Making Drupal Recommend with KNIME

Dr. Heiko Hofer
Outline

- About me
- Drupal in 2014
- Recommendations for a movie shop
- Association rule mining
- What's next?
About me

- Freelance Developer, Hamburg
- Programming in Java since 2001
- Working with KNIME since 2009
- Web projects with Drupal since 2009

www.dr-heiko-hofer.de/en
Drupal in 2014

• A content management system (CMS) with focus on community sites
• Drupal powers around 6% of the web
• Data centric
• Modular System
• Drupal Commerce – open source eCommerce

• Founder: Dries Buytaert (2001)
Drupal Case Study 1

Why I KNIME... Being able to look into your data after each step is crucial for error handling, and this is preserved...

KNIME Reporting

KNIME lets you intuitively integrate all your data and with the free Report Designer it is easy to build comprehensive reports in your preferred document format.

More information.

+++ 7th KNIME User Group Meeting in Zurich: Feb 12-14, 2014 +++

KNIME's next UGM will be accompanied by an interesting week of presentations and workshops, offering a more intimate encounter with KNIME on Monday and Tuesday with our KNIME User Training sessions before the User Group Meeting itself on Wednesday and Thursday and followed by interesting workshops on Friday, for example with Dean Abbott talking about "Strategies for Building Predictive Models in KNIME". Register today to reserve your seat.

KNIME - Professional Open-Source Software

KNIME [naim] is a user-friendly graphical workbench for the entire analysis process: data access, data transformation,
Drupal Case Study 2
Drupal Case Study 3
Drupal Commerce Case Study 4
Drupal Commerce Case Study 5
Dries Buytaert – Portland 2013
The next big thing we must solve:
„Provide the right content, at the right place, at the right time“
What is right?

- What content is currently popular?
- What interest does the user have?

Learn from past behaviour

⇒ This is data mining
Success Story – Amazon
Success Story – Amazon

Customers Who Bought This Item Also Bought

Star Wars: Episode II - Attack of the Clones …
Ewan McGregor
⭐⭐⭐⭐⭐ (2,237)
DVD
$37.95  ✔️Prime

Star Wars: Episode III - Revenge of the Sith …
Ewan McGregor
⭐⭐⭐⭐⭐ (1,635)
DVD
$45.84  ✔️Prime

Star Wars Trilogy (A Hope / The Empire …
Mark Hamill
⭐⭐⭐⭐⭐ (3,765)
DVD
$107.25  ✔️Prime
Netflix Price 2006

- Provider of on-demand Internet streaming media
- Recommendation system based on ratings
- $1,000,000
Batch Mode

Drupal Database \hspace{4cm} KNIME Server

Read data

Store recommendations
Web API

Drupal

Send data

Read recommendations

KNIME Server
Movie Lens Data

- Dataset with five star ratings of movies
- University of Minnesota
- Publicly available
- MovieLens 100k
  - Collected from September 19th, 1997 to April 22nd, 1998.
  - Users: 943
  - Movies: 3,706
  - Rating: 100,000
Demo

- KNIME and Drupal – two great open source products working together
KNIME Workflow

Read data filter with good rating

File Reader

- u.data
- user watched
- movie ID with ratings

Pre-processing

- user ID, rating = 4,5
- list of movies

Association Rule Learner

- input: list of movies rated 4 or 5 by each user

Antecedent - Consequent

- put all antecedents in one column

Joiner

- join with consequent

Group By

- remove duplicated associations

Prepare for Drupal
Association rule mining

• An association rule is a pattern that states when A occurs, B occurs with certain probability
• People who like movie A will like movie B
• Tends to produce a large number of rules
• Restriction of used rules with minimal support and confidence
Association Rule Learner

Options

Itemset Mining

Column containing transactions: List*(item id)

Minimum support (0-1): 0.1

Underlying data structure: ARRAY

Output

Itemset type: FREE

Maximal itemset length: 10

Association Rules

Output association rules

Minimum confidence: 0.6
Association rule mining – support

- Fraction of user ratings that contain movie A and movie B

<table>
<thead>
<tr>
<th>User</th>
<th>Ratings</th>
<th>Support = Occurrence / Total Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ABD</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>AC</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BCD</td>
<td></td>
</tr>
</tbody>
</table>

Total Support = 5
Support \{AB\} = 2 / 5 = 40%
Support \{BC\} = 3 / 5 = 60%
Support \{ABC\} = 1 / 5 = 20%
Association rule mining – confidence

- Fraction of ratings containing movie A also contain movie B

<table>
<thead>
<tr>
<th>User</th>
<th>Ratings</th>
<th>Confidence = Occurrence ({X+Y}) / Occurrence ({X}) for (X \Rightarrow Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td>Confidence ({A \Rightarrow B}) = 2 / 3 = 66%</td>
</tr>
<tr>
<td>2</td>
<td>ABD</td>
<td>Confidence ({B \Rightarrow C}) = 3 / 4 = 75%</td>
</tr>
<tr>
<td>3</td>
<td>BC</td>
<td>Confidence ({AB \Rightarrow C}) = 1 / 2 = 50%</td>
</tr>
<tr>
<td>4</td>
<td>AC</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BCD</td>
<td></td>
</tr>
</tbody>
</table>
Association rule mining – confidence

- Fraction of ratings containing movie A also contain movie B

<table>
<thead>
<tr>
<th>User</th>
<th>Ratings</th>
<th>Confidence = Occurrence {X+Y} / Occurence {X} for X ⇒ Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ABD</td>
<td>Confidence {A ⇒ B} = 2 / 3 = 66%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confidence {B ⇒ C} = 3 / 4 = 75%</td>
</tr>
<tr>
<td>3</td>
<td>BC</td>
<td>Confidence {AB ⇒ C} = 1 / 2 = 50%</td>
</tr>
<tr>
<td>4</td>
<td>AC</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BCD</td>
<td></td>
</tr>
</tbody>
</table>
Association rule mining – confidence

- Fraction of ratings containing movie A also contain movie B

<table>
<thead>
<tr>
<th>User</th>
<th>Ratings</th>
<th>Confidence = Occurrence {X+Y} / Occurrence {X} for X \Rightarrow Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td>Confidence {A \Rightarrow B} = 2 / 3 = 66%</td>
</tr>
<tr>
<td>2</td>
<td>ABD</td>
<td>Confidence {B \Rightarrow C} = 3 / 4 = 75%</td>
</tr>
<tr>
<td>3</td>
<td>BC</td>
<td>Confidence {AB \Rightarrow C} = 1 / 2 = 50%</td>
</tr>
<tr>
<td>4</td>
<td>AC</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>BCD</td>
<td></td>
</tr>
</tbody>
</table>
Drupal + KNIME in 2014

- Functioning recommendation system
- Flexible due to KNIME user interface
- Applicable to large datasets

- Currently implemented as a batch system
What's next

- Use user history (AB \(\Rightarrow\) C)
- Include negative ratings (one and two stars)
- Broaden knowledge base (use similarity measure)

- Vision: Easy service like Google Analytics
Netflix Price 2012

- Blog Post: “We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment”
Netflix Price 2012

- Blog Post: “We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment”

- KNIME: Data Analytics Made Easy
Thank you for your attention!

Do you have any questions?