

INFS4203/7203 Data Mining Project

1. Introduction

Flickr (<http://www.flickr.com/>) is one of the most popular Web2.0 Image sharing platforms. In addition to being a popular website to manage and share personal photos, Flickr also provides interface for users to tag or annotate these resources with their own words. Tags not only can help distribute photos, but also they are usually used by modern search engines to describe and index the content of images.

Consider such a scenario: Peter is using Flickr to upload a new photo and he tags it with some words. It is impossible for him to come up with all the relevant words at one time. In order to assist Peter find more potentially relevant tags, what can Flickr do? Can we integrate some mechanisms into Flickr such that it can automatically recommend tags according to the existing tags?

2. Task:

1. Use Flickr API (<http://www.flickr.com/services/api/>) to collect image tag dataset.
 - 1.1. Data Collection
 - 1.1.1. The scale of dataset is related to the following mining task.
 - 1.1.2. Never download original images!!
E.g., Tags, ImageURLs, Users, ImageIDs, etc.
 - 1.2. Data Pre-processing
 - 1.2.1. Low-frequency and high-frequency terms should be removed.
 - 1.2.2. Miss-spelling words
 - 1.2.3. Data format (string or integer?)
2. Use Association Rule Mining to discover user tagging behaviours/patterns.
 - 2.1. Small dataset (limited tag vocabulary and potential associations)
 - 2.2. Large dataset (efficiency issues need to be considered!)
E.g., Group semantic similar transactions into several clusters, then perform ARM within each group.
3. Design a tag recommendation approach based on mined tagging patterns.
4. Present your work.

Note that since most of you are master and honour students, the main purpose of our project is to get you familiar with the research process rather than asking you guys to find a perfect solution. Do not worry about effectiveness of your work, just focus on how to apply what you've learned to handle a real research problem and coming up with some novel ideas!