Proposal - M.Sc. Research Project
by
Bill (Mufan) Li
Department of Statistical Sciences
University of Toronto
Phone: (416)843-9179
Email: mufan.li@mail.utoronto.ca

Assessing GPA Measures From a Collaborative Filtering Perspective

Recent developments in machine learning have made significant contributions to a wide range of fields that are not traditionally considered data science. Notably the Netflix competition have attracted a collective effort in developing models that greatly improved prediction of movie ratings by different users, creating the best movie recommendation system at the time [1]. This research project aims to apply the machine learning techniques in analyzing the student grade dataset used in [2], allowing for further understanding of student behavior for course selection, inference on a student's potential grade in courses not selected, finally allowing for an estimate of an equivalent grades by inferencing on the same set of courses. The potential application of the results include but limited to tuning difficulty of courses for fair grading, predicting demand of courses, and improving admission selection process.

Inference on missing data in the student grade dataset falls directly under a collaborative filtering problem, where a value or grade is assigned for some matches of students and courses. The greatest difficulty of this type of problem is the sparsity of data, where each student can only take a small subset of courses, leaving majority of potential course grades missing. Additionally, data is not even distributed among courses, where some courses can have few attendance. Therefore, common methods of inference will be difficult to perform with this type of problem. That being said, matrix factorization and restricted Boltzmann machines have been highly effective at collaborative filtering. This project intends to investigate these two techniques and the respective extensions with the student grade dataset.

Matrix factorization (MF), also known as singular value decomposition (SVD), decomposes a large low-rank matrix into a product of two low-rank matrices. Specifically for this problem, let A be the m-by-n matrix of grades, where A_{ij} corresponds to the grade of student i and course j, then we seek two matrices U m-by-d and V n-by-d such that

A UV' where ' here denotes transpose

By the Eckart-Young Theorem, given a best rank k-1 approximation of U and V, a best rank k approximation is obtained by adding a column to both U and V such that the product provides the best fit to the residual matrix A-UV'.

This result allows for iterative computation of U and V, reducing the complexity significantly. One approach, known as alternating least squares (ALS), fixes the column of interest of U while using least squares to fit the column of V, then fixing V to fit U, and alternates until convergence. The other approach calculates the gradient of the square error with respect to all elements of U and V, and perform steepest descent until convergence. While the optimization problem is non-convex, both approaches have been proven to reach satisfactory local minima.

On the other hand, restricted Boltzmann machines (RBM) is a completely different approach to the problem. A RBM is a Markov random field as a bipartite graph, where the joint probability follows a Boltzmann type distribution. The bipartite graph structure creates two layers without internal connections. One layer, called the visible layer, contain the values of grades for a specific student. These nodes are connected to the other layer, called the hidden layer, with symmetrical weighted connections. The joint distribution then follows as below:

[insert equations]

References

[1] A. Feuerverger, Y. He and S. Khatri Statistical Significance of the Netflix Challenge Statistical Science, 27:2, 2012, pp. 202-231

[2] M. Bailey, J. Rosenthal and A. Yoon Grades and incentives: assessing competing grade point average measures and postgraduate outcomes Studies in Higher Education, 2014 DOI: 10.1080/03075079.2014.982528