University of Zurich[UZH]

Department of Informatics

University of Zurich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zurich

Robert Jan Stucki

Switzerland

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, 1. Oktober 2012

SQL Implementation of QR Formalization and Householder Transformation

Informatik-Vertiefung (3 ECTS):

Work overview:

The aim of this Vertiefung is to investigate and implement the Singular Value Decomposition (SVD) method using SQL queries. SVD is a matrix decomposition method that decomposes a matrix $\mathbf{V}$ into three matrices $\mathbf{L}$, $\mathbf{\Sigma}$ and $\mathbf{R}^T$. The product of the three matrices is equal to $\mathbf{V}$. The SVD method is performed by using the Householder transformation and the QR factorization. Formally, a matrix $\mathbf{V} = [V_1|V_2|\ldots|V_n] \in \mathcal{R}^{m\times n}$ can be decomposed into a product of three matrices as follows: $SVD(\mathbf{V}) = \mathbf{L} \times \mathbf{\Sigma} \times \mathbf{R}^T$

Where:

- $\mathbf{\Sigma}$: is a $n \times n$ square diagonal matrix that contains strictly positive singular values of $\mathbf{V}$. The diagonal entries (σ_i) of $\mathbf{\Sigma}$ are the square root of the eigen values of $\mathbf{V}^T\mathbf{V}$ and are ranked in the decreasing order such that $\sigma_1 > \sigma_2 > \ldots > \sigma_n$.
- $\mathbf{L}$: is an $m \times n$ orthogonal matrix having as columns orthonormal eigen vectors of $\mathbf{V}\mathbf{V}^T$ ($L^TL = I$, where I is the identity matrix). The eigen vectors of L are computed by solving $Det(\sigma\mathbf{I} - \mathbf{V}\mathbf{V}^T) = 0$ where $Det(\mathbf{X})$ is the determinant of matrix $\mathbf{X}$.
- $\mathbf{R}$: is an $n \times n$ orthogonal matrix having as columns orthonormal eigen vectors of $\mathbf{V}^T\mathbf{V}$ ($R^TR = I$). The eigen vectors of $\mathbf{R}$ are computed by solving $Det(\sigma\mathbf{I} - \mathbf{V}^T\mathbf{V}) = 0$.

Work tasks:

1. Understand and implement the QR formalization and the Householder transformation that perform SVD using PL/SQL
2. Evaluate the scalability of the implemented decomposition on an Oracle server: horatio.ifi.uzh.ch. The test datasets are already loaded on the server.

3. Empirical comparison of the running time between the SQL implementation and a main memory implementation provided by the supervisor.
4. Report of 5-10 pages
5. Oral exam (approx. 25 min)

Literature:

1. Navas, M., and Ordonez, C., *Efficient computation of PCA with SVD in SQL*, in DMMT, 2009
2. Baker, K., *Singular Value Decomposition Tutorial*, Technical report, 2005

Task assignment and supervisor:

- Mourad Khayati

Starting date of thesis: 15.10.2012

Ending date of thesis: 11.12.2012

Exam date: December 11, 2012, 15:30 - 16:00

Exam place: BIN-2.E.13

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor