

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

January 14, 2013

Concepts

Taggen

Deprecated! 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. Taggen: Triple store database mit thesaurus. All Woerter werden gegen den thesaurus gematched. Wenn treffer dann wird das preferred Label als tag verwendet.

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 $M = U \cdot \Sigma \cdot V^T$. 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

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

Jedem preferred label sind n Synonyme zugeordnet. Cronjob, der alle 30min die Datenbank nach neuen eintraegen ueberprueft. Wenn ein neuer Eintrag vorhanden ist matcht er alle Woerter des Eintrags gegen die Woerter des thesaurus. Die preferred label der gematchten Woerter werden als tag verwendet.

Javascript Clientseitig

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

String match

http://en.wikipedia.org/wiki/Levenshtein_distance bei distance grösser 4 abbrechen

Bitap algorithm

0.0.6 Summaries

<+> SCENE : A Scalable Two-Stage Personalized News Recommendation System<+> Difference between news recommendation and qa-recommendation: for news recommendation the publishing date is very important therefore there are many new articles for every user. With this in mind it is important to know the subject that the user is interested in. However at qa-recommendation it is important to know similar users but the overall interest of the user is not that important because the subject of the recommended item is based upon the theme of the currently visited website. Quote: 'What differentiates our work from prior methods is that we model personalized news recommendation as a budgeted maximum coverage problem, i.e., the selection of one news item will influence the selection of the following news items.' Two-stage procedure at first the news articles are clustered