Personalization and User Modeling

Web Usage Mining, Association Rules

Patrick Siehndel(siehndel@L3S.de)

IVS Semantic Web Group

26.05.2014

Web Log Analyse, Association Rules

Given is cleaned Web server log.

1. Indentify the different *Sessions* in the log, assuming that a user is identifed by the IP-Address and that a session ends 30 minutes after the last page request.

2. Based on the identifed sessions create *Frequent Item Sets* with a *Support* higher than 30%. Use the *Apriori Algorithm*:

```
Init: Initial candidates are set to be the items
Loop:
   (i)    Combine n-element Frequent Item Sets to
          n+1-element Frequent Item Sets
   (ii)   Check whether Support of the n+1-element Frequent Item
          Sets is high enough
   (iii)  Stop if no new Frequent Item
          Sets are found, else go to (i)
```

3. Identify *Association Rules* based on the *Frequent Item Sets* with a *Confidence* higher than 60%.

4. Which information from the Web Log is not used by the Association Rules?

Web Log Analyse, Association Rules

Web Log:

```
10.0.0.127 [06/May/2013:09:03:01] ... "/shop/A.html"
10.0.0.127 [06/May/2013:09:08:11] ... "/shop/C.html"
10.0.0.127 [06/May/2013:09:09:56] ... "/shop/D.html"
10.0.0.143 [06/May/2013:09:10:23] ... "/shop/A.html"
10.0.0.127 [06/May/2013:09:12:34] ... "/shop/B.html"
10.0.0.143 [06/May/2013:09:26:53] ... "/shop/D.html"
10.0.0.156 [06/May/2013:09:39:08] ... "/shop/C.html"
10.0.0.143 [06/May/2013:09:40:12] ... "/shop/C.html"
10.0.0.127 [06/May/2013:09:45:42] ... "/shop/A.html"
10.0.0.127 [06/May/2013:09:48:45] ... "/shop/C.html"
10.0.0.127 [06/May/2013:09:56:12] ... "/shop/E.html"
10.0.0.156 [06/May/2013:09:59:16] ... "/shop/E.html"
```