

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

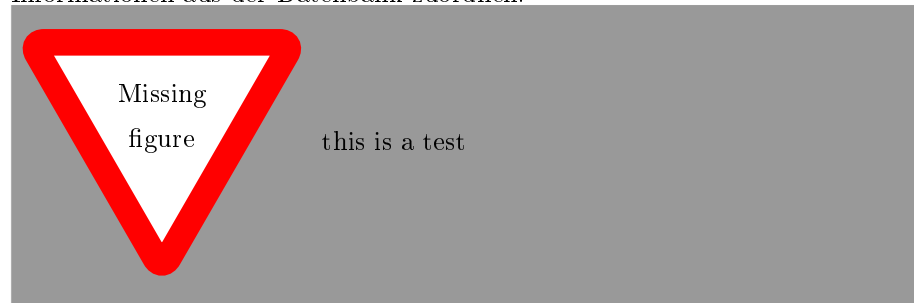
December 10, 2012

## Concepts

### Taggen

Mithilfe eines part of speech tagger (POS Tagger) alle begriffe eines Satzes taggen und ein minimales subset der Wörter eines Satzes gegen die Begriffe des ZBW Thesaurus Matchen und den Deskriptoren zuordnen. Somit kann man später den 20000 Synonymen die den 6000 Deskriptoren zugeordnet sind die Informationen aus der Datenbank zuordnen.

this is a test



<++>

### Rating

- Seiten Aufruf
- aktive Aufenthaltsdauer
- Scrollverhalten
- benutzerdefinierte Tasks

Ein Rating ist ein Wert zwischen 0 und 10. Null bedeutet nicht vorhanden, 10 bedeutet max rating wird in der Regel nur durch benutzerdefinierte tasks erreicht.

Jeder standard task hat einen standard Wert welcher beim Einbinden des skripts abgeändert werden kann.

## **Recommendation**

### **Collaborative recommendation**

#### **0.0.1 User-Based (Usually memory based)**

Finde Benutzer die ähnliche Bewertungen in der Vergangenheit abgegeben haben. Anschließend berechne für jedes Produkt p welches der Benutzer noch nicht gesehen hat einen Wert in Abhängigkeit der Bewertungen der ähnlichen Benutzer.

#### **0.0.2 Item-Based (Usually model based )**

Finde Objekte die ähnliche Bewertungen haben wie ein noch nicht gesehenes Produkt. Betrachte die Bewertungen des Benutzers für diese ähnlichen Produkte berechne den Durchschnitt.

#### **0.0.3 Matrix factorization ( Usually model based )**

SVD Matrizen Zerlegung. Zerlege eine Bewertungstabelle so, dass  $M = U * \text{Sigma} * V^T$ . Betrachte nur die obersten Spalten von U und V und approximiere mit diesen die Benutzer sowie die Items.

#### **0.0.4 Probabilistic Approach ( Usually model based)**

#### **0.0.5 Slope One ( Usually model based )**

Problem: Viele Daten viele Benutzer und vorraussichtlich viele neue Informationen/Benutzer pro Tag. Daher ist ein rein model basierter Ansatz nicht aktuell genug. Ein rein memory basierter Ansatz ist mit einer normalen Serverauslastung nicht möglich daher muss ein hybrider Ansatz gewählt werden.

## **Allgemeiner Aufbau**

### **Fragen Taggen**

Plugin mit python mithilfe des Natural language toolkits. Eine Frage bekommt eine many to many connection zu einem Datenbankeintrag mit einem Deskriptor.

Jedem Deskriptor sind n Synonyme/nicht Deskriptoren zugeordnet.

### **Javascript Server und Client seitig**

Javascript script in website einbinden angaben rating:

- id des objekts

- css selector für rating (Beispiel button wird geklickt starkes Rating von 0-10)
- standard Ratings können aktiviert/deaktiviert werden
- standard Rating Werte können verändert werden

Standard Ratings sind:

- Seiten Aufruf
- aktive Aufenthaltsdauer
- scrollverhalten

beim Verlassen der Seite werden die Daten an das recommendation system übertragen

Angaben Ausstreuen der Informationen:

- Angabe eines css selectors mit divboxen in denen die Information eingebettet werden soll.
- Angabe des thematischen Bezuges damit thematisch passende Informationen eingeblendet werden können.

## Ideen

Cronjob der automatisch ueberprueft ob neue datenbank eintraege bestehen wenn dieses zutrifft, tagge diese

## String match

[http://en.wikipedia.org/wiki/Levenshtein\\_distance](http://en.wikipedia.org/wiki/Levenshtein_distance) bei distance groesser 4 abbrechen

Bitap algorithm