lena Wiese

PII: S2214-5796(17)30308-8
DOI: <https://doi.org/10.1016/j.bdr.2018.05.001>
Reference: BDR 96

To appear in: *Big Data Research*

Received date: 15 October 2017
Revised date: 8 March 2018
Accepted date: 3 May 2018

Please cite this article in press as: A. Tashkandi et al., Efficient In-Database Patient Similarity Analysis for Personalized Medical Decision Support Systems, *Big Data Res.* (2018), <https://doi.org/10.1016/j.bdr.2018.05.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Efficient In-Database Patient Similarity Analysis for Personalized Medical Decision Support Systems

Araek Tashkandi^{a,1,*}, Ingmar Wiese^a, Lena Wiese^a

^a*Georg-August-University Goettingen, Institute of Computer Science, Goldschmidtstr. 7, 37077 Goettingen, Germany.*

Abstract

Patient similarity analysis is a precondition to apply machine learning technology on medical data. In this sense, patient similarity analysis harnesses the information wealth of electronic medical records (EMRs) to support medical decision making. A pairwise similarity computation can be used as the basis for personalized health prediction. With n patients the amount of $\binom{n}{2}$ similarity calculations is required. Thus, analyzing patient similarity leads to data explosion when exploiting big data. By increasing the data size the computational burden of this analysis increases. A real-life medical application may exceed the limits of current hardware in a fairly short amount of time. Finding ways to optimize patient similarity analysis and handling this data explosion is the topic of this paper.

Current implementations for patient similarity analysis require their users to have knowledge of complex data analysis tools. Moreover, data pre-processing and analysis are performed in synthetic conditions: the data are extracted from the EMR database and then the data preparation and analysis are processed in external tools. After all of this effort the users might not experience a superior performance of the patient similarity analysis. We propose methods to optimize the patient similarity analysis in order to make it scalable to big data. Our method was tested against two real datasets and a low execution time was accomplished. Our result hence benefits a comprehensive medical decision support system. Moreover, our implementation comprises a balance between performance and applicability: the majority of the workload is processed within a database management system to enable a direct implementation on an EMR database.

Keywords: Patient Similarity Analysis, Medical Data Analysis, Medical Decision Support System, Personalized Medicine, In-Database Analytics, Column Stores

1. Introduction

The increasing amount and scale of available electronic medical records (EMRs) raises the interest to gain valuable knowledge from them. Knowledge discovered from EMRs can favor patients

*Corresponding author

Email addresses: araek.tashkandi@cs.uni-goettingen.de (Araek Tashkandi), asatashkandi@kau.edu.sa (Araek Tashkandi), ingmar.wiese@stud.uni-goettingen.de (Ingmar Wiese), wiese@cs.uni-goettingen.de (Lena Wiese)

¹Also affiliated to King Abdulaziz University, Faculty of Computing and Information Technology, 21589 Jeddah, Kingdom of Saudi Arabia.

Download English Version:

<https://daneshyari.com/en/article/10225732>

Download Persian Version:

<https://daneshyari.com/article/10225732>

[Daneshyari.com](https://daneshyari.com)