World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:12, No:03, 2018

Restricted Boltzmann Machines and Deep Belief Nets for Market Basket Analysis: Statistical Performance and Managerial Implications

Authors : H. Hruschka

Abstract : This paper presents the first comparison of the performance of the restricted Boltzmann machine and the deep belief net on binary market basket data relative to binary factor analysis and the two best-known topic models, namely Dirichlet allocation and the correlated topic model. This comparison shows that the restricted Boltzmann machine and the deep belief net are superior to both binary factor analysis and topic models. Managerial implications that differ between the investigated models are treated as well. The restricted Boltzmann machine is defined as joint Boltzmann distribution of hidden variables and observed variables (purchases). It comprises one layer of observed variables and one layer of hidden variables. Note that variables of the same layer are not connected. The comparison also includes deep belief nets with three layers. The first layer is a restricted Boltzmann machine based on category purchases. Hidden variables of the first layer are used as input variables by the second-layer restricted Boltzmann machine which then generates second-layer hidden variables. Finally, in the third layer hidden variables are related to purchases. A public data set is analyzed which contains one month of real-world point-of-sale transactions in a typical local grocery outlet. It consists of 9,835 market baskets referring to 169 product categories. This data set is randomly split into two halves. One half is used for estimation, the other serves as holdout data. Each model is evaluated by the log likelihood for the holdout data. Performance of the topic models is disappointing as the holdout log likelihood of the correlated topic model – which is better than Dirichlet allocation - is lower by more than 25,000 compared to the best binary factor analysis model. On the other hand, binary factor analysis on its own is clearly surpassed by both the restricted Boltzmann machine and the deep belief net whose holdout log likelihoods are higher by more than 23,000. Overall, the deep belief net performs best. We also interpret hidden variables discovered by binary factor analysis, the restricted Boltzmann machine and the deep belief net. Hidden variables characterized by the product categories to which they are related differ strongly between these three models. To derive managerial implications we assess the effect of promoting each category on total basket size, i.e., the number of purchased product categories, due to each category's interdependence with all the other categories. The investigated models lead to very different implications as they disagree about which categories are associated with higher basket size increases due to a promotion. Of course, recommendations based on better performing models should be preferred. The impressive performance advantages of the restricted Boltzmann machine and the deep belief net suggest continuing research by appropriate extensions. To include predictors, especially marketing variables such as price, seems to be an obvious next step. It might also be feasible to take a more detailed perspective by considering purchases of brands instead of purchases of product categories.

Keywords : binary factor analysis, deep belief net, market basket analysis, restricted Boltzmann machine, topic models
Conference Title : ICOR 2018 : International Conference on Operations Research
Conference Location : Paris, France
Conference Dates : March 15-16, 2018