Content and Context-Aware Recommender Systems for Business

C

Prageet Aeron
https://orcid.org/0000-0003-3957-5912
Management Development Institute, Gurgaon, India

INTRODUCTION

Author has already emphasized upon the importance of recommender systems for today's firms irrespective of domain such as grocery, apparel, electronic goods, entertainment services etc in the first part of this series of two articles. Today most retailing companies have proprietary recommender systems that are often highly customized and already into the next generation in terms of bringing together many different algorithms together. Today's recommender systems are no longer standalone proof of concept systems trying to gain credibility but rather are considered to be a major source of revenue booster for the organizations. Popular literature estimate suggests that almost 70% of content consumed online by Netflix viewers is through personal recommendations and according to Gomez-Uribe & Hunt (2016) they could save almost USD 1.2 billion by avoiding cancellations. Similarly, popular literature reports almost 30-35% revenues of Amazon originating through its recommender systems. It is now well established that these systems help in better customer engagement, upselling, customer retention and in boosting revenues.

Author started his discussion with basic collaborative systems and its variants in the earlier chapter and also emphasized on the growing need and logic (Aggarwal, 2015) for combining various methods. This second chapter of the series takes the on from exactly the point where we left off. Author shall start his discussion from content based systems and its variants, then move on to knowledge based systems and ensembles, subsequently author discusses context aware systems and evaluation methods for recommender systems. Finally, the author concludes the chapter with a brief discussion and emphasizing on the need for more work on recent topics in the field.

BACKGROUND

Content Based Recommender Systems

As the name rightly suggests the content-based recommender systems focus on data from users' own ratings as well as attributes or features of the product to make recommendations. As a result, these systems are inherently incapable of identifying novel recommendations and are likely to be recommended products/services more in line with past usage only. By extension of the same argument, content based systems are more appropriate for systems with detailed attribute information irrespective of the knowledge representation available for describing the attributes. Since structured data representation can be considered as subsets of unstructured data representation therefore the analysis of unstructured data can

DOI: 10.4018/978-1-7998-9220-5.ch165

be considered as more generic. Therefore, we look at the analysis of unstructured data only. Natural language processing/text processing and regression models usually form the primary workhorse of content based systems. Like any other recommender systems three parts of content based system comprise, pre-processing of raw data, learning model implementation and prediction utility.

Pre-processing

This involves transforming either text or any other data into vector-based attributes. Usually this is a domain and context centric activity which may differ as per the requirements. Transforming of text based structure into a term-document matrix or identifying most appropriate vector attributes based on a combination of term frequency and inverse document frequency (TF_IDF) is now well understood and this is just one of the ways to deal with basic data at hand. Use of stop-word removal is akin to feature selection and use of TF_IDF is an example of feature weighing in the above setting. However, based on the context methods are likely to change (Gini index or entropy based methods). Further, in addition to vector space representation there may be a need for multi-dimensional attributes from structured data which may include price, and other formal characteristics.

Learning

Learning models will require a training data dsr and testing data dse. These may represent attributes together with user rating as class or weight as per the availability. The training and testing data may even be pure documents or music albums depending upon the context. The model trained on dsr are then tested on dse which can be considered unlabeled and therefore the problem translates into a simple regression or classification kind of problem. Using any similarity metric such as cosine or Jaccard etc. k-similar documents from dse corresponding to each dsr can be identified and average rating assigned to each. As is evident for each example or each user in dsr the process operates at an individual level making it rather complex and computationally intensive. As an improvement to the above process a k-means clustering approach can be used to lower the computational effort. As a part of this approach each independent rating from training set dsr the entire dsr is clustered into a group of x. Therefore, for y independent ratings in dsr this leads to x*y groups to work with which is much lesser as compared to total number in dsr. This leads to reformulation of the vector-space for documents by aggregation across the x documents as the new vector-space representation. Subsequently, any target row of training data-set or document from dsr could be taken up and any k<x closest items could be identified (Aggarwal, 2015; Zhai & Aggarwal, 2012). In addition to above Naive Bayes methods may be applied to learning very effectively with appropriate smoothing methods (Laplace or others) taking care of rare elements or events with zero probability. Also rule-based classifiers have also been successfully utilized in certain cases. Next author focuses on a regression based method for learning in the context of content based recommender systems.

Regression Method

Depending upon the target variable in the data at hand ordinary regression, logistic regression, probit or multinomial probit or ordered probit or any other variant may be successfully used for learning. Here we discuss only the ordinary regression method together with regularization. Consider dsr as the d x t matrix, with documents or datarows and t terms or words as per the context. Assume $\overline{y}, \overline{W}$ to be the