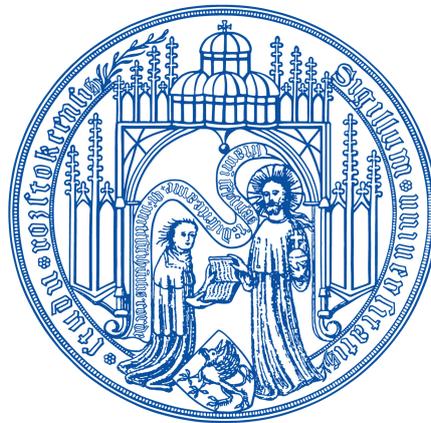

Big Data Analytics für die effiziente Aktivitätserkennung und -vorhersage in Assistenzsystemen

Dissertation
zur
Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)
der Fakultät für Informatik und Elektrotechnik
der Universität Rostock



vorgelegt von
Dennis Marten, geb. am 04.05.1988 in Berlin
aus Rostock

Rostock, den 14. Juni 2021

Zusammenfassung

Die kontinuierlichen technischen Fortschritte der vergangenen Jahre haben das enorme Potenzial datengetriebener Anwendungen, wie etwa Assistenzsysteme, aufgezeigt. Aufgrund enormer Datenmengen ist zur Analyse und Verarbeitung dieser oftmals eine massive Parallelisierung nötig. Entgegen dem bestehenden Trend spezifischer Neuentwicklungen werden in parallelen Datenbanksystemen bereits seit mehreren Jahrzehnten transparente parallele Verarbeitungen großer Datenmengen unterstützt und optimiert. Daraus motiviert wird in dieser Arbeit untersucht, inwiefern parallele relationale Datenbanksysteme für Methoden der Aktivitätserkennung und -vorhersage in Assistenzsystemen gewinnbringend eingesetzt werden können. Der Fokus liegt hierbei auf der effizienten und skalierbaren Umsetzung und Komposition von Basisoperatoren der linearen Algebra. Dies ermöglicht neben der Umsetzung zugehöriger Machine-Learning-Verfahren die Einbeziehung zahlreicher weiterer Methoden des wissenschaftlichen Rechnens. Für die potenzielle Umsetzung solcher werden daher zahlreiche Aspekte diskutiert und experimentell ausgewertet. Hierzu zählen unter anderem die Wahl geeigneter Datenbanksysteme, Datenrepräsentationen und zugehörige Übersetzungen in SQL und der Einfluss von Indexstrukturen. Für die effiziente Intraoperatorparallelisierung wird zudem eine Anfragetechnik vorgestellt, die die Verarbeitung einzelner Lineare-Algebra-Operatoren deutlich beschleunigen kann. Zur Einordnung der Güte des datenbankinternen Ansatzes werden Laufzeitvergleiche einzelner Operatoren und ganzer Methoden mit etablierten Lineare-Algebra-Bibliotheken und Big-Data-Ansätzen präsentiert und analysiert.

Continuous technological advances in recent years have highlighted the potential of data-driven applications such as assistance systems. Due to enormous amounts of data, massive parallelization is often required in order to process such applications in time. Contrary to the existing trend of specific new developments, transparent parallel processing of large amounts of data has already been supported and optimized in parallel database systems for several decades. Motivated by this, the present thesis investigates to what extent parallel relational database systems can be profitably used for activity recognition and prediction methods in assistance systems. The focus is on efficient and scalable implementations of basic linear algebra operators and their composition. In addition to the realization of associated machine learning methods, this enables the inclusion of numerous other methods of scientific computing. For the potential implementation of such, numerous aspects are therefore discussed and experimentally evaluated. These include the choice of suitable database systems, data representations and associated translations in SQL, as well as the influence of index structures. For efficient intraoperator parallelization, a query technique is presented that can significantly speed up the processing of single linear algebra operators. To prove the usability of the intra-database approach, runtime comparisons of individual operators and entire methods with established linear algebra libraries and big data environments are presented and analyzed.