Editorial

Lorna Uden

FCET,
Staffordshire University,
The Octagon, Beaconside,
Stafford, ST18 0AD, UK
E-mail: L.uden@staffs.ac.uk

Biographical notes: Lorna Uden is Professor Emeritus of IT Systems in the Faculty of Computing, Engineering and Technology at Staffordshire University. Her research interests include technology learning, HCI, mobile learning, activity theory, knowledge management, web engineering, multimedia, e-business, service science and innovation, Semantic Web, software as a service (SaaS) and problem-based learning.

Welcome to V7N3 of this journal. The first paper is by Panagiotis Giannikopoulos and Costas Vassilakis entitled 'A distributed recommender system architecture'. These authors argue that contemporary internet architectures including server farms and blog aggregators, web log data may be scattered among multiple cooperating peers. In order to perform content personalisation through provision of recommendations on such architectures, it is necessary to employ a recommendation algorithm. However the majority of such algorithms are centralised, necessitating excessive data transfers and exhibiting performance issues when the number of users or the volume of data increases.

In this paper, they propose an approach where the clickstream information is distributed to a number of peers, who cooperate for discovering frequent patterns and for generating recommendations, introducing

a architectures that allow the distribution of both the content and the clickstream database to the participating peers

b algorithms that allow collaborative decisions on the recommendations to the users, in the presence of scattered log information.

The proposed approach may be employed in various domains, including digital libraries, social data, server farms and content distribution networks. They have used an extension of the FGP algorithm into the P2P domain, which allows distributed web logs to be mined efficiently for frequent patterns in a shared-nothing scheme. A further extension was the incorporation of log distribution according to histogram statistics to achieve better load balancing and smaller algorithm completion times. The proposed extensions have been implemented and benchmarked, using log files of various sizes and different numbers of peers, and the performance gains have been quantified in different network settings. The algorithm has limitations and further work is needed to address this.

The second paper is, 'A data-centric approach to feed search in blogs' by Flora S. Tsai. This paper studied the blog distillation and feed search task in the TREC Blog Track, which was designed to search for the relevant feeds which have a principal and recurring interest in a particular topic or query. In this paper, the author used a novel

data-centric approach to achieve good results compared to others in the TREC Blog Track. According to the author, the proposed data-centric approach which converts the blogs to a database format and then processes them directly has proved to be successful by obtaining the highest MAP of 0.4083, which is better than the best results reported in the 2007 TREC Blog Track. Tsai also argues that her approach is promising. For future work, a complete data conversion can be performed so that it will be more accurate and perhaps further improve the performance.

The third paper is 'Mining potential research synergies from co-authorship graphs using power graph analysis' by Iraklis Varlamis and George Tsatsaronis. The authors of this paper proposed a novel representation model for bibliographic data, which combines co-authorship and content similarity information, and allows for the formation of scientific networks. Using a graph visualisation tool from the biological domain, they were able to provide comprehensive visualisations that help to uncover hidden relations between authors and suggest potential synergies between researchers or groups. The contribution of their approach lies in the use of a graph reduction method that facilitates the efficient visualisation of the dense co-authorship graph, the identification of potential research synergies based on the analysis of the power graph, and the ranking of potential co-author pairs by similarity of interests. More specifically, they have demonstrated how the use of power graph analysis can uncover potential future research synergies between authors. According to these authors, this modular approach helps them to avoid the burden of finding the optimal clustering and classification scheme for bibliographic data organisation. As a proof of concept, they demonstrated some of the capabilities of their approach in the DBLP data. Further work is to apply the same approach to more bibliographic networks as well as to other social networks, and to study the evolution of the graphs over time based on the comparison of different graph snapshots taken in different years.

The final paper is, 'Tagging users based on Twitter lists' by Yuto Yamaguchi, Toshiyuki Amagasa and Hiroyuki Kitagawa. This paper addresses the problem of tagging users in Twitter, one of the most popular microblogs. Instead of analysing tweets, the authors of the paper proposed a method for tagging Twitter users using Twitter lists. The Twitter list is provided as an official function to make and share user lists (abbreviated as 'lists'). Users make lists when they want to group user accounts that seem to share a major topic. Users included in a list are called list members. A list name represents an appropriate topic about which members frequently post. The list name can be regarded as a tag generated by the user who made the list and that suggests the associated topic. According to these authors, the proposed method extracts terms from list names and tags list members using extracted terms by calculating MI between terms and users. Experimental results showed that the proposed method could assign appropriate tags to users. This indicates that employing lists to tag users and using MI to calculate the degree of relevance between terms and users were effective.

To deal with the dataset sparsity problem, the authors plan to propagate tags from users who have many tags to similar users who have a few tags. There are still many problems that need to be addressed before the method can be used effectively.