Hybrid TRS-PSO Clustering Approach for Web2.0 Social Tagging System

Hannah Inbarani H, Department of Computer Science, Periyar University, Salem, India

Selva Kumar S, Department of Computer Science, Periyar University, Salem, India

Ahmad Taher Azar, Benha University, Benha, Egypt

Aboul Ella Hassanien, Cairo University, Cairo, Egypt, & Computers and Information Faculty, Beni Suef University, Beni Suef, Egypt, & Scientific Research Group in Egypt (SRGE), Giza, Egypt

ABSTRACT

Social tagging is one of the important characteristics of WEB2.0. The challenge of Web 2.0 is a huge amount of data generated over a short period. Tags are widely used to interpret and classify the web 2.0 resources. Tag clustering is the process of grouping the similar tags into clusters. The tag clustering is very useful for searching and organizing the web2.0 resources and also important for the success of Social Bookmarking systems. In this paper, the authors proposed a hybrid Tolerance Rough Set Based Particle Swarm optimization (TRS-PSO) clustering algorithm for clustering tags in social systems. Then the proposed method is compared to the benchmark algorithm K-Means clustering and Particle Swarm optimization (PSO) based Clustering technique. The experimental analysis illustrates the effectiveness of the proposed approach.

Keywords: Bookmark Selection, Clustering, K-Means Clustering, PSO Clustering, Tolerance Rough Set

1. INTRODUCTION

Web sites have introduced a collection of refined techniques known as Web 2.0. Social tagging is one of the important features of Web 2.0. Social tagging systems allow users to annotate resources with free-form tags (Guandong Xu, 2013). Tagging is a popular way to interpret web 2.0 websites. The tags are collected by the user's favorites or interests in the social bookmarking website. In Social Bookmarking systems, Tagging can be seen as the act of

linking of entities such as users, resources and tags (Caimei Lu, 2011). When a user applies a tag to a resource in the system, a multilateral relationship between the user, the resource and the tag is formed. It helps user better way to understand and distribute their collections of attractive objects (Gupta M et.al 2010) (Figure 1).

Feature Selection (FS) is an essential part of knowledge discovery. It is a process which attempts to select features which are more informative (Velayutham, 2011). FS is separated into the supervised and unsupervised categories.

Figure 1. Multilateral relationship
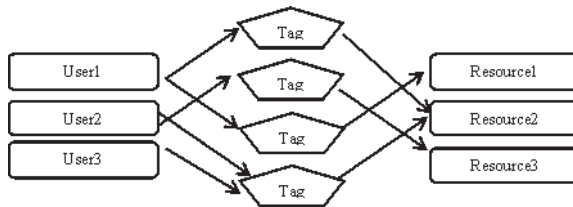
When class labels of the data are existing we use supervised feature selection, otherwise the unsupervised feature selection is applicable (Jothi, 2012). In this paper, the feature selection technique is applied for Bookmark Selection (BMS). The goal of Bookmark selection is to find out a marginal bookmarked URL subset from a Web 2.0 data while retaining a suitably high accuracy in representing the original bookmarks (Kumar SS, 2013 ; Inbarani et.al 2014b). The BMS is a must due to the profusion of noisy, irrelevant or misleading bookmarks added to web 2.0 sites. BMS is also used to increase the clustering accuracy and reduce the computational time of clustering algorithms. Web 2.0 user-generated tagged bookmark data usually contains some irrelevant bookmarks which should be removed before knowledge extraction. In this paper, Unsupervised Quick Reduct (USQR) method is used for selection of tagged bookmarks since there are no class labels for web 2.0 tagged bookmark data.

Clustering is one of the important tasks in data mining. Clustering is considered as an interesting approach for finding similarities in data and putting similar data into groups (Selvakumar, 2013). Clustering partitions a dataset into several groups such that the similarity within a group is larger than that among groups (Hammouda, 2000). Tag clustering is the process of grouping similar tags into the same cluster and is important for the success of Social Systems. The goal of clustering tags is to find frequently used tags from the tagged bookmarks. On the tag clustering, similar tags are clustered based on tag weights associated with bookmarks (Kumar SS, 2013).

In this paper, we proposed a TRS-PSO clustering algorithm for social systems and the techniques are implemented and tested against a various social tagging dataset. The performance of these techniques is compared based on 'goodness of clustering' evaluation measures.

The proposed work consists of

- **Data Extraction**: Fetching data from social systems and the Data set are converted into matrix representation. "Delicious" (del.icio.us) is a famous social bookmarking web service for storing, sharing, and determining web bookmarks.
- **Data Formatting:** Data formatting consists of mapping the tags and bookmarks based on tag weights represented in matrix format.
- **Bookmark selection**: BMS is the progression of selecting more useful tagged bookmarks from a set of bookmarks associated with tags.
- **Tag Clustering**: To cluster relevant tags based on tag weights associated with selected bookmarks

The rest of this paper is organized as follows: Section 2 presents some of the related work in web 2.0 tag clustering and feature selection. Section 3 Present Methodology of this research work. In Section 4, the experimental results have been reported. And the conclusion has been addressed in Section 5.

Related Content

Fuzzy Decision Support System for Coronary Artery Disease Diagnosis Based on Rough Set Theory
Noor Akhmad Setiawan (2014). *International Journal of Rough Sets and Data Analysis (pp. 65-80).*
www.igi-global.com/article/fuzzy-decision-support-system-for-coronary-artery-disease-diagnosis-based-on-rough-set-theory/111313?camid=4v1a

Attribute Reduction Using Bayesian Decision Theoretic Rough Set Models
Sharmistha Bhattacharya Halder and Kalyani Debnath (2014). *International Journal of Rough Sets and Data Analysis (pp. 15-31).*
www.igi-global.com/article/attribute-reduction-using-bayesian-decision-theoretic-rough-set-models/111310?camid=4v1a

Forecasting Exchange Rates: A Chaos-Based Regression Approach
Ahmed Radhwan, Mahmoud Kamel, Mohammed Y. Dahab and AboulElla Hassanien (2015). *International Journal of Rough Sets and Data Analysis (pp. 38-57).*
www.igi-global.com/article/forecasting-exchange-rates/122778?camid=4v1a

Online Distance Education and Embedded Librarianship Integration
Robin Phelps-Ward, Thalia Mulvihill, Lisa Jarrell and Brenda Yates Habich (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 2249-2257).*
www.igi-global.com/chapter/online-distance-education-and-embedded-librarianship-integration/112636?camid=4v1a