Paper Review Summary

Review Summary – 1

Title: A Review Paper on Feature Selection Methodologies and Their Applications

Authors: Ms. Swetha Srivastava, Ms. Nikita Joshi, Ms. Madhvi Gaur

Web Link: http://www.ijerd.com/paper/vol7-issue6/H07065761.pdf

Summary:

The paper discusses about different methods for feature selection and their applications. Feature selection is one of the stage for preprocessing to reduce the dimensionality. The author discusses about supervised feature selection method. Objective functions are used to evaluate the selected subset of features and is divided into three categories – Filters, Wrapper, Embedded. Certain search strategies are used to search the attributes and that include exhaustive search, Heuristic search, forward selection, backward elimination, randomized search strategy. Another feature selection method is unsupervised feature selection method, which is the procedure of arranging the objects into natural classes whose members are similar to each other, identified by a given metrics. There are wide variety of applications of text classification, sentiment analysis, image retrieval, etc and the paper discusses a point or two about how feature selection helps in the above fields. In conclusion, this paper provides a good summary of the necessity of feature selection, and its method and applications, which gave us a direction initially to start with the project

Review Summary – 2

Title: Yelp User Rating Prediction

Authors: Yifei Feng (yife@standford.edu) and Zhengli Sun (zsun4@standford.edu)

Web Link:
http://cs229.stanford.edu/proj2014/Yifei%20Feng,%20Zhengli%20Sun,%20Yelp%20User%20Rating%20Prediction.pdf

Summary:

This paper aims to create a recommender system for yelp users by predicting the users' future ratings of businesses using the dataset provided by Yelp Dataset Challenge. The authors of this paper use three different approaches – Content-based filtering, colloborative filtering and hybrid approach. This paper analyzed the sparsity of the data and tried to reduce it by implementing certain conditions, for it be used in the prediction model. They used Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to validate the performance of the prediction. For Content based approach, the authors used about 10 features and predicted the rating score from a user to a business by matching user's interests with description and attributes of the business. Standard techniques like logistic regression and Classification and Regression Tree (CARD) were used. Collaborative filtering approach was performed using models such as k-nearest neighbor model. Hybrid approach targeted to combine these two approaches to test for better performances. In collaborative Filtering, same features were used for methods like item-based K Nearest Neighbours (KNN) and matrix factorization (A factor method that performs well on sparsely populated data). As part of Hybrid approach, a combination of these two approaches were used. Lowest RMSE was recorded for Binary Decision Tree Regression and KNN combined, which came to 0.87.

Review Summary – 3

Title: A Neural Network Model for Bankruptcy Prediction

Authors: Marcus D. Odom and Ramesh Sharda

Web Link: http://ieeexplore.ieee.org/abstract/document/5726669/

Summary:

This paper implements a neural network model for predicting bankruptcy and compares the result of this prediction with a more traditional model, multivariate discriminant analysis, that is often used for bankruptcy prediction. The prediction was carried out using some significant financial ratios like Working capital/Total Assets, Retained Earnings/Total assets etc. Using Discriminant Analysis an accuracy of 86.84% was achieved. But the drawback using this method was that, the validity was assessed on the same data that was used for model formulation. The authors then used Neural network with a three perceptron network consisting of an input,ouput and one hidden layer. The input and hidden layer consisted of 5 nodes, each for one of the financial ratios. The output node consisted of a single node that would give a Boolean output. Using this design, the authors were correctly able to predict bankruptcy and non-bankruptcy for the dataset. Then a plot of the confusion matrix was made, for each model – Neural network and Discriminant analysis for each variations of the of train and test dataset. The results obtained showed strong inclination towards using neural networks for future prediction problems.