Erdi TAC 2017.03.27
1501214127

TOPICS IN QUANTITATIVE FINANCE PROJECT PROPOSAL

CLUSTERING KEYWORDS

What is the project about?

This project will be a part of a bigger project called Siphtor. Siphtor is an app that matches the news according to users' interests. It also recommends news articles taking the user's reading behaviors into account.

The ultimate goal of the project for this class is to create keywords clusters to use in the recommendation system.

What will I do?

In this project I will scrap (crawl) at least 10,000 articles from 30 most famous and trusted news websites. Then I'll divide the articles into paragraphs. After that I'll use Open Calais to extract keywords from each paragrahp. Open Calais gives relevance scores for each keyword. I'll find a method to place the keywrods on a 2 dimensional plane or a 3 dimentional space. This will be the thoughest part of the project.

After I'm done with placing the keywords I'll use one of the clustering algorithms to create clusters of keywords.

How to use keywords clusters for recommendation?

We have an interest table that consists of keywrods for each user. We add keywords according to user's click, like, dislike, share etc. actions. Essentially we want to find what keywords are related to our user's keywords that are in his interest table. If we have clusters we can simply take a keyword from the interest table, find its cluster and get all the kwywrods in the cluster as "related keywords". Then we can find the articles that scores high for "related kwywords". These articles will be the articles we recommend.

P.S.

The websites I choose are diversified in terms of the topics they cover. For example I chose sports, political, fashion, cooking, hobbies news sites so that I can have more and more different keywords.