

Detecção de efeitos colaterais indesejáveis na Internet das coisas - um estudo de caso no Home Network System

Apresentação de Monografia

Heron Sanches Gonçalves Pires Ferreira

Universidade Federal da Bahia

Departamento de Ciência da Computação

Orientadora: Profa. Dra. Daniela Barreiro Claro

Contato: heronsanches@dcc.ufba.br

31 de outubro de 2016

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
 - Resultados
- 5 Conclusão

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
 - Resultados
- 5 Conclusão

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de *25 terabytes/dia* (??).
- Governos e corporações também produzem milhares de **documentos** todos os dias, tais como relatórios, formulários, pesquisas de opiniões e etc.
- ??) ressalta que este cenário está além dos limites humanos para o uso e compreensão.

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de **25 terabytes/dia** (??).
- Governos e corporações também produzem milhares de **documentos** todos os dias, tais como relatórios, formulários, pesquisas de opiniões e etc.
- ??) ressalta que este cenário está além dos limites humanos para o uso e compreensão.

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de **25 terabytes/dia** (??).
- Governos e corporações também produzem milhares de **documentos** todos os dias, tais como relatórios, formulários, pesquisas de opiniões e etc.
- ??) ressalta que este cenário está além dos limites humanos para o uso e compreensão.

Introdução

- O avanço da tecnologia tem proporcionado um **aumento gigantesco** na quantidade de **dados armazenados**.
- A rede social Facebook produz mais de **25 terabytes/dia** (??).
- Governos e corporações também produzem milhares de **documentos** todos os dias, tais como relatórios, formulários, pesquisas de opiniões e etc.
- ??) ressalta que este cenário está além dos limites humanos para o uso e compreensão.

Introdução

- ??) enfatizam que instituições estão sobrecarregadas com o processamento desse montante de dados.
- Os dados possuem diversos tipos e formatos, sendo armazenados de forma estruturada ou não estruturada.

Exemplos

documentos de textos, planilhas, áudios, imagens, vídeos, documentos HTML e etc.

Introdução

- ??) enfatizam que instituições estão sobrecarregadas com o processamento desse montante de dados.
- Os dados possuem diversos tipos e formatos, sendo armazenados de forma estruturada ou **não estruturada**.

Exemplos

documentos de textos, planilhas, áudios, imagens, vídeos, documentos HTML e etc.

Introdução

- ??) enfatizam que instituições estão sobrecarregadas com o processamento desse montante de dados.
- Os dados possuem diversos tipos e formatos, sendo armazenados de forma estruturada ou **não estruturada**.

Exemplos

documentos de textos, planilhas, áudios, imagens, vídeos, documentos HTML e etc.

Introdução

- Dados estruturados já possuem mecanismos eficientes de armazenamento e recuperação.
- Documentos textuais por serem não estruturados são recuperados através de Sistemas de Recuperação da Informação (SRI).

Exemplos

Duckduckgo, Jus Brasil, IEEExplore, ACM, Google e etc

Introdução

- Dados estruturados já possuem mecanismos eficientes de armazenamento e recuperação.
- **Documentos textuais** por serem **não estruturados** são recuperados através de Sistemas de Recuperação da Informação (SRI).

Exemplos

Duckduckgo, Jus Brasil, IEEExplore, ACM, Google e etc

Introdução

- Demanda crescente para desenvolvimento e aprimoramento de métodos que possam processar e **extrair padrões** de **dados textuais**.
- A extração de padrões de documentos textuais é o principal objetivo da Mineração de Textos (MT).

Introdução

Vários desafios estão presentes no processo de extração de padrões de documentos textuais, entre eles destaca-se:

- Não estruturados.
- Naturalmente **imprecisos** e **incertos**.
- Abordam um ou mais temas.
- **Alta dimensionalidade**.
- Dados **esparsos**.

Exemplos

Uma coleção de documentos pode conter 100.000 palavras, enquanto um documento pode conter apenas algumas centenas (??).

Introdução

Vários desafios estão presentes no processo de extração de padrões de documentos textuais, entre eles destaca-se:

- Não estruturados.
- Naturalmente **imprecisos** e **incertos**.
- Abordam um ou mais temas.
- **Alta dimensionalidade**.
- Dados **esparsos**.

Exemplos

Uma coleção de documentos pode conter 100.000 palavras, enquanto um documento pode conter apenas algumas centenas (??).

Introdução

Vários desafios estão presentes no processo de extração de padrões de documentos textuais, entre eles destaca-se:

- Não estruturados.
- Naturalmente **imprecisos** e **incertos**.
- Abordam um ou mais temas.
- **Alta dimensionalidade**.
- Dados **esparsos**.

Exemplos

Uma coleção de documentos pode conter 100.000 palavras, enquanto um documento pode conter apenas algumas centenas (??).

Introdução

Vários desafios estão presentes no processo de extração de padrões de documentos textuais, entre eles destaca-se:

- Não estruturados.
- Naturalmente **imprecisos** e **incertos**.
- Abordam um ou mais temas.
- **Alta dimensionalidade**.
- Dados **esparsos**.

Exemplos

Uma coleção de documentos pode conter 100.000 palavras, enquanto um documento pode conter apenas algumas centenas (??).

Introdução

Vários desafios estão presentes no processo de extração de padrões de documentos textuais, entre eles destaca-se:

- Não estruturados.
- Naturalmente **imprecisos** e **incertos**.
- Abordam um ou mais temas.
- **Alta dimensionalidade**.
- Dados **esparsos**.

Exemplos

Uma coleção de documentos pode conter 100.000 palavras, enquanto um documento pode conter apenas algumas centenas (??).

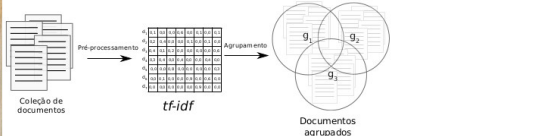


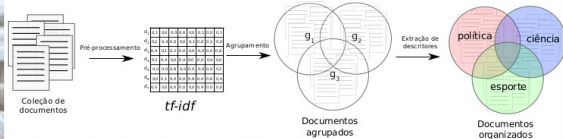
Coleção de
documentos





GOOD DATA?





Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- Grupos com significado relevante.
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- Considerar a imprecisão e a incerteza.
- Reduzir a influência de documentos ruidosos.

Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- **Grupos com significado relevante.**
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- Considerar a imprecisão e a incerteza.
- Reduzir a influência de **documentos ruidosos**.

Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- **Grupos com significado relevante.**
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- Considerar a imprecisão e a incerteza.
- Reduzir a influência de **documentos ruidosos**.

Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- **Grupos com significado relevante.**
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- Considerar a imprecisão e a incerteza.
- Reduzir a influência de **documentos ruidosos**.

Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- **Grupos com significado relevante.**
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- Considerar a imprecisão e a incerteza.
- Reduzir a influência de **documentos ruidosos**.

Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- **Grupos com significado relevante.**
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- **Considerar a imprecisão e a incerteza.**
- Reduzir a influência de **documentos ruidosos**.

Introdução

O agrupamento é muito importante neste processo e possui uma série de desafios:

- Agrupar de acordo com a similaridade.
- **Grupos com significado relevante.**
- Escalável para grandes coleções (*Big Data*).
- Baixo custo computacional.
- Estimar os parâmetros dos algoritmos.
- **Considerar a imprecisão e a incerteza.**
- Reduzir a influência de **documentos ruidosos**.

Introdução

Citação

[...] não é esperado que um único método de agrupamento atenda todas as exigências para todos os conjuntos de dados [...] (??).

Introdução

Existem diversos métodos de agrupamento na literatura, os quais destacam-se:

- *Fuzzy C-Means* (FCM)
- *Possibilistic C-Means* (PCM)
- *Possibilistic Fuzzy C-Means* (PFCM)

Introdução

A partir das investigações conduzidas foi proposto dois métodos de extração de descritores:

- **Possibilistic Description Comes Last (PDCL)**
- **Mixed - Possibilistic Fuzzy Description Comes Last (Mixed-PFDCL) (Híbrido)**

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
 - Resultados
- 5 Conclusão

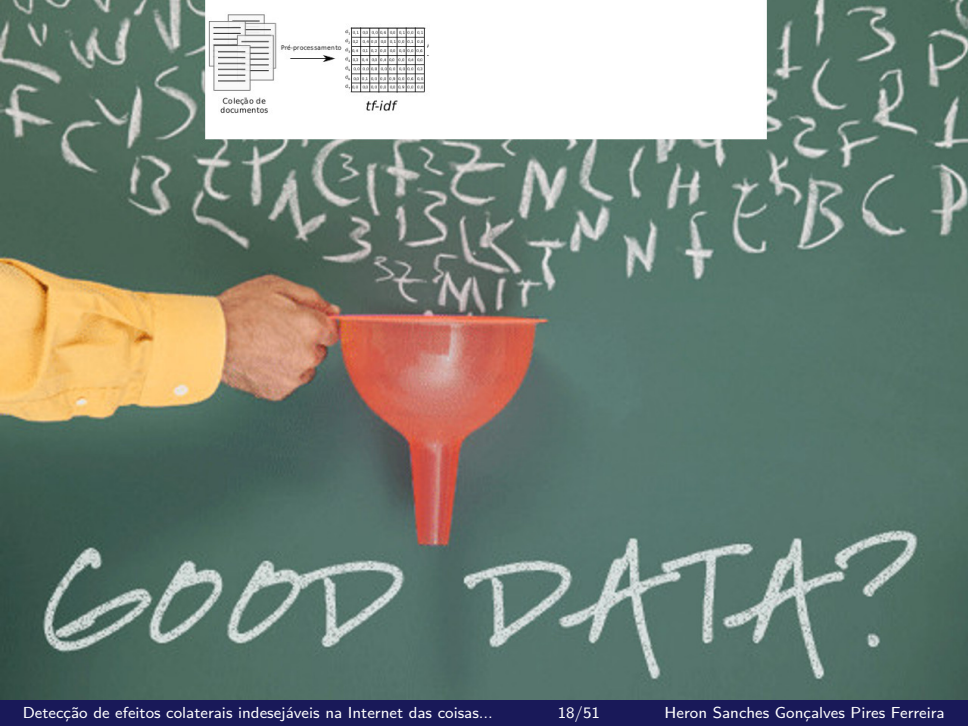


Diagram illustrating the process of document collection and transformation into a matrix:

Coleção de documentos → Pré-processamento → *tf-idf*

d_1	0.1	0.3	0.4	0.5	0.2	0.1
d_2	0.4	0.2	0.2	0.3	0.2	0.3
d_3	0.4	0.1	0.2	0.2	0.4	0.4
d_4	0.2	0.1	0.3	0.4	0.3	0.4
d_5	0.1	0.1	0.4	0.2	0.2	0.2
d_6	0.4	0.3	0.2	0.2	0.1	0.2

GOOD DATA?

Detecção de efeitos colaterais indesejáveis na Internet das coisas... 18/51 Heron Sanches Gonçalves Pires Ferreira

Pré-processamento

- Remoção de espaços.
- Expansão de abreviações.
- Remoção de *stopwords* (pronomes, artigos e etc.).
- Lematização (Casa → Cas).
- Estruturação dos documentos (TF-IDF).

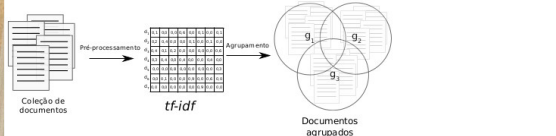
	<i>termo₁</i>	<i>termo₂</i>	<i>termo₃</i>
<i>doc₁</i>	1	3	4
<i>doc₂</i>	9	2	0

Tabela: Exemplo matriz docs x termos



	<i>termo₁</i>	<i>termo₂</i>	<i>termo₃</i>
<i>doc₁</i>	0.1	0.6	1.0
<i>doc₂</i>	0.9	0.4	0.0

Tabela: Exemplo matriz tf-idf



Agrupamento

- Organizar objetos similares em um mesmo grupo.
- Grupos crisp x fuzzy
- Coeficiente de similaridade de cosseno.
- Validação do agrupamento com o método silhueta fuzzy.

Agrupamento

- Organizar objetos similares em um mesmo grupo.
- Grupos crisp x fuzzy
- Coeficiente de similaridade de cosseno.
- Validação do agrupamento com o método silhueta fuzzy.

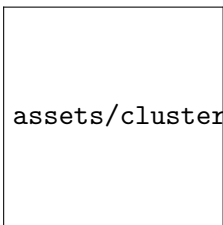


Imagem: Grupos crisp

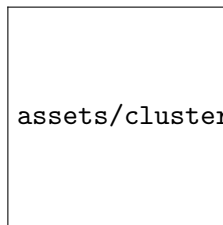


Imagem: Grupos fuzzy

Agrupamento

- Organizar objetos similares em um mesmo grupo.
- Grupos crisp x fuzzy
- Coeficiente de similaridade de cosseno.
- Validação do agrupamento com o método silhueta fuzzy.

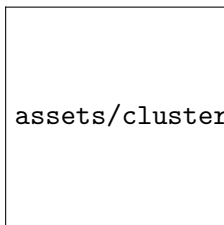


Imagem: Grupos crisp

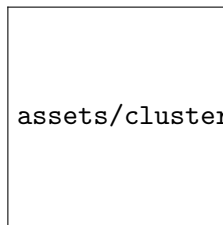


Imagem: Grupos fuzzy

Agrupamento

- Organizar objetos similares em um mesmo grupo.
- Grupos crisp x fuzzy
- Coeficiente de similaridade de cosseno.
- Validação do agrupamento com o método silhueta fuzzy.

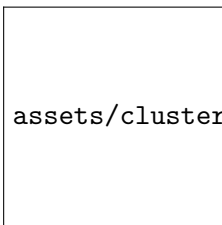


Imagem: Grupos crisp

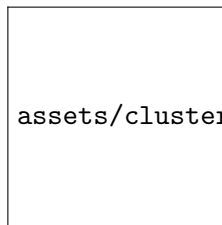


Imagem: Grupos fuzzy

Agrupamento (FCM) (??)

- Graus de pertinência.
- **Restrição probabilística.**
- Problema com ruídos.

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	0,5	0,5	1,0
<i>doc₂</i>	0,5	0,5	1,0

Tabela: Pertinências FCM

assets/fcm_problem.pdf

Imagem: Problema dos ruídos

Agrupamento (PCM) (??)

- Graus de tipicidade.
- Remoção da restrição probabilística.
- Problema dos grupos coincidentes.

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	0,7	0,7	1,4
<i>doc₂</i>	0,2	0,2	0,4

Tabela: Tipicidades PCM

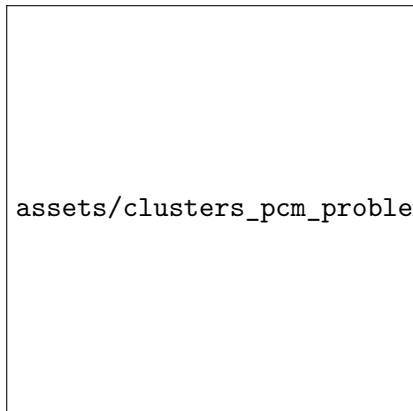
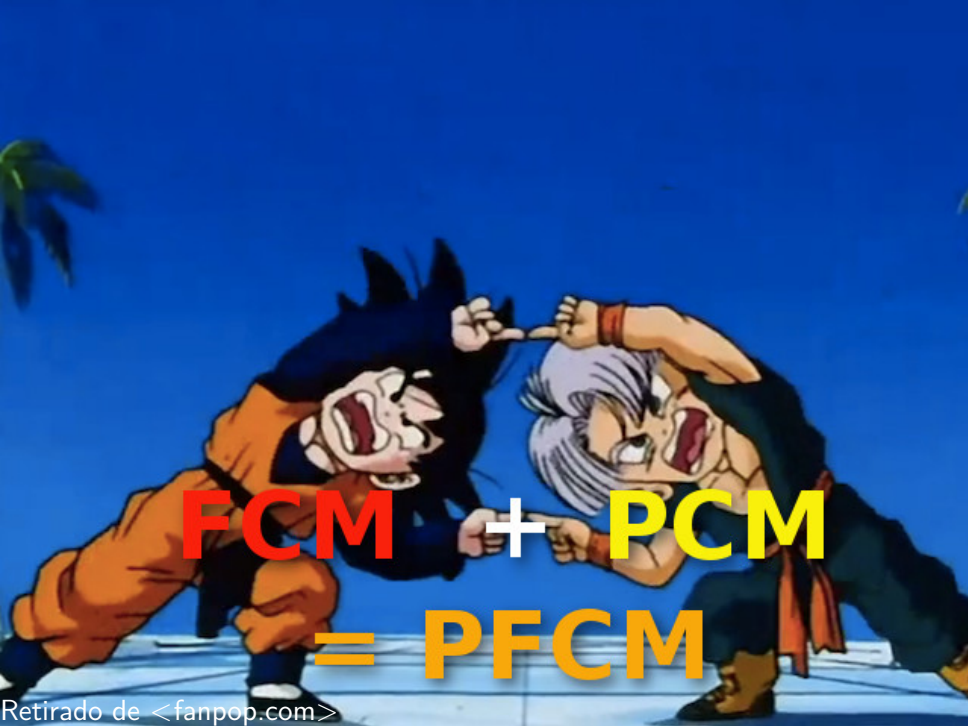


Imagem: Grupos coincidentes



FCM + PCM
= PFCM

Agrupamento (PFCM) (??)

- Pertinências e tipicidades.
- Robustez.
- Parâmetros de ponderação a e b .

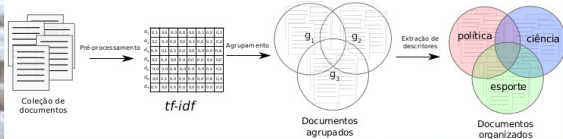
assets/samples_pfcmm.png

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	0,5	0,5	1,0
<i>doc₂</i>	0,5	0,5	1,0

Tabela: Pertinências PFCM

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	0,7	0,7	1,4
<i>doc₂</i>	0,2	0,2	0,4

Tabela: Tipicidades PFCM



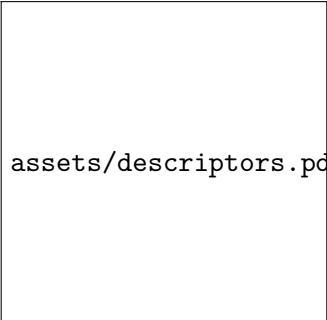
Extração de descritores

- Atribuir significados aos grupos.
- Manual ou **Automatizada**.
- Abordagens de conhecimento interno e externo.
- Durante o agrupamento (*Description Comes First* - DCF)
- **Após o agrupamento** (*Description Comes Last* - **DCL**).
- Método *Soft Organization - Fuzzy Description Comes Last* (SoftO-FDCL) (??).

assets/descriptors.pdf

Extração de descritores

- Atribuir significados aos grupos.
- Manual ou **Automatizada**.
- Abordagens de conhecimento interno e externo.
- Durante o agrupamento (*Description Comes First* - DCF)
- **Após o agrupamento** (*Description Comes Last* - **DCL**).
- Método *Soft Organization - Fuzzy Description Comes Last* (SoftO-FDCL) (??).



assets/descriptors.pdf

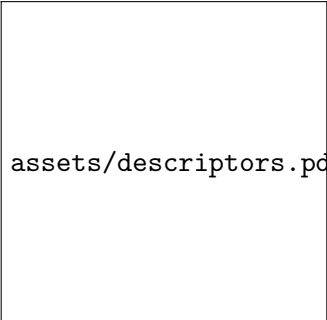
Extração de descritores

- Atribuir significados aos grupos.
- Manual ou **Automatizada**.
- Abordagens de conhecimento interno e externo.
- Durante o agrupamento (*Description Comes First* - DCF)
- **Após o agrupamento** (*Description Comes Last* - **DCL**).
- Método *Soft Organization - Fuzzy Description Comes Last* (SoftO-FDCL) (??).

assets/descriptors.pdf

Extração de descritores

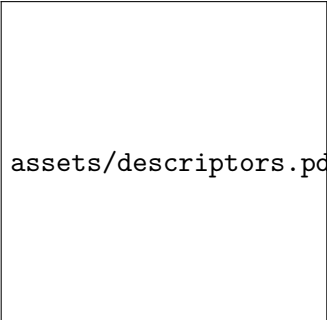
- Atribuir significados aos grupos.
- Manual ou **Automatizada**.
- Abordagens de conhecimento interno e externo.
- Durante o agrupamento (*Description Comes First* - DCF)
- **Após o agrupamento** (*Description Comes Last* - **DCL**).
- Método *Soft Organization - Fuzzy Description Comes Last* (SoftO-FDCL) (??).



assets/descriptors.pdf

Extração de descritores

- Atribuir significados aos grupos.
- Manual ou **Automatizada**.
- Abordagens de conhecimento interno e externo.
- Durante o agrupamento (*Description Comes First* - DCF)
- **Após o agrupamento** (*Description Comes Last* - **DCL**).
- Método *Soft Organization - Fuzzy Description Comes Last* (SoftO-FDCL) (??).



assets/descriptors.pdf

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
 - Resultados
- 5 Conclusão

Trabalhos relacionados

- ??) propões uma abordagem incremental hierárquica para construção de tópicos.
- ??) e ??) explora otimizações para *Big Data*.
- ??) propõe uma maneira de estabilizar a inicialização do agrupamento.
- ??) utiliza o agrupamento ainda na fase de pré-processamento.

Trabalhos relacionados

- ??) propões uma abordagem incremental hierárquica para construção de tópicos.
- ??) e ??) explora otimizações para *Big Data*.
- ??) propõe uma maneira de estabilizar a inicialização do agrupamento.
- ??) utiliza o agrupamento ainda na fase de pré-processamento.

Trabalhos relacionados

- ??) propões uma abordagem incremental hierárquica para construção de tópicos.
- ??) e ??) explora otimizações para *Big Data*.
- ??) propõe uma maneira de estabilizar a inicialização do agrupamento.
- ??) utiliza o agrupamento ainda na fase de pré-processamento.

Trabalhos relacionados

- ??) propões uma abordagem incremental hierárquica para construção de tópicos.
- ??) e ??) explora otimizações para *Big Data*.
- ??) propõe uma maneira de estabilizar a inicialização do agrupamento.
- ??) utiliza o agrupamento ainda na fase de pré-processamento.

Trabalhos relacionados

- ??) combina os algoritmos genéticos no agrupamento para evitar os mínimos locais.
- ??) propõe uma medida de similaridade com informações semânticas.
- ??) traz uma abordagem de extração de descritores independente do agrupamento.

Trabalhos relacionados

- ??) combina os algoritmos genéticos no agrupamento para evitar os mínimos locais.
- ??) propõe uma medida de similaridade com informações semânticas.
- ??) traz uma abordagem de extração de descritores independente do agrupamento.

Trabalhos relacionados

- ??) combina os algoritmos genéticos no agrupamento para evitar os mínimos locais.
- ??) propõe uma medida de similaridade com informações semânticas.
- ??) traz uma abordagem de extração de descritores independente do agrupamento.

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
 - Resultados
- 5 Conclusão

Coleções textuais

Coleção	docs	termos	classes	% zeros	n-gramas
Opinosis	51	842	3	95,73%	1-grama
20newsgroups	2000	11028	4	99,11%	1-grama
Hitech	600	6925	6	97,93%	1-grama
NSF	1600	2806	16	99,76%	1-grama
WAP	1560	8070	20	98,51%	1-grama
Reuters-21578	1052	3925	43	98,55%	1-grama

Tabela: Características das coleções textuais utilizadas nesta pesquisa

Refinamento com PFCM

assets/process_pfcmm.pdf

Refinamento com PFCM

Coleção	# classes	FCM	PCM	PFCM
Opinosis	3	3	3	3
20Newsgroup	4	2	2	2
Hitech	6	6	5	5
NSF	16	11	2	16
WAP	20	14	5	16
Reuters-21578	43	22	11	36

Tabela: Quantidade ótima de grupos determinada através do método da silhueta fuzzy para cada algoritmo de agrupamento

Refinamento com PFCM

Coleção	docs	termos	FCM	PCM	PFCM
Opinosis	51	842		✓	
20newsgroups	2000	11028			✓
Hitech	600	6925	✓		
NSF	1600	2806	✓		
WAP	1560	8070			✓
Reuters-21578	1052	3925	✓		

Tabela: Sumário dos resultados da classificação dos descritores

Refinamento com PFCM

Método	<i>crisp</i> ₁	<i>crisp</i> ₂	<i>crisp</i> ₃
FCM	drive, display, control, car , work	import, model , problem, unit, design	breakfast, concierge , coffee, food, inn
PCM	read, problem, car , work, found	turn, size, qual- ity, review, fea- ture	extreme, drive, point, reason, run
PFCM μ	drive, control, version, car , work	read, complete, device , display, size	breakfast, pleasant, concierge , coffee, clean
PFCM λ	club, immacu- late, towel, pil- low, fridge	housekeep, tourist, tea, smoke, london	bottle, adult, food, reserve, dinner

Tabela: Descritores extraídos com os métodos de agrupamento FCM,

Refinamento com PFCM - Discussão

- FCM e PFCM capturaram melhor a estrutura das coleções.
- Capacidade de adaptação do método SoftO-FDCL.
- Descritores fuzzy mais significativos.
- **Descritores possibilísticos pouco significativos.**
- Dimensionalidade aparenta influenciar os resultados.

Método SoftO-FDCL (??)

	$\mu(d_i, g_j) \geq \delta, \forall d_i$	$\mu(d_i, g_j) < \delta, \forall d_i$
$t_k \in d_i, \forall d_i$	<i>ganhos</i>	<i>ruídos</i>
$t_k \notin d_i, \forall d_i$	<i>perdas</i>	<i>rejeitos</i>

Tabela: Matriz de contingência do método SoftO-FDCL

$$\text{precisão}(t_k, g_j) = \frac{\text{ganhos}}{\text{ganhos} + \text{ruídos}} \quad (1)$$

$$\text{recuperação}(t_k, g_j) = \frac{\text{ganhos}}{\text{ganhos} + \text{perdas}} \quad (2)$$

$$f1(t_k, g_j) = \frac{2 * \text{precisão}(t_k, g_j) * \text{recuperação}(t_k, g_j)}{\text{precisão}(t_k, g_j) + \text{recuperação}(t_k, g_j)} \quad (3)$$

Método SoftO-FDCL (??)

	<i>grupo₁</i>	<i>grupo₂</i>
<i>termo₁</i>	0.85	0.75
<i>termo₂</i>	0.35	0.95
<i>termo₃</i>	0.25	0.55
<i>termo₄</i>	0.80	0.65
<i>termo₅</i>	0.50	0.50
<i>termo₆</i>	0.30	0.24
<i>termo₇</i>	0.10	0.83

Tabela: Pontuação dos termos obtidas com a medida f1

Método SoftO-FDCL (??)

	<i>grupo₁</i>	<i>grupo₂</i>
<i>termo₁</i>	0.85	0.75
<i>termo₂</i>	0.35	0.95
<i>termo₃</i>	0.25	0.55
<i>termo₄</i>	0.80	0.65
<i>termo₅</i>	0.50	0.50
<i>termo₆</i>	0.30	0.24
<i>termo₇</i>	0.10	0.83

Tabela: Descritores de maior pontuação em cada grupo

O limiar é adequado?

$$\delta = \frac{1}{\text{total de grupos}} = \frac{1}{2} = 0,5 \quad (4)$$

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	0,4	0,6	1,0
<i>doc₂</i>	0,8	0,2	1,0

Tabela: Pertinências PFCM

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	0,6	0,9	1,5
<i>doc₂</i>	0,4	0,1	0,5

Tabela: Tipicidades PFCM

Convertendo tipicidades em pertinências

Tipicidades → Pertinências

$$\lambda'(d_i, g_j) = \frac{\lambda(d_i, g_j)}{\sum_{k=1}^c \lambda(d_i, g_k)} \quad (5)$$

	<i>grupo₁</i>	<i>grupo₂</i>	total
<i>doc₁</i>	$\frac{0,6}{1,5} = 0,4$	$\frac{0,9}{1,5} = 0,6$	1,0
<i>doc₂</i>	$\frac{0,4}{0,5} = 0,8$	$\frac{0,1}{0,5} = 0,2$	1,0

Tabela: Tipicidades → Pertinências

Método PDCL

	$\lambda'(d_i, g_j) \geq \delta, \forall d_i$	$\lambda'(d_i, g_j) < \delta, \forall d_i$
$t_k \in d_i, \forall d_i$	<i>ganhos</i>	<i>ruídos</i>
$t_k \notin d_i, \forall d_i$	<i>perdas</i>	<i>rejeitos</i>

Tabela: Matriz de contingência do método PDCL

Ponderando os ganhos, ruídos, perdas e rejeitos

$$\text{ganhos}(t_k, g_j) = \sum_{d_i \in D'} \lambda(d_i, g_j) \quad (6)$$

$$D' = \{d_i \mid \text{SE } \lambda'(d_i, g_j) \geq \delta \text{ E } t_k \in d_i \text{ PARA } \forall d_i\} \quad (7)$$

Método Mixed-PFDCL

`assets/process_pfdcl.pdf`

Método Mixed-PFDCL

`assets/process_pfdcl.pdf`

Método Mixed-PFDCL

`assets/process_pfdcl.pdf`

Resultados

	PCM		PFCM	
Coleção	SoftO-FDCL	PDCL	SoftO-FDCL	Mixed
Opinosis		✓		✓
20newsgroups	✓	✓		✓
Hitech		✓		✓
NSF	✓	✓		✓
WAP		✓		✓
Reuters-21578		✓		✓

Tabela: Sumário dos resultados da classificação dos descritores extraídos com os métodos SoftO-FDCL, PDCL e Mixed-PFDCL

Resultados

	<i>grupo₁</i>		<i>grupo₂</i>	
Método	termo	pontuação	termo	pontuação
SoftO-FDCL	caf	0,923077	caf	0,923077
	floor	0.888889	floor	0.888889
	food	0.880000	food	0.880000
	coffe	0.857143	coffe	0.857143
	concierge	0.846154	concierge	0.846154
PDCL	bathro	0.894716	make	0.800980
	food	0.888785	time	0.789846
	concierge	0.860127	nice	0.779338
	supermarket	0.856632	feature	0.778138
	chain	0.856632	easy	0.768564

Tabela: 5 termos de maior pontuação obtidos extraídos com os métodos Soft-FDCL e PDCL da coleção Opinosis com o algoritmo PCM

Resultados - Discussão

- Demonstram a adequação da interpretação proposta.
- O método SoftO-FDCL pode gerar pontuações similares a partir das tipicidades.
- Os métodos propostos resolvem o problema de pontuações similares.
- Os métodos PDCL e Mixed-PFDCL superaram o método SoftO-FDCL.
- O método Mixed-PFDCL se mostrou adequado para a organização híbrida de documentos.

Conteúdo

- 1 Introdução
- 2 Fundamentação Teórica
 - Pré-processamento
 - Agrupamento (FCM,PCM,PFCM)
 - Extração de descritores
- 3 Trabalhos relacionados
- 4 Abordagem proposta
 - Refinamento com PFCM
 - Método PDCL
 - Método Mixed-PFDCL
 - Resultados
- 5 Conclusão

Conclusão

- A organização flexível de documentos envolve muitos campos de estudo.
- Detalhamento dos métodos de agrupamento FCM, PCM, PFCM e HFCM.
- É possível aprimorar todas as etapas do processo.
- Impactos do PFCM na organização flexível de documentos.

Conclusão

- A organização flexível de documentos envolve muitos campos de estudo.
- Detalhamento dos métodos de agrupamento FCM, PCM, PFCM e HFCM.
- É possível aprimorar todas as etapas do processo.
- Impactos do PFCM na organização flexível de documentos.

Conclusão

- A organização flexível de documentos envolve muitos campos de estudo.
- Detalhamento dos métodos de agrupamento FCM, PCM, PFCM e HFCM.
- É possível aprimorar todas as etapas do processo.
- Impactos do PFCM na organização flexível de documentos.

Conclusão

- A organização flexível de documentos envolve muitos campos de estudo.
- Detalhamento dos métodos de agrupamento FCM, PCM, PFCM e HFCM.
- É possível aprimorar todas as etapas do processo.
- Impactos do PFCM na organização flexível de documentos.

Conclusão

- **Propriedades do limiar δ .**
- A estratégia híbrida se mostrou adequada, produzindo bons descritores.
- Os métodos propostos obtiveram bons resultados.
- Publicação de artigo científico na conferência FUZZ-IEEE.

Conclusão

- **Propriedades do limiar δ .**
- **A estratégia híbrida se mostrou adequada, produzindo bons descritores.**
- Os métodos propostos obtiveram bons resultados.
- Publicação de artigo científico na conferência FUZZ-IEEE.

Conclusão

- **Propriedades do limiar δ .**
- **A estratégia híbrida se mostrou adequada, produzindo bons descritores.**
- **Os métodos propostos obtiveram bons resultados.**
- Publicação de artigo científico na conferência FUZZ-IEEE.

Conclusão

- **Propriedades do limiar δ .**
- **A estratégia híbrida se mostrou adequada, produzindo bons descritores.**
- **Os métodos propostos obtiveram bons resultados.**
- **Publicação de artigo científico na conferência FUZZ-IEEE.**

Referências I