

Noise Reduction in Web Data: A Learning Approach Based on Dynamic User Interests

Authors : Julius Onyancha, Valentina Plekhanova

Abstract : One of the significant issues facing web users is the amount of noise in web data which hinders the process of finding useful information in relation to their dynamic interests. Current research works consider noise as any data that does not form part of the main web page and propose noise web data reduction tools which mainly focus on eliminating noise in relation to the content and layout of web data. This paper argues that not all data that form part of the main web page is of a user interest and not all noise data is actually noise to a given user. Therefore, learning of noise web data allocated to the user requests ensures not only reduction of noisiness level in a web user profile, but also a decrease in the loss of useful information hence improves the quality of a web user profile. Noise Web Data Learning (NWDL) tool/algorithm capable of learning noise web data in web user profile is proposed. The proposed work considers elimination of noise data in relation to dynamic user interest. In order to validate the performance of the proposed work, an experimental design setup is presented. The results obtained are compared with the current algorithms applied in noise web data reduction process. The experimental results show that the proposed work considers the dynamic change of user interest prior to elimination of noise data. The proposed work contributes towards improving the quality of a web user profile by reducing the amount of useful information eliminated as noise.

Keywords : web log data, web user profile, user interest, noise web data learning, machine learning

Conference Title : ICKDDM 2018 : International Conference on Knowledge Discovery and Data Mining

Conference Location : London, United Kingdom

Conference Dates : January 18-19, 2018