Symbolic Data Analysis Tools for Recommendation Systems

Byron Leite Dantas Bezerra
bldb@dsc.upe.br

Francisco Assis Tenório Carvalho
fatc@cin.ufpe.br

Centro de Informática – CIn/UFPE
Recife – Pernambuco – Brasil

KindaKarma.com
Howdy, byron.leite Logout FAQ Stats
We are all but threads in the fabric of the universe.
Hate It
Dislike It
Like It
Love It
Things You Might Like
Movies -> Star Wars: Episode VI - Return of the Jedi
Movies -> Incredibles, The
Movies -> Star Wars: Episode V - The Empire Strikes Back
Movies -> Shawshank Redemption, The
Movies -> Indiana Jones and the Last Crusade
Hate Dislike Like Love 2¢
New Stuff That Needs Ratings
Games -> Kirby and The Amazing Mirror
Music -> Crosse, Clay
Music -> Walker, T-Bone
No Ratings
Find Something
Find
Browse by Category
Authors
Children's
Classics
Fantasy
Horror
Humorists
Mainstream
Mystery
Non-Fiction
Romance
Sci-Fi
Video Games
DS
GameCube
GBA
PC
Playstation 2
PSP
XBox
Movies
Action/Adventure
Animation
Comedy
Documentary
Drama
Family
Horror
Musical
Mystery/Thriller
Sci-Fi/Fantasy
Music
Alternative
Christian
Country
Electronica
Folk
Hip-Hop
Jazz/Blues
Metal
Pop
Punk
Politics
Commentators
Congressmen
Presidents
Senators
Ads
amazon.com
and you're done.
Microsoft Xbox - Game console
(Prices May Change)
Privacy Information
Ads by Goooooogle
Looking for an Agent?
"32 agents & editors requested to read my book using Book Blaster!"
www.scriptblaster.com
Wholesale Merchandise
Locate Wholesale Suppliers Free Wholesale Sources
www.wholesaledir.com

Summary

- Problem Definition
- Personalization based on Symbolic Data Analysis
  - CMBF
  - SMCF
  - HMBF
- Experimental Evaluation

Problem Definition

- Key issues
  - Which kind of information should be added in the user profile?
  - How to acquire information about the user preference?
  - How to represent the user profile in computer memory?
  - How to recommend items to the user based on his profile?
  - How much information we need about the user in order to delivery good recommendations?
  - How Symbolic Data Analysis can be a powerful tool to the Recommendation Systems field?

Case Study

.: Movie Recommendation :.

- Items attributes

Movie	Director	Cast	Genre
F_1	D3	A1,A3,A4,A5	G1
F_2	D5	A4,A6,A8,A9	G2
F_3	D7	A2,A3,A7,A8	G3
F_4	D3	A3,A5,A6,A7	G1
F_5	D2	A1,A2,A7,A8	G3

- Rates matrix

	F_1	F_2	F_3	F_4	F_5
Brícia	5	$\varnothing$	2	5	$\varnothing$
Bryan	3	2	2	$\varnothing$	5
Elaine	1	4	5	3	5
Vanessa	4	$\varnothing$	4	$\varnothing$	5

Personalization based on Modal Symbolic Profiles

Content Modal
Based Filtering System

Social Modal
Collaborative Filtering System

Hybrid Modal
Based Filtering System

.: CMBF :.

Content Modal Based Filtering System

.: CMBF :.

- Steps:
  1. Build the user profile
     1. Pre-processing
     2. Generalization
  2. Compare the modal symbolic user profile with the symbolic description of each item in the target repository
  3. Build a personalized list to the user based on the similarity scores obtained in the previous step

.: CMBF :.

(Step 1.1 - preprocessing)

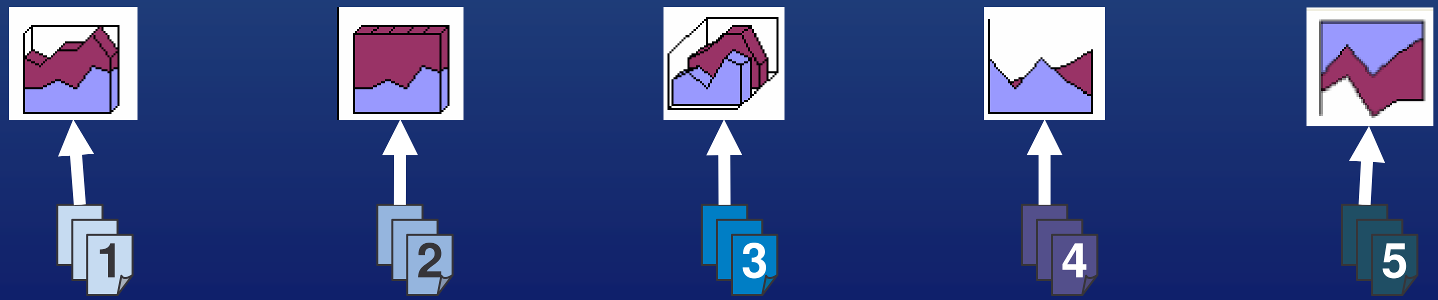
Filme	Diretor	Elenco	Gênero
F_1	D3	A1,A3,A4,A5	G1
F_2	D5	A4,A6,A8,A9	G2
F_3	D7	A2,A3,A7,A8	G3
F_4	D3	A3,A5,A6,A7	G1
F_5	D2	A1,A2,A7,A8	G3

Filme	$\tilde{X}_{F_i}^{Director}$	$\tilde{X}_{F_i}^{Cast}$	$\tilde{X}_{F_i}^{Genre}$
F_1	({D3},(1.0))	({A1,A3,A4,A5},(¼, ¼, ¼, ¼))	({G1},(1.0))
F_2	({D5},(1.0))	({A4,A6,A8,A9},(¼, ¼, ¼, ¼))	({G2},(1.0))
F_3	({D7},(1.0))	({A2,A3,A7,A8},(¼, ¼, ¼, ¼))	({G3},(1.0))
F_4	({D3},(1.0))	({A3,A5,A6,A7},(¼, ¼, ¼, ¼))	({G1},(1.0))
F_5	({D2},(1.0))	({A1,A2,A7,A8},(¼, ¼, ¼, ¼))	({G3},(1.0))

.: CMBF :.

(Step 1.2 - generalization)

- For each level of user preference available in the system, we associate a **set of modal symbolic variables**, where each variable concerns with some attribute in the movie domain



.: CMBF :.

(Step 1.2 - generalization)

	F_1	F_2	F_3	F_4	F_5
Bryan	3	2	2	∅	5

Movie	Director	Cast	Genre
y_{Bryan_1}	∅	∅	∅
y_{Bryan_2}	({D5,D7},(½,½))	({A2,A3,A4,A6,A7,A8,A9}, (1/8,1/8,1/8,1/8,1/8,1/4,1/8))	({G2,G3},(½,½))
y_{Bryan_3}	({D3},(1.0))	({A1,A3,A4,A5},(¼, ¼, ¼, ¼))	({G1},(1.0))
y_{Bryan_4}	∅	∅	∅
y_{Bryan_5}	({D2},(1.0))	({A1,A2,A7,A8},(¼, ¼, ¼, ¼))	({G3},(1.0))

.: CMBF :.

Comparing the user profile with some item (Step 2)

- Preprocess each target item i in the repository, building the modal symbolic descriptions of each one

$$\tilde{x}_i = (\tilde{X}_i^1, \ldots, \tilde{X}_i^p) \quad \text{where} \quad \tilde{X}_i^j = \tilde{X}_j(i) = (S_j(i), q_j(i))$$

- The following function measures the similarity between the user profile u and the item i

$$\Phi(u,i) = \frac{1}{|L|} * \sum_{g \in L} \frac{2 * \rho_g * (1 - \phi(y_{u_g}, \tilde{x}_i)) - \wp_{\max} - \wp_{\min}}{\wp_{\max} - \wp_{\min}}$$

.: CMBF :.

Comparing the user profile with some item (Step 2)

- The dissimilarity function ϕ takes into account the differences in the **support** and the associated **weight distributions**.

$$\phi(y_{u_g}, \tilde{x}_i) = \frac{1}{p} \sum_{j=1}^{p} \frac{1}{2} \left[\phi_{cf}(S_j(u_g), S_j(i)) + \phi_{cd}(q_j(u_g), q_j(i)) \right]$$

.: CMBF :.

Suggesting Items (Step 3)

- Sort the items of the repository according to their respective scores produced in the previous step.

.: HMBF :.

Hybrid Modal Based Filtering System

.: HMBF :.

- Steps:
  1. Build the user profile ← **CMBF = HMBF**
  2. Compute the similarity between the active user and other users in the community
  3. Select the h nearest neighbors
  4. Build the personalized list for the active user based on the suggestions of his neighbors

Collaborative Approach

.: HMBF :.

Comparing User Profiles (Step 2)

- The comparison between ***u*** e ***v*** is accomplished through the similarity function:

$$\psi(u,v)=\frac{1}{|L|}\sum_{g\in L}\left(1-\phi(y_{u_g},y_{v_g})\right)$$

- Where:

$$\phi(y_{u_g},y_{v_g})=\frac{1}{p}\sum_{j=1}^{p}\frac{1}{2}\left[\phi_{cf}(S_j(u_g),S_j(v_g))+\phi_{cd}(q_j(u_g),q_j(v_g))\right]$$

.: HMBF :.

Comparing User Profiles (Step 2)

*After execution of step 2 for all users, we go forward with **step 3** by selecting the h best users according Ψ.*

.: HMBF :.

Suggesting Items (Step 4)

- After computing the neighborhood of the active user, we select the most valued items for this users and recommend them for the active user, taking into account the level of similarity between the active user and a given neighbor.

$$\Phi(u,i) = \bar{r}_u + \frac{\sum_{v=1}^{h} \left(r_{v,i} - \bar{r}_v\right) * \psi(u,v)}{\sum_{v=1}^{h} \psi(u,v)}$$

.: SMCF :.

Social Modal
Collaborative Filtering System

.: SMCF :.

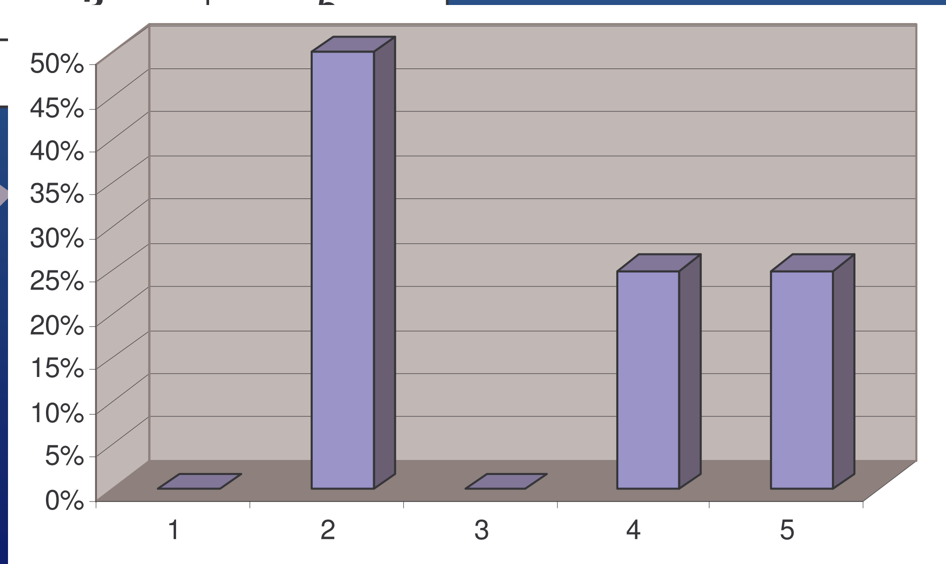
- Steps:
  1. Build the user profile
  2. Compute the similarity between the active user and other users in the community
  3. Select the h nearest neighbors
  4. Build the personalized list for the active user based on the suggestions of his neighbors

HMBF ≠ SMCF

HMBF = SMCF

.: SMCF :.

	F_1	F_2	F_3	F_4	F_5
Brícia	5	∅	2	5	∅
Bryan	3	2	2	∅	5
Elaine	1	4	5	[illegible]	[illegible]
Vanessa	4	∅	4	[illegible]	[illegible]



.: SMCF :.

(Step 1.1 – preprocessing)

	F_1	F_2	F_3	F_4	F_5
Brícia	5	∅	2	5	∅
Bryan	3	2	2	∅	5
Elaine	1	4	5	3	5
Vanessa	4	∅	4	∅	5

		i = F_3
X_{F_3}	$m_{F_3}^1$	({}, 1)
	$m_{F_3}^2$	({"Bryan", "Brícia"}, 2)
	$m_{F_3}^3$	({}, 3)
	$m_{F_3}^4$	({"Vanessa"}, 4)
	$m_{F_3}^5$	({"Elaine"}, 5)

.: SMCF :.

(Step 1.1 – preprocessing)

	F_1	F_2	F_3	F_4	F_5
Brícia	5	$\varnothing$	2	5	$\varnothing$
Bryan	3	2	2	$\varnothing$	5
Elaine	1	4	5	3	5
Vanessa	4	$\varnothing$	4	$\varnothing$	5

	Descrição Simbólica Modal :: $\tilde{x}_i = (\tilde{X}_i)$
$\tilde{x}_{F_1}$	$(\{1,2,3,4,5\}, (0.25,0.00,0.25,0.25,0.25))$
$\tilde{x}_{F_2}$	$(\{1,2,3,4,5\}, (0.00,0.50,0.00,0.50,0.00))$
$\tilde{x}_{F_3}$	$(\{1,2,3,4,5\}, (0.00,0.50,0.00,0.25,0.25))$
$\tilde{x}_{F_4}$	$(\{1,2,3,4,5\}, (0.00,0.00,0.50,0.00,0.50))$
$\tilde{x}_{F_5}$	$(\{1,2,3,4,5\}, (0.00,0.00,0.00,0.00,1.00))$

.: SMCF :.
(Step 1.2 - generalization)

	F_1	F_2	F_3	F_4	F_5
Bryan	**3**	**2**	**2**	**∅**	**5**

$\tilde{x}_{F_1}$	$(\{1,2,3,4,5\}, (0.25,0.00,0.25,0.25,0.25))$
$\tilde{x}_{F_2}$	$(\{1,2,3,4,5\}, (0.00,0.50,0.00,0.50,0.00))$
$\tilde{x}_{F_3}$	$(\{1,2,3,4,5\}, (0.00,0.50,0.00,0.25,0.25))$
$\tilde{x}_{F_4}$	$(\{1,2,3,4,5\}, (0.00,0.00,0.50,0.00,0.50))$
$\tilde{x}_{F_5}$	$(\{1,2,3,4,5\}, (0.00,0.00,0.00,0.00,1.00))$

y_{Bryan_1}	$(\{1,2,3,4,5\} , (0,0,0,0,0))$
y_{Bryan_2}	$(\{1,2,3,4,5\} , (0,0.50,0,0.375,0.125))$
y_{Bryan_3}	$(\{1,2,3,4,5\} , (0.25,0,0.25,0.25,0.25))$
y_{Bryan_4}	$(\{1,2,3,4,5\} , (0,0,0,0,0))$
y_{Bryan_5}	$(\{1,2,3,4,5\} , (0,0,0,0,1.0))$

.: SMCF :.

Comparing User Profiles (Step 2)

*After execution of step 2 for all users, we go forward with **step 3** by selecting the h best users according Ψ.*

.: SMCF :.
Suggesting Items (Step 4)

- Similar to step 4 of the HMBF!

$$\Phi(u,i) = \bar{r}_u + \frac{\sum_{v=1}^{h} \left(r_{v,i} - \bar{r}_v\right) * \psi(u,v)}{\sum_{v=1}^{h} \psi(u,v)}$$

Experimental Evaluation

Experimental Evaluation

- Objective: investigate the quality of personalized recommendations taking into account the following issues
  - Size of the user community
  - Number of items in the user profile
  - User preferences aquiring process
- Domain: movie recommendations
- Database: EachMovie + IMDB
- Methods: CMBF, SMCF, HMBF, kNN-CF, kNN-CB
- Metrics: half-life, precision, recall, f-measure, speed
- Methodology
  - inverted 10-fold cross validation X holdout

Experimental Evaluation

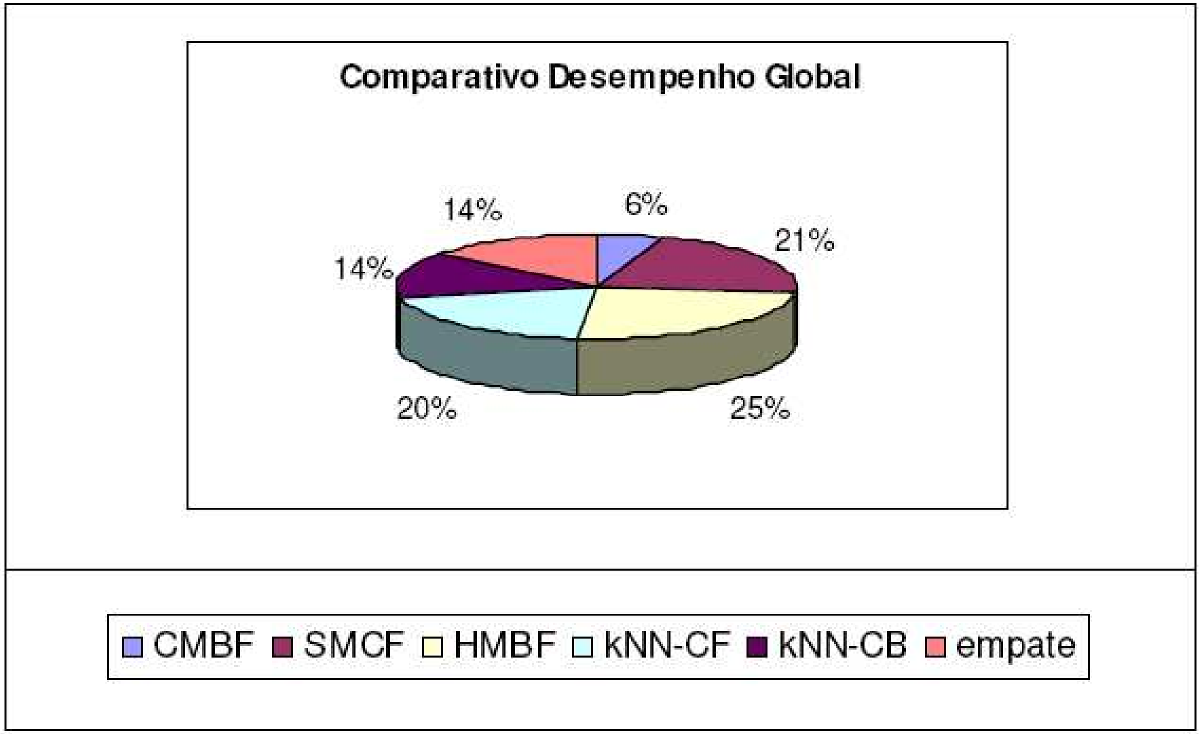


Figura 51 - Gráfico de pizza que apresenta a proporção de vitórias de cada método globalmente, ou seja, consolidando os resultados globais de cada cenário.

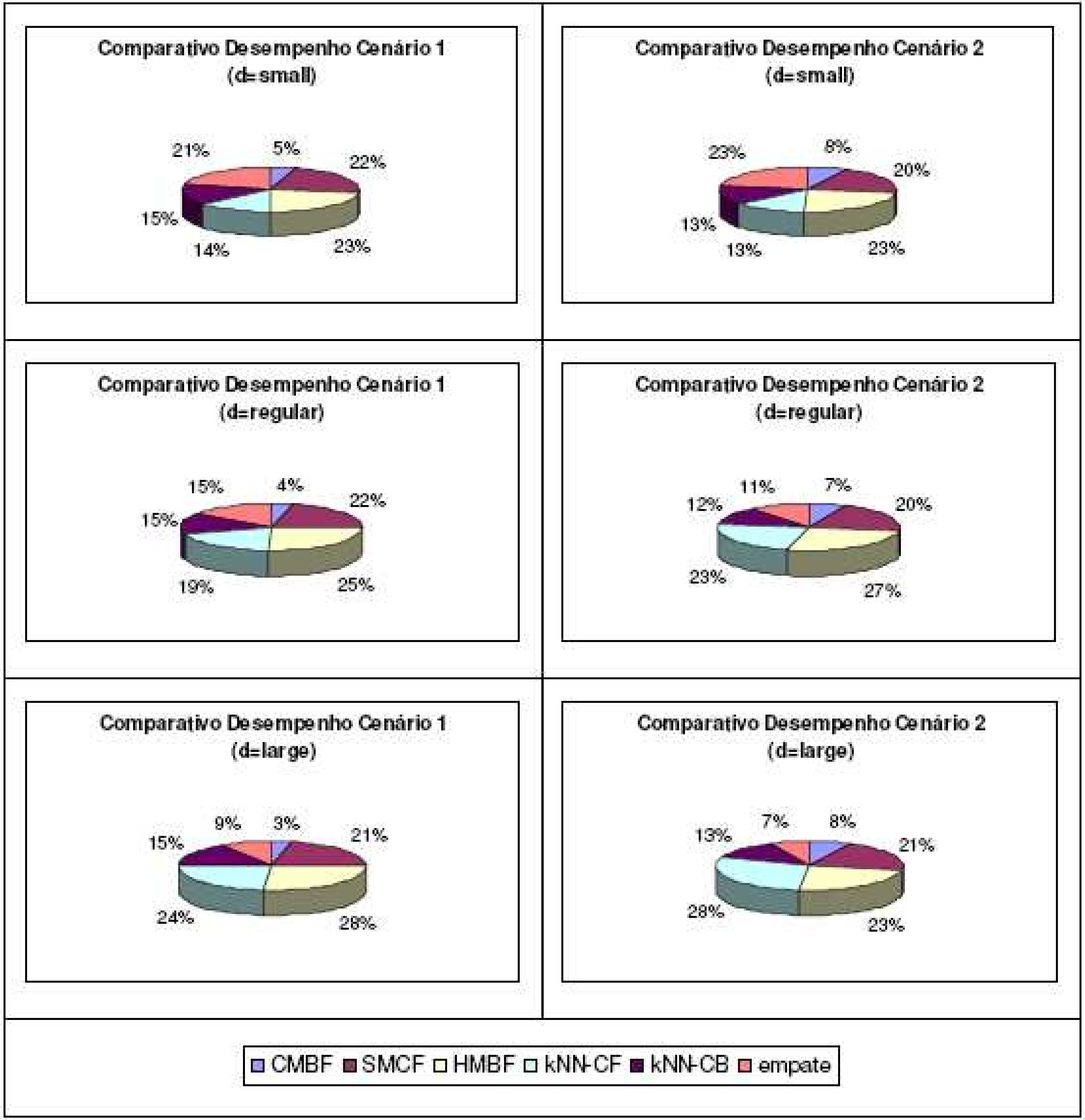
Symbolic Data Analysis Tools for Recommendation Systems

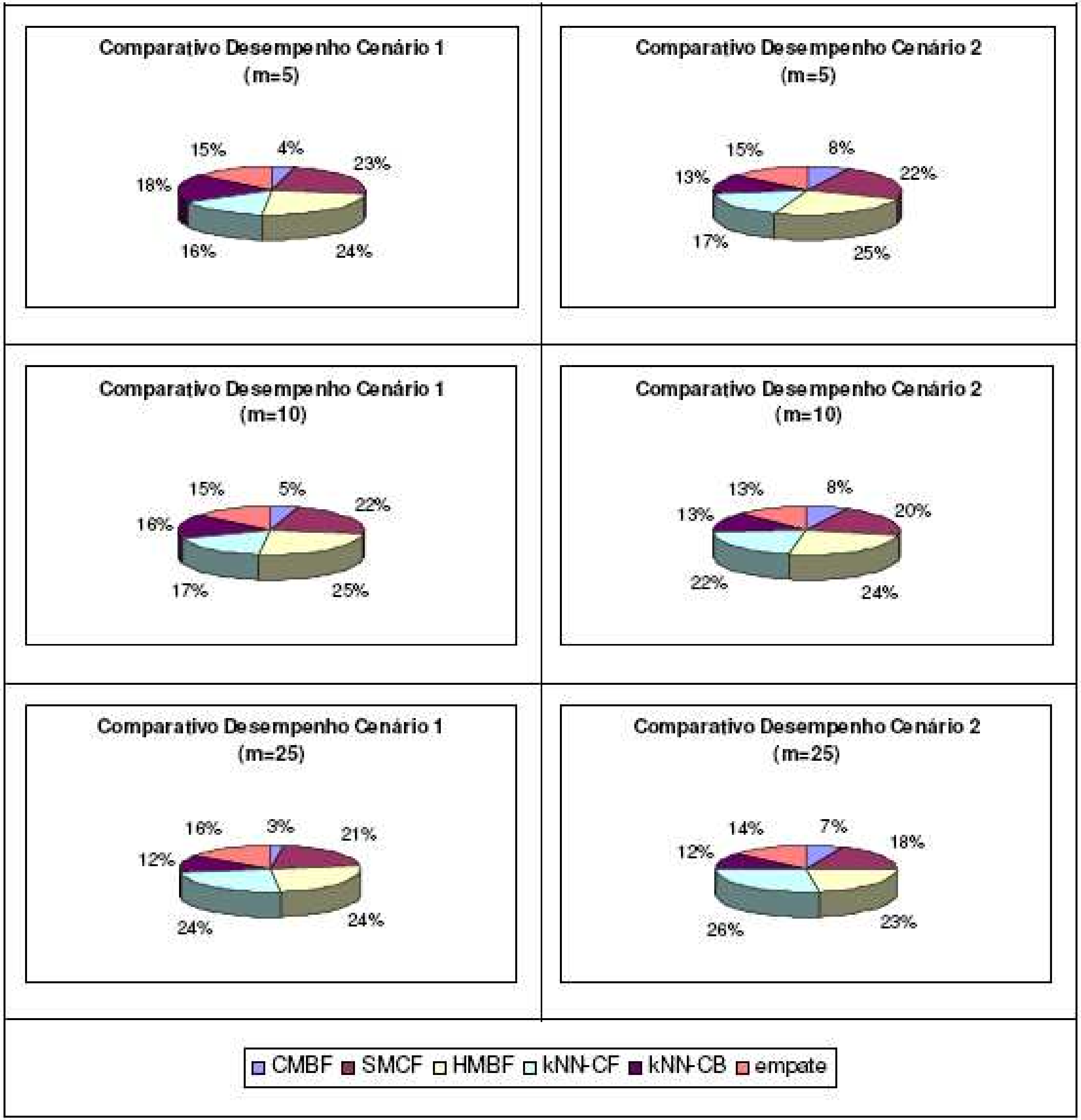
?

Byron Leite Dantas Bezerra
bldb@dsc.upe.br

Francisco Assis Tenório Carvalho
fatc@cin.ufpe.br

Centro de Informática – CIn/UFPE
Recife – Pernambuco – Brasil


Comparativo Desempenho Cenário 1 (d=small)
21%
5%
22%
15%
14%
23%
Comparativo Desempenho Cenário 2 (d=small)
23%
8%
20%
13%
13%
23%
Comparativo Desempenho Cenário 1 (d=regular)
15%
4%
22%
15%
19%
25%
Comparativo Desempenho Cenário 2 (d=regular)
11%
7%
12%
20%
23%
27%
Comparativo Desempenho Cenário 1 (d=large)
9%
3%
15%
21%
24%
28%
Comparativo Desempenho Cenário 2 (d=large)
7%
8%
13%
21%
28%
23%
CMBF SMCF HMBF kNN-CF kNN-CB empate


Comparativo Desempenho Cenário 1
(m=5)
15%
4%
23%
18%
16%
24%
Comparativo Desempenho Cenário 2
(m=5)
15%
8%
22%
13%
17%
25%
Comparativo Desempenho Cenário 1
(m=10)
15%
5%
22%
16%
17%
25%
Comparativo Desempenho Cenário 2
(m=10)
13%
8%
20%
13%
22%
24%
Comparativo Desempenho Cenário 1
(m=25)
16%
3%
21%
12%
24%
24%
Comparativo Desempenho Cenário 2
(m=25)
14%
7%
18%
12%
26%
23%
CMBF
SMCF
HMBF
kNN-CF
kNN-CB
empate

Avaliação Experimental