Hybrid TRS-FA 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

ABSTRACT

Social tagging is one of the vital attributes of WEB2.0. The challenge of Web 2.0 is a gigantic measure of information created over a brief time. Tags are broadly used to interpret and arrange the web 2.0 assets. Tag clustering is the procedure of grouping the comparable tags into clusters. The tag clustering is extremely valuable for researching and organizing the web2.0 resources furthermore critical for the achievement of Social Bookmarking frameworks. In this paper, the authors proposed a hybrid Tolerance Rough Set Based Firefly (TRS-Firefly-K-Means) clustering algorithm for clustering tags in social systems. At that stage, the proposed system is contrasted with the benchmark algorithm K-Means clustering and Particle Swarm optimization (PSO) based Clustering technique. The experimental analysis outlines the viability of the suggested methodology.

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

1. INTRODUCTION

WWW sites have represented an accumulation of refined techniques known as Web 2.0. Social tagging is one of the critical peculiarities of Web 2.0. Social tagging systems allow users to use resources with free-form tags (Guandong, 2013). Tagging is a well known approach to translate 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 taken as the demonstration of linking up of users, resources and tags (Sami, 2011). When a user employs a tag to a resource in the system, a multilateral relationship between the user, the resource and the tag is made. It helps user better way to understand and disseminate their collections of attractive objects (Gupta et.al 2010) (Figure 1).

Feature Selection (FS) is a critical bit of knowledge discovery. It is a procedure which endeavors to choose characteristics which are more useful (Velayutham, 2011). FS is separated into the supervised and unsupervised categories. 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 objective of Bookmark selection is to find out a marginal bookmarked URL subset from a Web 2.0 data while retaining

DOI: 10.4018/ijrsda.2015010105
a suitably high accuracy in representing the original bookmarks (Kumar, 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 Reduce (USQR) method is utilized for selection of tagged bookmarks since there are no class labels for web 2.0 tagged bookmark data.

Clustering is one of the essential tasks in data mining. Clustering is considered as an intriguing methodology 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 objective of clustering tags has been to discover often utilized tags from the tagged bookmarks. On the tag clustering, similar tags are clustered based on tag weights associated with bookmarks (Kumar, 2013).

Firefly algorithm (FA) is a swarm-based algorithm that can be applied for solving optimization problems. In this paper, we concentrate on a clustering algorithm utilizing the Tolerance Rough sets concept that are consolidated into the original firefly algorithm to enhance the execution. We proposed a TRS-Firefly-K-Means clustering algorithm for social systems and the techniques are implemented and tested against a various social tagging dataset. The implementation of these techniques is compared based on ‘goodness of clustering’ evaluation measures.

The proposed work comprises of

1. **Information Extraction:** Fetching information 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.

2. **Data Formatting:** Data formatting comprises of mapping the tags and bookmarks focused around tag weights represented in matrix format.

3. **Bookmark Selection:** BMS is the progression of selecting more useful tagged bookmarks from a set of bookmarks associated with tags.

4. **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 study. In Section 4, the experimental results have been covered. And the conclusion has been addressed in Section 5.
16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/article/hybrid-trs-fa-clustering-approach-for-web20-social-tagging-system/122780?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

3D Media Architecture Communication with SketchUp to Support Design for Learning
www.igi-global.com/chapter/3d-media-architecture-communication-with-sketchup-to-support-design-for-learning/112657?camid=4v1a

The Way We Work: Past, Present, and Future
www.igi-global.com/chapter/way-work-past-present-future/65312?camid=4v1a

Modified Distance Regularized Level Set Segmentation Based Analysis for Kidney Stone Detection
www.igi-global.com/article/modified-distance-regularized-level-set-segmentation-based-analysis-for-kidney-stone-detection/133531?camid=4v1a
A Multitrait-Multimethod Analysis of the End User Computing Satisfaction and Computer Self-Efficacy Instruments


[www.igi-global.com/chapter/multitrait-multimethod-analysis-end-user/30341?camid=4v1a](www.igi-global.com/chapter/multitrait-multimethod-analysis-end-user/30341?camid=4v1a)