A Contribution to Rating and Recommendation Systems: Concepts, Development and Evaluation

Oliver Diestel

January 16, 2013

Overview

Website and Rating page 1 general overview what do we display on which page? image qa / website we display questions on website

Rating

- Click
- active time on page

If a user stays shorter or longer on a page than the average user - it is a high or low rating.
We rate QA pages and webpages, however we only recommend qa pages.

Tagging

- Standard-Thesaurus Wirtschaft, tripple store
- Stemming: Compute the root of a word.
- Computing Levenshtein distance: Calculate the distance between two words.

Principle

Every question will be tagged with a prefered label from the ZBW Thesaurus. For every word in the question find the root of the word and match it against the thesaurus. If we find a matching prefered label, Done. If we find a matching synonym take the preferd label as tag, Done.

Recommendation

Three part recommendation. Pipeline architecture

- Calculate similarity between items
- Add similar ratings to a user/rating vector.
- Calculate singular value decomposition
- Find similar users

Input/Output

Input: ⟨user_id, item_id, rating⟩, [List of tags], number of recommendations
Output: Sorted list of Items with avg. rating that correspond with the pref. input label

- Item: id, name, tag
- User: id, name

Item-Based Algorithm

Find items that have similar ratings. Calculate the similarity values for every item.

Cosinus Similarity

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} * \vec{b}}{|\vec{a}| * |\vec{b}|}$$

Take the differences of the average rating behaviour of the user into account.

Advanced Cosinus Similarity

$$sim(a, b) = \frac{\sum_{u \in U}(r_{u,a} - \overline{r_u})(r_{u,b} - \overline{r_u})}{\sqrt{\sum_{u \in U}(r_{u,a} - \overline{r_u})^2} \sqrt{\sum_{u \in U}(r_{u,b} - \overline{r_u})^2}}$$

Calculate predictions for similar items.

User u, Product p $pred(u, p) = \frac{\sum_{i \in ratedItems(u)} sim(i,p) * r_{u,i}}{\sum_{i \in ratedItems(u)} sim(i,p)}$

Fill recommendation vector of each user with similar item ratings

Singular Value Decomposition

Create a SVD with the matrix $M = U * \Sigma * V^t$

- U dimension m x m, orthogonal matrix, spans the column space of matrix m
- Σ dimension m x n, diagonal matrix having only r nonzero entries
- V dimension n x n, orthogonal matrix, spans the row space of matrix m

Σ Properties

Diagonal entries of Σ have the property, that $\sigma > 0$ and $\sigma_i \geq \sigma_{i+1}$

Calculate cosinus similarity between users

Find Questions

Find questions that are hightly rated by similar users that match the current topic

Display Questions

Add a tag and a div-placeholder on every webpage. Display top questions for the current user on the webpage.