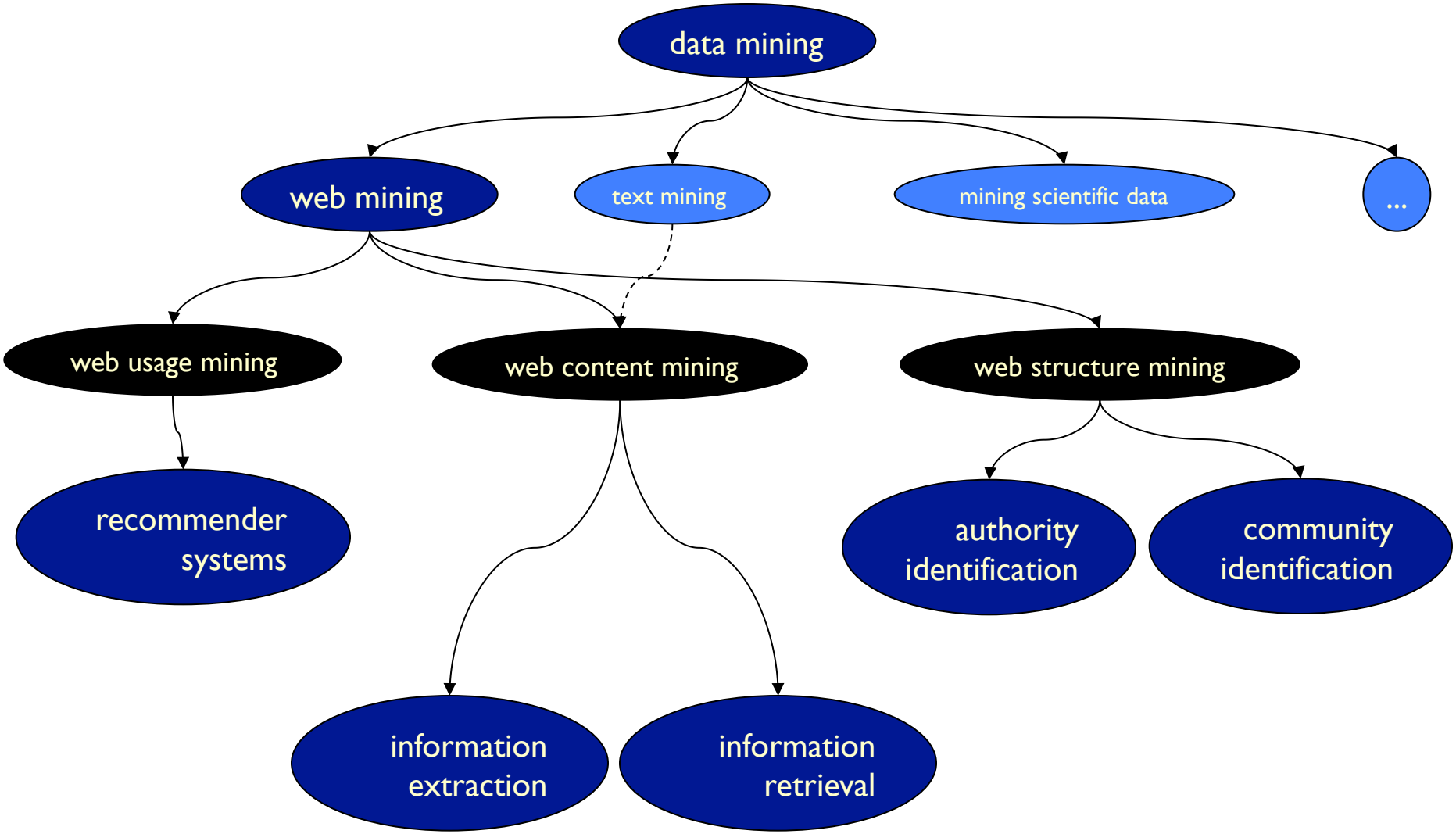
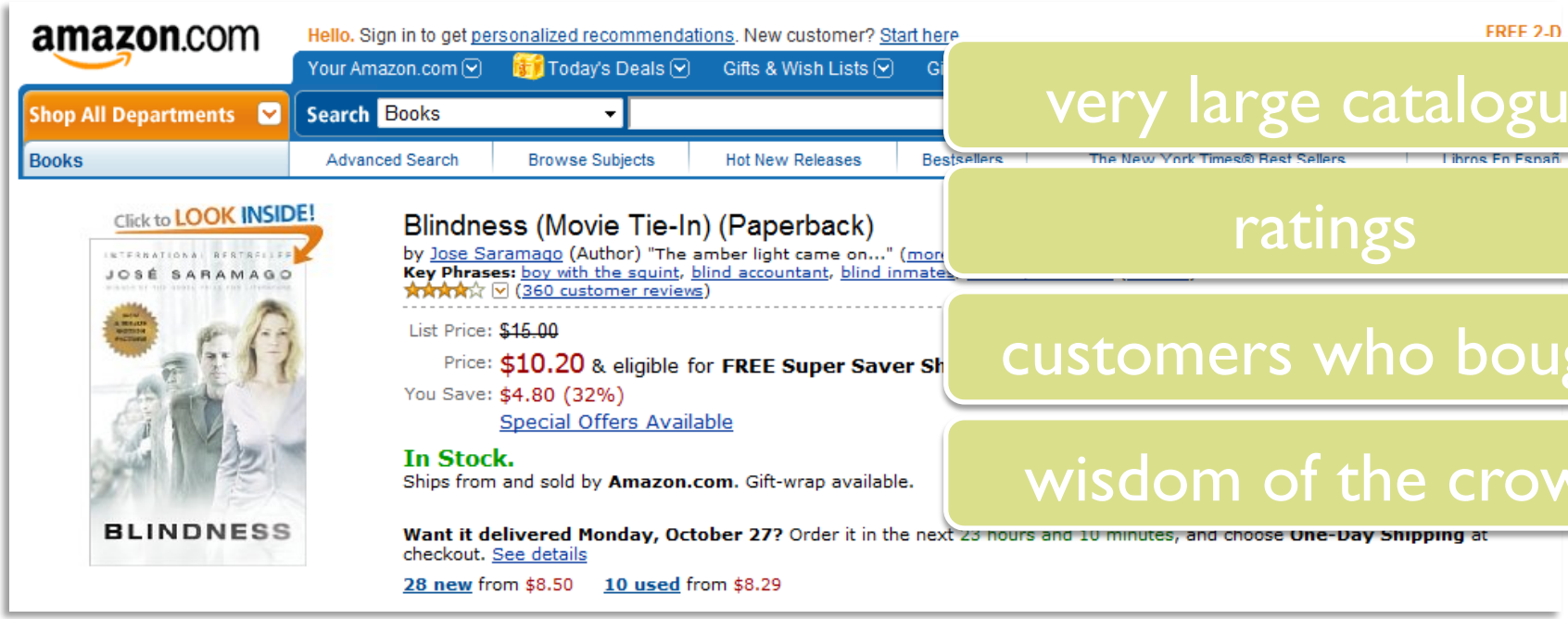
Web Mining: Recommendation with Association Rules

Alípio Jorge, DCC-FC, Universidade do Porto
amjorge@fc.up.pt

Data Mining – a structured view



Recommender Systems

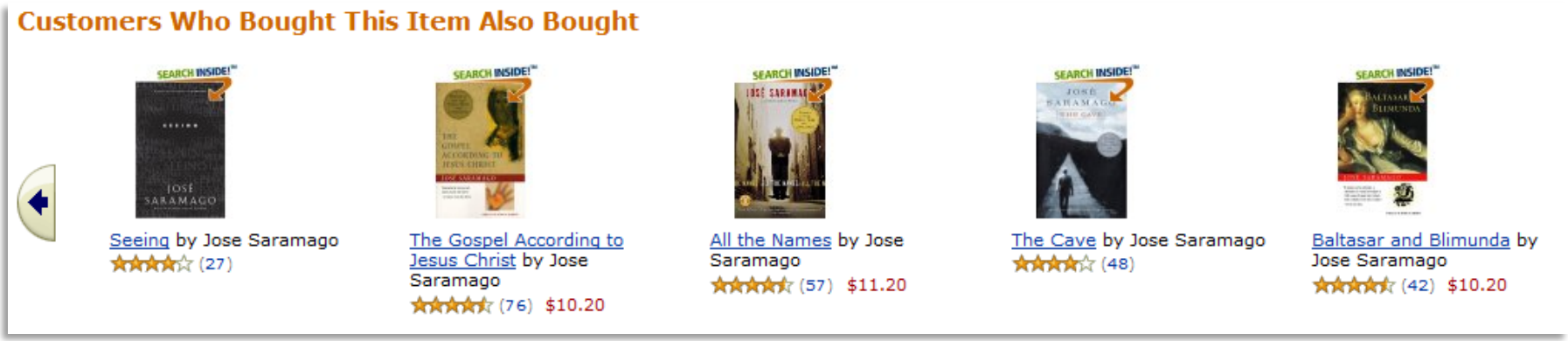


very large catalogue

ratings

customers who bought

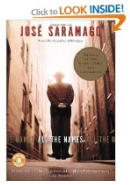
wisdom of the crowd



Recommender systems are about

- Information filtering
- Web intelligence
- Data mining
- Big Data
- Sales

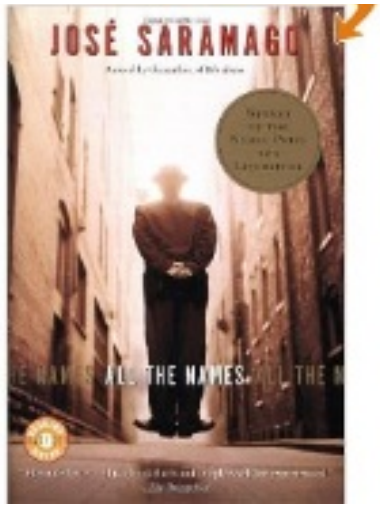
What is a recommender system?



Profiling

- Content filtering

Item



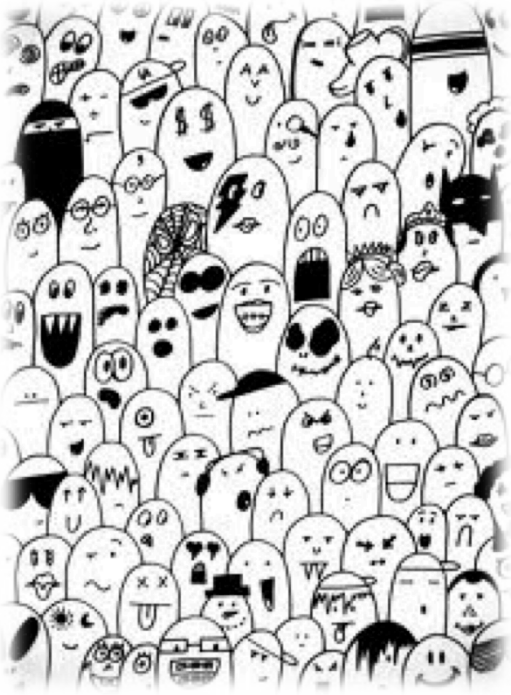
I like awarded novelists

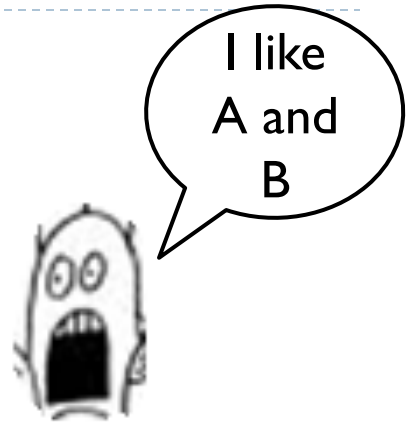
User

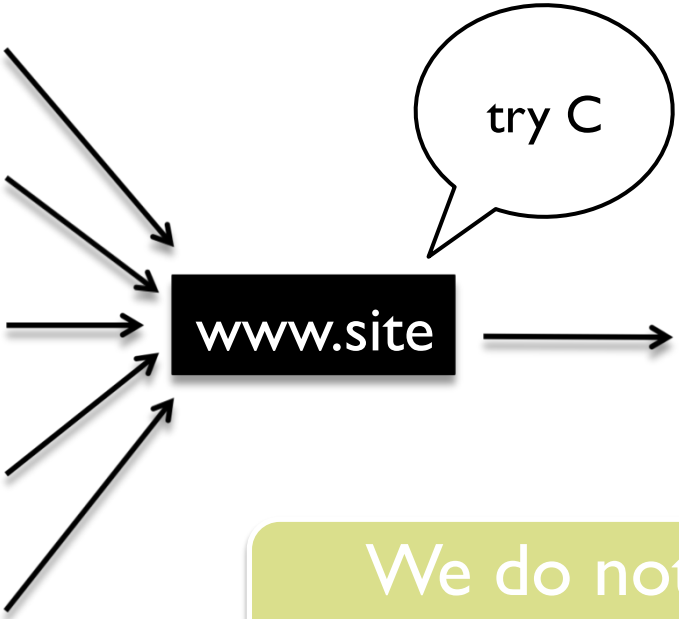
Novel of an awarded author

Profiling

- Collaborative filtering
  - The wisdom of the crowd
  - The tyranny of the average







Users who like A and B like C

We do not need to know the content

Profiling

- Hybrid
- Background information
- Social links
- Context

Example

Activity

- "Selling" MOOCs in a university
  - how to profile users ?
  - which data to keep ?
- Simplest recommendation strategies ?
  - suggest the simplest strategies
  - how would they work?

Profiling

Profiling - Collaborative Filtering

- Ratings – the 5 star system
  - usage data
  - very sparse

User 1 ratings:
A: ★
B: ★★★
G: ★★★★

USER	ITEM	RATING
1	A	1
1	B	3
1	G	4
2	A	4
2	C	2
3	B	4
3	G	5
3	F	3
3	I	4
4	B	5
4	C	4
5	G	3
5	F	4
5	I	5
5	J	3
6	A	5
6	C	3

Profiling - Collaborative Filtering

- Binary (or unary) settings

User 1 interactions:
A
B
G

Data is cheaper

Quality may be a problem

Implicit ratings

USER	ITEM
1	A
1	B
1	G
2	A
2	C
3	B
3	G
3	F
3	I
4	B
4	C
5	G
5	F
5	I
5	J
6	A
6	C

Recommender Systems on The Web

- Help facilitate navigation
  - suggest links to follow, items to view
  - try to guess the preferences of users/consumers

amjorge@fc.up.pt

Web Usage Mining: Recommendation

- Stating the problem
  - Given the current activity of a user (the active user) recommend relevant items
- Recommender algorithm
  - produces a recommender model by learning from past actions of the users
- Recommender model
  - relates the active **session** with potentially interesting items
- Session
  - set/sequence of items "transactioned" by an user in a period of time.
- Simple recommenders
  - top ... clustering

amjorge@fc.up.pt

Activity: simple recommenders

```
# top pages
sort(table(d$PAGE),decreasing=T)

# user clusters (two)
cl <-cutree(hclust(dm),k=2) # cl is a vector with cluster labels

# get top pages per cluster
sort(table(d$PAGE[cl[d$USER]==1]),decreasing=T) # top is C, A, B
sort(table(d$PAGE[cl[d$USER]==2]),decreasing=T) # top is F, G, I

# get top 2 recommendations for  user U
U<-3
sort(table(d$PAGE[cl[d$USER]==cl[U]]),decreasing=T)[1:2]
# removing articles known by the user
setdiff(names(sort(table(d$PAGE[cl[d$USER]==cl[U]]),decreasing=T)),
   names(dm[U,which(dm[U,]!=0)]))
```

Resources

- Articles
  - Resnick , P., Varian, H. R ., Recommender Systems, Communications of the ACM, Vol. 40, No. 3, March 1997.

Association Rules

for recommendation

amjorge@fc.up.pt

Web Usage Mining: Association

- Example task
  - we want to be able to predict which pages the visitor is most interested at some point of the session
  - with this knowledge we can provide Recommendations, improve usability, improve sales/loyalty.

- Strategy
  - we can look for pages that tend to be accessed in the same sessions and look for sets of pages that predict other sets of pages
  - this is done using Association Rule discovery
  - the model, here, is a set of association rules

amjorge@fc.up.pt

Web usage mining: AR-based recommender

- Modelling (training)
  - from historic DB build set of rules involving items
  - use a very low support (10/|DB|)
  - confidence
    - >50% → every recommendation is more likely to be relevant than not
    - <50% → riskier recommendations
    - >>50% → more no-recommendation situations
  - use other AR filters
  - model M is the set of rules

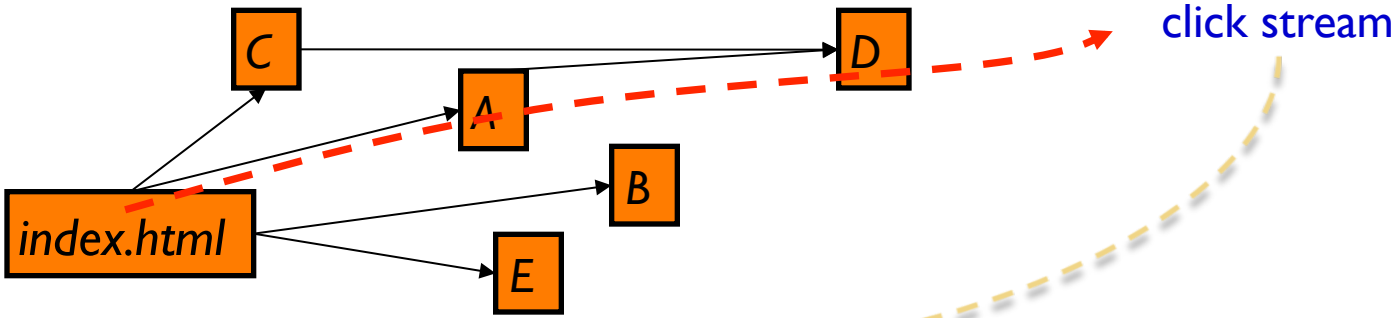
amjorge@fc.up.pt

Web usage mining: AR-based recommender

- Recommending (deploying)
  - for an active user "vieweing" items Obs
  - look for rules A→C such that
    - A is a subset of Obs
  - disregard rules with C in Obs
  - sort rules by confidence (descending)
    - if they have the same confidence prefer higher support
    - if needed prefer simpler rules
  - for a given N
    - recommend different consequents of top N rules

amjorge@fc.up.pt

Recommending with Association Rules



Obs.: A D

Rules.:

A B E → D (conf: 0,8)

A E → D (conf: 0,7)

A D → F (conf: 0,6)

A → D (conf: 0,5)

D → X (conf: 0,4)

Recommendations (top 2):

F (0,6)

X (0,4)

amjorge@fc.up.pt

Applying Association Rules in Web Mining

- Product / Item automatic recommendation
  - cross-selling, up-selling
- improve site navigation
  - recommended links
- product bundling

amjorge@fc.up.pt

Case "Management Portal"

- a Web portal for managers with articles, products, data
- restricted access
  - subscription
- we have detailed access records
  - weblogs
- data collected for one year
- business goals
  - increase frequency and length of visits
  - increase the number of articles read
  - build article bundles

amjorge@fc.up.pt

Data (pre-processed)

user	artigo
2	As-voltas-que-o-crédito-dá
2	O-valor-da-inovação-ou-vice-versa
2	Como-reconciliar-o-Marketing-e-as-Operações
2	Gestão-em-oito-lições
3	Chairman-e-CEO---um-cargo-ou-dois
4	A-guerra-pelo-talento
4	Steve-Ballmer-Um-computador-para-cada-membro-da-família
4	Universitários-trocam-cafés-por-portáteis
6	Novos-empresários-para-o-comércio
6	Retalhistas-com-vida-facilitada
6	F-C--Porto-lidera-transferências-na-pré-temporada
6	O-que-está-a-dar-no-retalho---Parte-I
6	Rotas-Úteis---Retalho
6	Leroy-Merlin-expande-se-para-sul-com-150-milhões-para-investir
6	aQuem-está-empregado-tem-muitos-direitosa
6	Grandes-superfícies-perdem--EUR-20-milhões
6	Leroy-Merlin-quer-investir-150-milhões-de-euros-até-2013
6	Modelo-Continente-com-vários-pedidos-de-licenciamento
7	IKEA-monta-casa-em-Portugal
9	Turismo-mundial-registou-a-maior-quebra-de-sempre
9	Rotas-Uteis
9	Rotas-Úteis---Marketing
10	Rotas-Uteis
10	138-projectos-aprovados-pelo-Programa-Operacional

amjorge@fc.up.pt

Examples of recommendations

- Example 1
  - Seen:
    - Medidas-de-combate-à-fraude-e-evasão-fiscal
  - Recommended:
    - Impacto-das-medidas-fiscais-da-Proposta-de-Lei-do-Orçamento-do-Estado-para-2005 (0.97)
    - Principais-alterações-em-sede-de-IRS (0.75)
    - Rotas-Uteis (0.28)

- Example 2
  - Seen:
    - Medidas-de-combate-à-fraude-e-evasão-fiscal
    - Peter-Cohan-Não-penso-que-haja-uma-retoma
  - Recommended :
    - Impacto-das-medidas-fiscais-da-Proposta-de-Lei-do-Orçamento-do-Estado-para-2005 (0.97)
    - O-valor-de-Peter-Cohan (0.75)
    - Principais-alterações-em-sede-de-IRS (0.75)
    - Rotas-Uteis (0.28275)

amjorge@fc.up.pt

Applying Association Rules for recommendation (cross-selling)

- Step by step
  - User reads articles A
  - Site knows the rule A=> B
  - The rule has a certain confidence (>20%)
  - Site displays articles B to the user
  - User chooses whether to follow recommendation or not

- Notes
  - The rules are discovered from user activity
  - Discovery is off-line
  - Rule application is on-line

amjorge@fc.up.pt

Activity (R)

USER	PAGE
1	A
1	B
1	C
2	A
2	C
3	B
3	G
3	F
3	I
4	B
4	C
5	G
5	F
5	I
5	J
6	A
6	C

- Consider the data on the side
  - read data (copy and…)
    - d<-read.table(file("clipboard"),header=TRUE)
      - macs: read.table(pipe("pbpaste"),…)
    - or paste into a file and read.table / read.csv
  - generate rules using carenR
    - library(carenR)
    - rls<-caren(d,min.sup=0.17,min.conf=0.5,Bas=T)
    - ar.pp(rls)
  - generate rules into a compiled model and recommend…
    - rls.m<-caren(d,min.sup=0.17,min.conf=0.5,Bas=T,prm=T)
    - pr<-predict.caren(rls.m,data.frame(user=7,item=c('I','F')), Bas=T)

amjorge@fc.up.pt

Activity: AR with bigger data (R)

```
d<-read.csv('pe_2004_cestos_limpos.bas',sep=' ', header=F,
  col.names=c("user","article"))

# construir o modelo de recomendação
# NOTA: agora o objecto 'rls' já não é legível
# pois trata-se de uma referência a um binário
rls <- caren(d,min.sup=0.01,min.conf=0.3, Bas=T,prm=T)

# prever preferências com base no comportamento

# suponhamos que o utilizador viu um artigo
obs<-data.frame(user=1,item='Medidas-de-combate-à-fraude-e-
  evasão-fiscal')

# os 5 artigos mais prováveis são:
predict.caren(rls,obs,Bas=T,Top=5)
```


amjorge@fc.up.pt

Activity: AR with bigger data (R)

```
# com outras observações
obs<-data.frame(
       user=c(1,1),
       item=c(
       'Medidas-de-combate-à-fraude-e-evasão-fiscal',
       'Peter-Cohan-Não-penso-que-haja-uma-retoma'))

# os 5 artigos mais prováveis são:
predict.caren(rls,obs,Bas=T,Top=5)
```


amjorge@fc.up.pt

Activity (R) – package recommenderlab

USER	PAGE
1	A
1	B
1	C
2	A
2	C
3	B
3	G
3	F
3	I
4	B
4	C
5	G
5	F
5	I
5	J
6	A
6	C

```
library(recommenderlab)
d<-read.table(pipe("pbpaste"),header=TRUE)
# d<-read.table(file("clipboard"),header=TRUE)
d<-table(d$USER,d$PAGE)
cn<-colnames(d)
m<-matrix(d,nrow(d),ncol(d))
colnames(m)<-cn
rownames(m)<-rownames(d)
m<-as(m,"realRatingMatrix")
mb<-binarize(m,minRating=1)
# build model
model<-Recommender(mb,"AR")
# see rules
inspect(model@model$rule_base)
```

Activity (R) – package recommenderlab

```
# Now produce recommendations to the active user
activeuser<-matrix(0,1,ncol(d))
colnames(activeuser)<-cn
as(activeuser,"matrix")
activeuser[1,"C"]<-1
aum<-as(activeuser,"realRatingMatrix")
aumb<-binarize(aum,minRating=1)
as(predict(model,aumb,n=2),"matrix")
```

Activity (R) – package recommenderlab

```
# Producing topN recommendations
# using the POPULAR method instead of AR
model<-Recommender(mb,"POPULAR")
show(as(model@model$topN,"list"))
as(predict(model,aumb,n=2),"matrix")

# you can also try the RANDOM method
model<-Recommender(mb, "RANDOM")

# view existing methods and parameters
recommenderRegistry$get_entry_names()
```

Resources

Resources

- Book
  - Web Data Mining, Bing Liu
- Papers
  - Mobasher, B., Dai, H., & Luo, T. (2001). Effective personalization based on association rule discovery from web usage data. ACM WIDM01. http://dl.acm.org/citation.cfm?id=502935