Ministério da
Ciência e Tecnologia

sid.inpe.br/mtc-m19/2010/12.03.13.37-RPQ

MINERAÇÃO DE DADOS PARA IDENTIFICAR AGRUPAMENTOS DE ESTAÇÕES METEOROLÓGICAS USANDO DADOS HISTÓRICOS DE PRECIPITAÇÃO

José Roberto Motta Garcia

Relatório final da disciplina Princípios e Aplicações de Mineração de Dados (CAP-359) do Programa de Pós-Graduação em Computação Aplicada, ministrada pelo professor Dr. Rafael Santos

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/38N29FH >

INPE
São José dos Campos
2010

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/6921
Fax: (012) 3208-6919
E-mail: pubtc@sid.inpe.br

Ministério da
Ciência e Tecnologia

sid.inpe.br/mtc-m19/2010/12.03.13.37-RPQ

MINERAÇÃO DE DADOS PARA IDENTIFICAR AGRUPAMENTOS DE ESTAÇÕES METEOROLÓGICAS USANDO DADOS HISTÓRICOS DE PRECIPITAÇÃO

José Roberto Motta Garcia

Relatório final da disciplina Princípios e Aplicações de Mineração de Dados (CAP-359) do Programa de Pós-Graduação em Computação Aplicada, ministrada pelo professor Dr. Rafael Santos

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/38N29FH >

INPE
São José dos Campos
2010

RESUMO

O controle de qualidade de dados climatológicos recebidos por estações meteorológicas é requisito essencial para que as pesquisas baseadas nesses dados sejam confiáveis. Uma forma de realizar este controle é comparar cada dado recebido com dados de outras estações com comportamentos similares. Este trabalho tem como objetivo identificar agrupamentos de estações meteorológicas com base na similaridade de dados históricos de precipitação segundo uma janela de tempo previamente selecionada. Para isso serão apresentados como os dados foram tratados para que esta janela fosse exposta assim como o algoritmo de agrupamento e a sua visualização espacial.

LISTA DE FIGURAS

Pág.

SUMÁRIO

Pág.

1 INTRODUÇÃO

O controle de qualidade dos dados oriundos de estações meteorológicas e recebidos pelo CPTEC não possuem garantia de estarem 100% corretos. Embora a quantidade de dados suspeitos ser em pequena quantidade, cabe aos administradores criar mecanismos de aferição para minimizar esse acontecimento. Muitos dos dados podem sofrer com sujeira no trajeto da transmissão, podem ser delatados erroneamente por equipamentos defeituosos e até mesmo podem ser submetidos a erros de digitação e/ou medição manual dos sensores.

Para amenizar este problema o CPTEC possui alguns mecanismos de verificação e classificação dos dados recebidos. Um deles se baseia em limites máximos e mínimos mensais das variáveis meteorológicas que atuam em quadrantes geográficos delimitados por latitude e longitude e são estabelecidos por meteorologistas. Veja Figura 1.1.

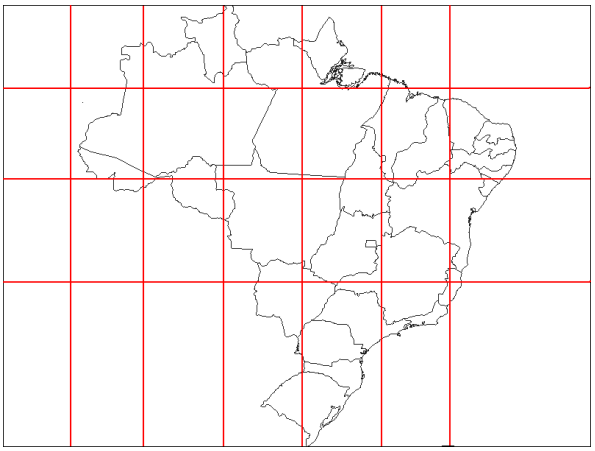

Figura 1.1 – Representação não precisa dos quadrantes em que o controle de qualidade dos dados climatológicos do CPTEC está baseado

É sabido, porém que a natureza não é compatível com o tipo de divisão acima definido, o que gera a necessidade de elaboração de outros mecanismos que ajudem a aumentar a confiabilidade dos dados.

Outra abordagem de resolver este problema é a comparação dos dados recebidos por uma estação com os dados recebidos de estações com comportamentos similares, assim, este trabalho tem como objetivo encontrar similaridades entre o comportamento das estações meteorológicas para que possa haver esta comparação segundo uma determinada forma de classificação.

O objetivo inicial deste trabalho era investigar estas similaridades nas regiões Sul e Sudeste do Brasil para realizar comparações dos resultados com (CARDOSO & SILVA DIAS, 2002), porém para produzir um resultado no tempo hábil da disciplina foram considerados apenas dados históricos de precipitação no estado de São Paulo. As dificuldades encontradas durante a execução deste assim como sua continuidade em trabalhos futuros serão relatadas mais adiante.

2 O TRABALHO

Como este trabalho se baseia em dados históricos logo imagina-se que o acervo de dados é muito grande e portanto serão aplicadas técnicas de descoberta de conhecimento em banco de dados (ou *KDD – Knowledge Discovery in Databases*) conforme as etapas definidas em (Fayad, 1996).

São ao todo 9 (nove) etapas iterativas para a completa realização do trabalho e a iteratividade se dá no momento em que pode-se voltar para etapas anteriores a qualquer momento em se julgue necessário. As etapas 1 a 4 foram refeitas algumas vezes para que os dados se tornassem manipuláveis e projetáveis.

2.1. Compreensão do domínio da aplicação, do conhecimento existente e objetivos do usuário final

A etapa de escolha do domínio da aplicação surgiu a partir do conhecimento do trabalho de (CARDOSO & SILVA DIAS, 2002) e da facilidade de obtenção dos dados para a realização do trabalho.

2.2. Escolha de subconjunto dos dados

O plano original era trabalhar com os dados de precipitação diária das regiões Sul e Sudeste do Brasil. Para tanto foi construída uma aplicação em Java para extrair os dados do Banco de Dados Climatológico do CPTEC, que está armazenado no sistema Oracle e convertê-los para um banco de dados de acesso apenas local desenvolvido sob o sistema PostgresSQL.

2.3. Limpeza e pré-processamento dos dados

Após uma verificação inicial do conteúdo do acervo do CPTEC notou-se que havia mais de 115 (cento e quinze) milhões de dados diários originados de inúmeras estações meteorológicas, em todos os tempos.

É sabido que o domínio deste acervo é a América Latina e se estende até parte da Antártida, com um predomínio de 96% de dados em território brasileiro. Para evitar que a memória fosse esgotada em uma recuperação do conjunto completo de dados, esta foi dividida em fatias de latitude com variações de 2 graus, sem se importar com a longitude, sendo que os limites das latitudes mínima e máxima fossem suficientes para englobar os estados que compõem as regiões Sul e Sudeste, a Figura 2.1 mostra como a recuperação dos dados foi dividida.

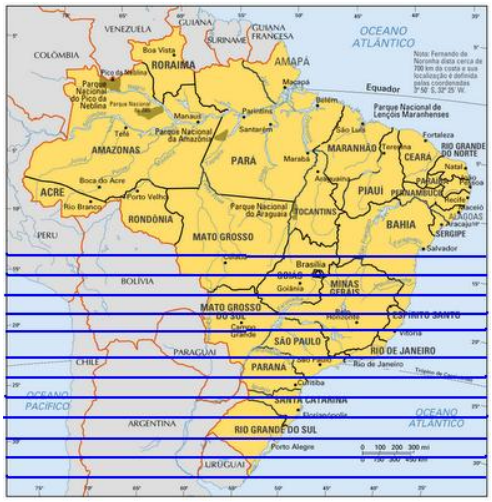

Figura 2.1 – Representação não precisa das faixas de recuperação de dados de precipitação do banco de dados do CPTEC
Fonte: adaptado de http://forumdefesa.com/forum/viewtopic.php?f=11&t=1776&start=60

Inicialmente a aplicação de transferência de dados de um banco para outro era feita de forma direta, ou seja, assim que o conjunto de dados da faixa de latitude fosse recuperado do banco Oracle ele já era armazenado no banco PostgreSQL entretanto este processo consumia muito tempo (quase 24 hs) e houve problemas de incompatibilidade de horário pois o banco Oracle é parado toda madrugada para backup o que gerava perda de conexão do programa de transferência.

A solução encontrada foi de gerar arquivos CSV e teve muito mais desempenho, gastando apenas algumas poucas horas (+/- 5 hs). Uma vez de posse dos arquivos, a tarefa foi de lê-los e gravar no banco de dados PostgreSQL destino. A tabela destino contém a estrutura abaixo informada:

```
lat, n, 7, 4
lon, n, 7, 4
data, date
valor, n 10,2
```

2.4. Redução e projeção dos dados

Motivados pelo fato de que um dos requisitos básicos para realizar um trabalho de mineração de dados é partir de um conjunto de dados coerente e completo sob pena de obter resultados errados, pensou-se numa maneira de projetá-los em forma de “mapa de pixels” para que o período com quantidade de dados mais razoável para a realização do trabalho pudesse ser escolhido visualmente, este deveria ser construído conforme os requisitos abaixo:

- Deveria ser visualizado em tela
- Deveria ter em seu eixo Y todas as estações meteorológicas envolvidas
- Deveria ter em seu eixo X, uma representação do espaço temporal
- Cada dado do banco deveria ser plotado como um pixel na seguinte posição:
    - Na linha referente a estação meteorológica a qual pertence (referenciado pelo conjunto lat + lon)
    - Na coluna referente a data de quando foi reportado

A princípio esta parecia ser uma boa solução para visualizar o “mapa dos dados”, porém ao realizar a conta da granularidade do eixo X verificou-se que esta ficaria muito grande que seriam necessárias algumas reduções no conjunto de dados até que ele ficasse praticável para o trabalho, segue abaixo uma relação e detalhamento de fases destas reduções.

FASE: 1 – GRAVAR DADOS DOS ARQUIVOS ASCII
DOMÍNIO ESPACIAL DA TABELA: Todos os dados entre latitude -40 e -10
TOTAL DE LINHAS DA TABELA: 72 milhões
TEMPO APROXIMADO DAS OPERAÇÕES SQL: 24 hs
PONTOS NO EIXO X: 122 anos x 12 x 30 = ~44 mil
PROCEDIMENTO RESPONSÁVEL: Programa Java que leu arquivos ASCII e gravou na tabela

FASE: 2 – REDUZIR ESPACIALIDADE
DOMÍNIO ESPACIAL DA TABELA: Todos os dados das regiões Sul e Sudeste
TOTAL DE LINHAS DA TABELA: 44 milhões
TEMPO APROXIMADO DAS OPERAÇÕES SQL: 4 hs
PONTOS NO EIXO X: 122 anos x 12 x 30 = ~44 mil
PROCEDIMENTO RESPONSÁVEL: Programa Java que leu arquivos ASCII gravou na tabela comparando localização geográfica através do PostGIS (POSTGIS, 2010).

FASE: 3 – REDUZIR ESPACIALIDADE
DOMÍNIO ESPACIAL DA TABELA: Todos os dados das regiões Sudeste
TOTAL DE LINHAS DA TABELA: 30 milhões
TEMPO APROXIMADO DAS OPERAÇÕES SQL: 1 h
PONTOS NO EIXO X: 122 anos x 12 x 30 = ~44 mil
PROCEDIMENTO RESPONSÁVEL: Programa Java que leu arquivos ASCII gravou na tabela comparando localização geográfica através do PostGIS.

FASE: 4 – REDUZIR ESPACIALIDADE E ACUMULAR MENSALMENTE
DOMÍNIO ESPACIAL DA TABELA: Todos os dados do estado de São Paulo
TOTAL DE LINHAS DA TABELA: 1 milhão
TEMPO APROXIMADO DAS OPERAÇÕES SQL: 1 h
PONTOS NO EIXO X: 122 anos x 12 = 1464
PROCEDIMENTO RESPONSÁVEL: Programa Java que leu arquivos ASCII gravou na tabela comparando localização geográfica através do PostGIS e comando SQL que acumulou dados diários em mensais e contabilizando quantidade de dias existentes.

Uma vez que foi conseguido que o eixo X tivesse uma melhor granularidade para ser mostrada, foi construído um aplicativo em Java que atende os requisitos do Mapa de Pixels e possui algumas funcionalidades que melhoram a operação do aplicativo. O Aplicativo foi batizado, a princípio, como PrecipExplorer e o mapa de pixel foi construído utilizando programação com Java2D (SUN, 2010).

A Figura 2.2 mostra as funcionalidades essenciais que ajudam na escolha do melhor período de dados além do Mapa de Pixels do período completo dos dados considerados após todas as reduções.

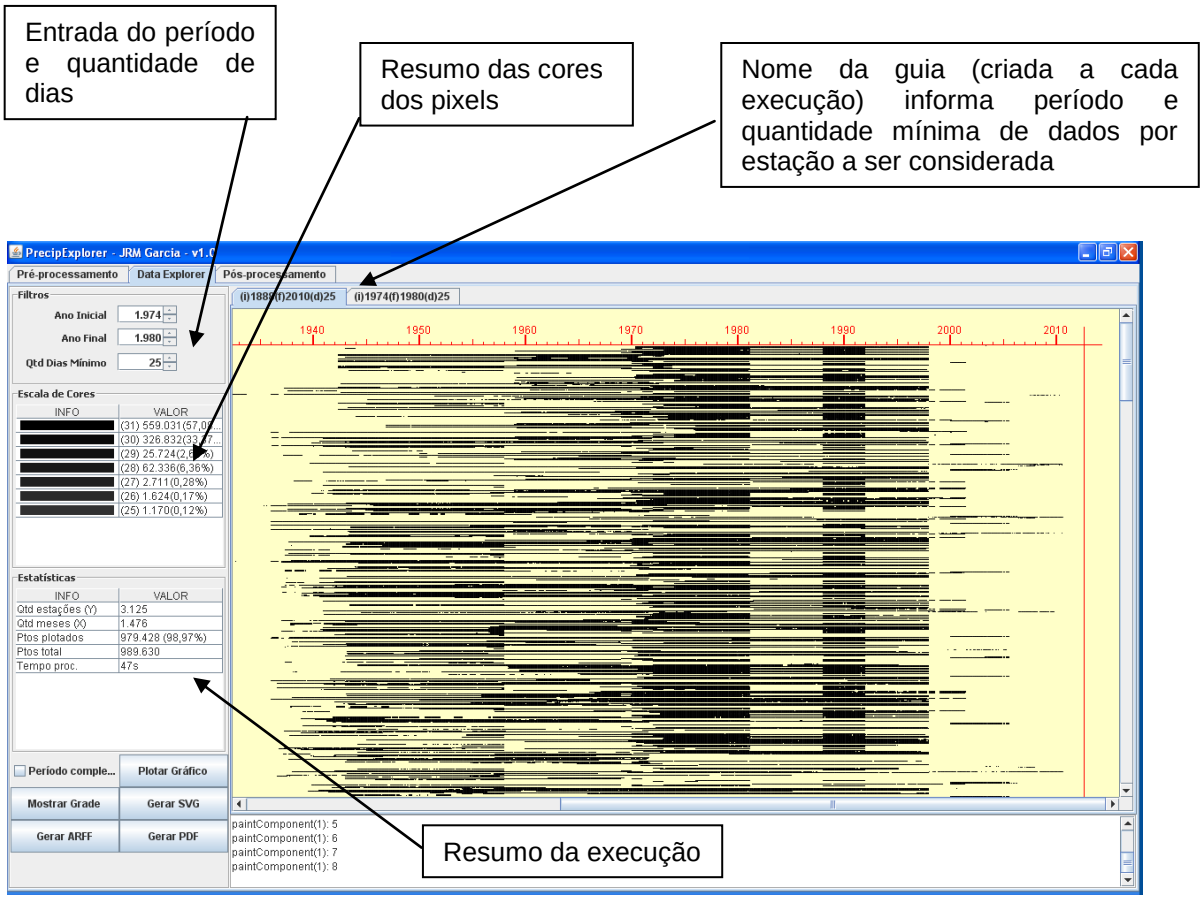

Figura 2.2 – Janela do mapa dos dados de precipitação mensal do CPTEC no período de 1888 a 2010 considerando apenas estações com mais de 25 dias

Dentre outras particularidades fora do escopo deste trabalho, pode ser notado que, um dos períodos que poderia ser usado para a análise de similaridade entre estações era a janela entre 1974 e 1980. O resultado da execução é mostrado pela Figura 2.3.

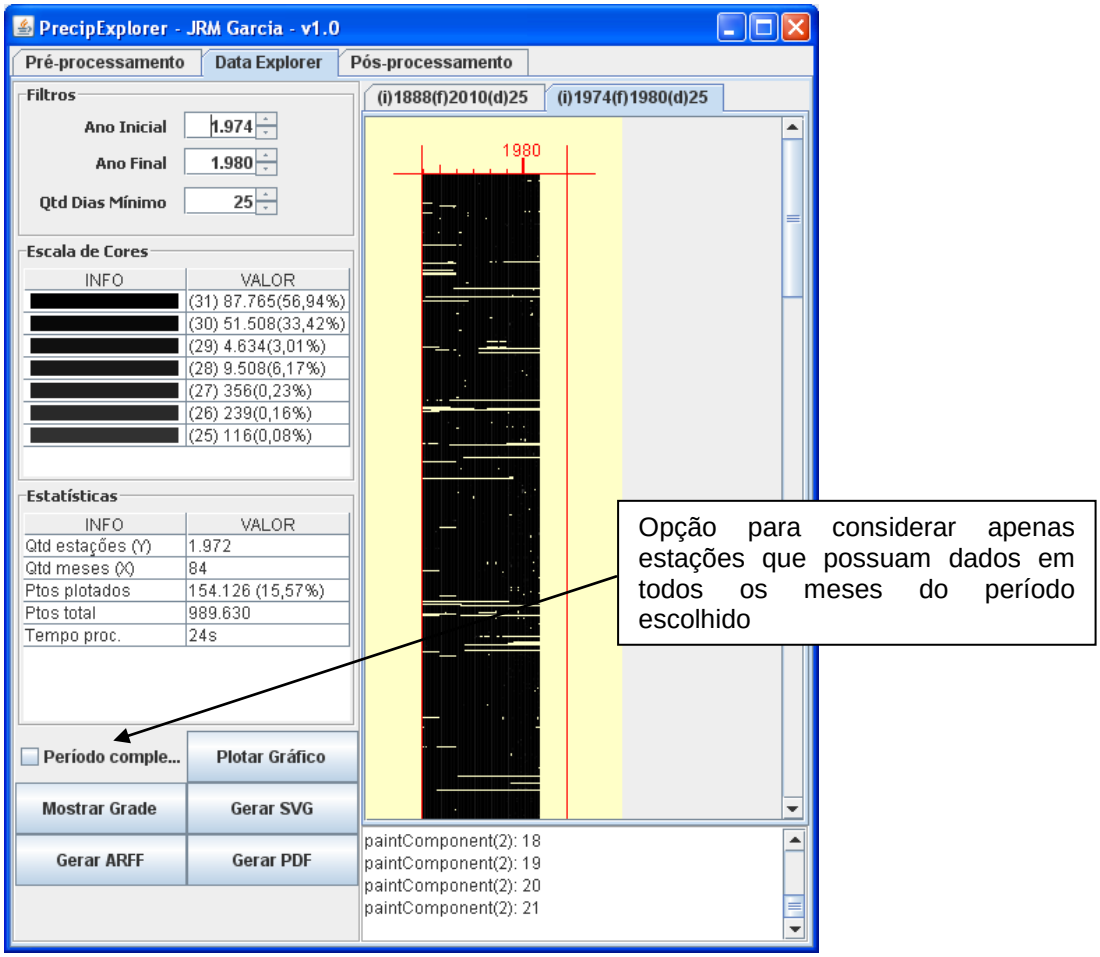

Figura 2.3 – Janela do mapa dos dados de precipitação mensal do CPTEC no período de 1974 a 1980 considerando apenas estações com mais de 25 dias

Como este é um trabalho de Mineração de Dados, utilizamos para isso o software WEKA (WAIKATO, 2010). Uma das formas que ele aceita dados de *input* é através de arquivos ARFF. O PrecipExplorer foi programado para gerar este arquivo a partir do Mapa de Pixels que ele está visualizando no momento e, como podemos notar na Figura 2.3, há linhas ou pedaços de linhas em branco espalhadas no mapa em questão, que significam períodos que não possui dados (menos que 25 dias no mês, no caso). Considerando que um bom trabalho de mineração de dados começa por um conjunto de dados coerentes a tarefa seria “rodar” novamente este período e pedir que o PrecipExplorer considere apenas estações que possuam dados em todos os meses do período solicitado. O resultado é visualizado na Figura 2.4.

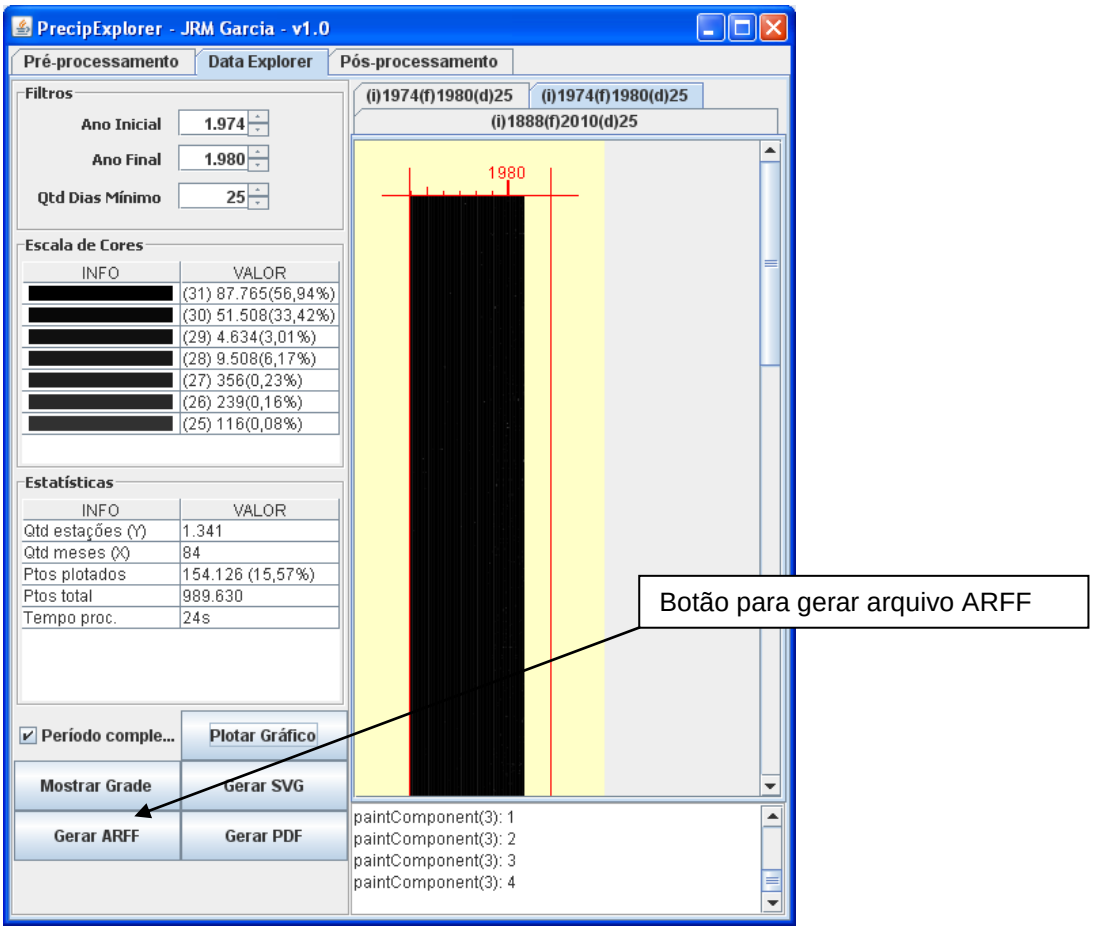

Figura 2.4 – Janela do mapa dos dados de precipitação mensal do CPTEC no período de 1974 a 1980 considerando apenas estações com mais de 25 dias e que possuem dados em todos os meses do período escolhido

Conforme é indicado no Resumo da Execução da Figura 2.4, são 1.341 estações (que formam o eixo Y) e 84 meses (que formam o eixo X).

2.5. Escolha da tarefa de mineração dos dados

O arquivo ARFF possui uma estrutura toda particular e o arquivo gerado pode ser assim resumido:

```
@relation preciptation0    # Rótulo da execução

@attribute 197401 numeric  # Início da relação de atributos
@attribute 197402 numeric
.....
@attribute 198011 numeric
@attribute 198012 numeric  # Fim da relação de atributos

@data                      # Início da seção de dados
264.8,72.4,439,121.8,17.1,61.6,0,...,249.3,278.3
230.3,100.1,342.4,91.6,19.7,70.7,...,152,272.3
...
359.9,118.4,308.7,105,5.5,56.9,...,201.9,163.2
```

A relação de atributos refere-se a quantidade de colunas de dados que serão analisadas, ou seja, o arquivo exemplo utilizado possui 84 atributos referentes aos meses entre os anos 1974 e 1980 (inclusive), e são identificados pela máscara AAAAMM. Assim, cada linha de dado deve conter 84 ocorrências de valores.

Após a criação do arquivo ARFF é necessário carregá-lo no WEKA, conforme a Figura 2.5.

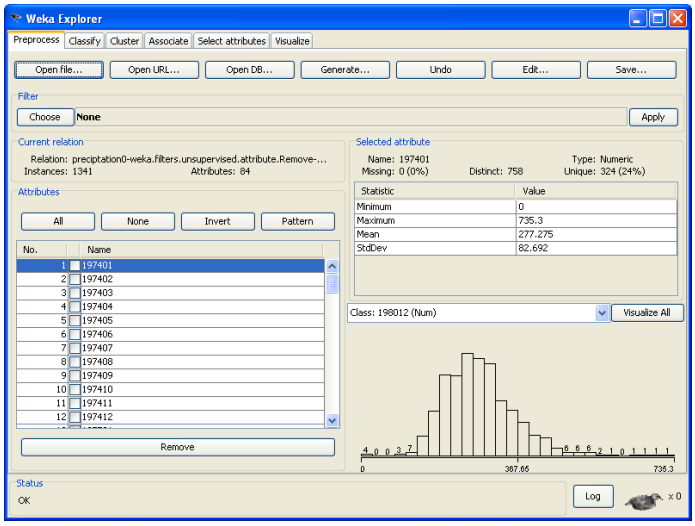

Figura 2.5 – Tela do WEKA mostrando o arquivo ARFF carregado

O objetivo do trabalho é investigar similaridades entre estações meteorológicas e, portanto, realizar agrupamentos de acordo com os critérios adotados. Esta tarefa é realizada pela aba Cluster no WEKA, conforme a Figura 2.6.

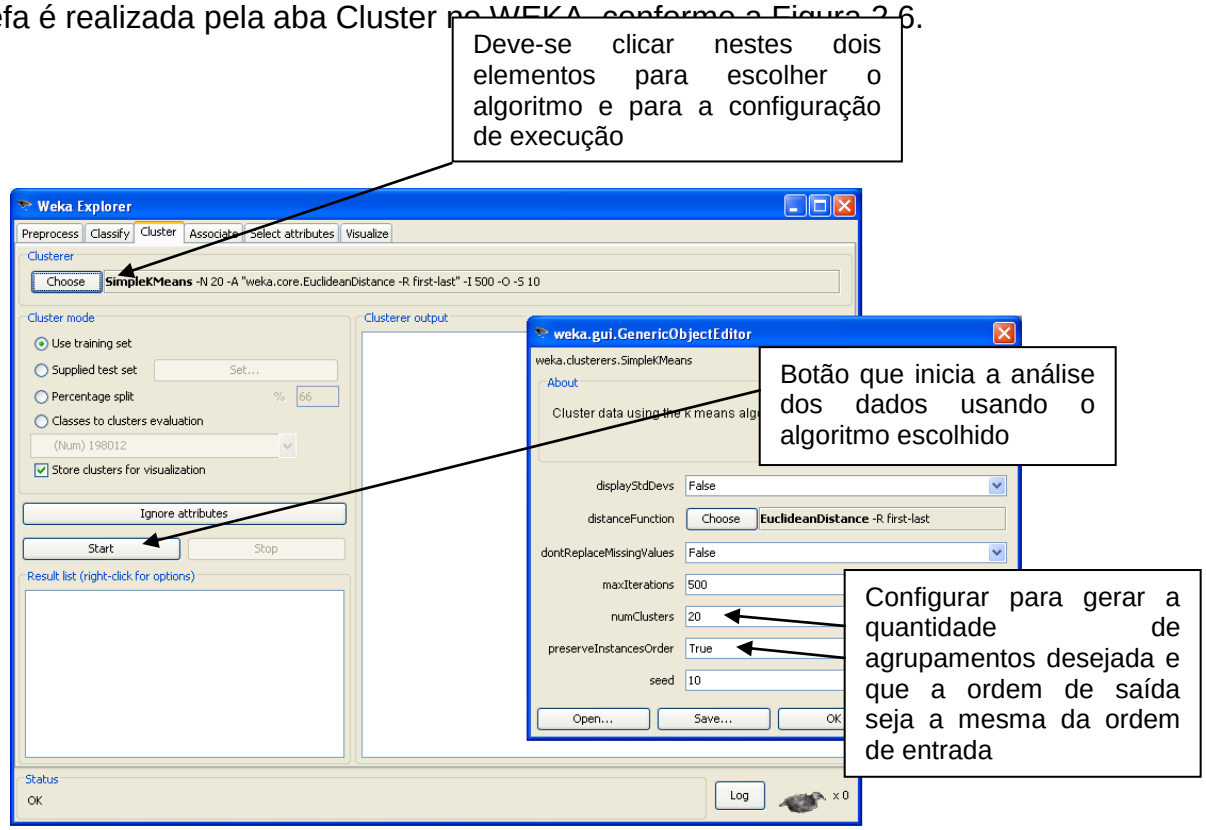

Figura 2.6 – Tela do WEKA mostrando detalhes da operação de clusterização e do algoritmo escolhido

2.6. Escolha do algoritmo de mineração dos dados

Conforme ilustrado na Figura 2.6, deve-se clicar no botão “Choose” para escolher o algoritmo que será responsável para fazer os agrupamentos, em nosso trabalho utilizamos o K-Médias (ALAG, 2009) que recebe o nome de *SimpleKMeans* no WEKA.

2.7. Mineração dos dados

Para realizar a mineração de dados, deve-se levar em consideração que:

- O arquivo ARFF que o PrecipExplorer gera, contem a latitude e a longitude como primeiro e segundo atributo de cada linha, para que se tenha uma referência de qual estação meteorológica os dados se referem;

- Este arquivo é carregado no WEKA e as informações de latitude e longitude são retiradas manualmente do arquivo para que os números não influenciem no resultado da aplicação do algoritmo;

A partir da escolha do algoritmo alguns passos ainda eram necessários antes de iniciar a mineração de dados propriamente dita:

2.7.1. Configuração da execução

Antes de executar com este novo conjunto de dados (sem latitude e longitude) deve-se optar por manter a ordem de saída de acordo com a ordem de entrada dos dados, como na mostra a Figura 2.6. Isso é necessário para garantir que a ordem de saída dos dados seja a mesma do arquivo original gerado pelo PrecipExplorer (com latitude e longitude), com isso pode-se depois unir os dois arquivos e gerar uma saída completa com a localização geográfica das

estações meteorológicas e a identificação do cluster a qual ela pertence (gerada após executar o WEKA).

2.7.2. Execução

A mineração dos dados pode ser então iniciada pelo botão *Start* como mostrou a Figura 2.6. Após a execução, que leva alguns instantes, uma entrada (linha) é criada no quadro *Result List*, juntamente com o detalhamento no quadro *Output*, ver Figura 2.7.

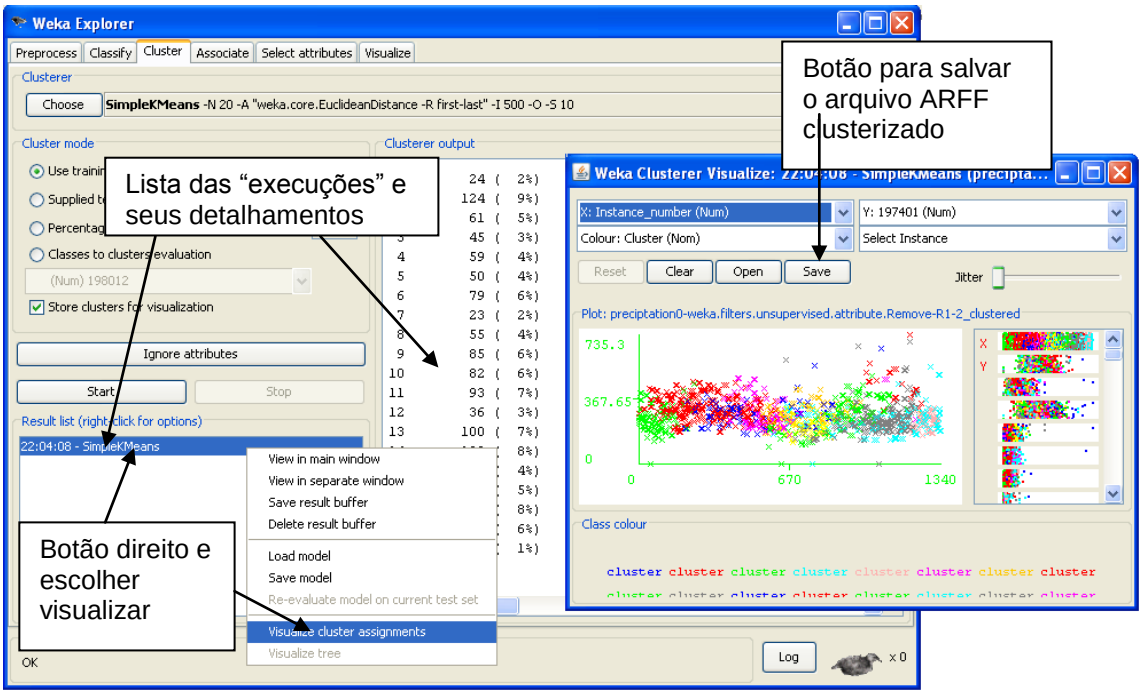

Figura 2.7 – Tela do WEKA mostrando como proceder com a gravação do arquivo com a clusterização

2.7.3. Salvar o arquivo clusterizado gerado

Após a execução deve-se clicar com o botão direito do mouse na entrada criada e escolher *Visualize cluster assignments*. A tela que surge não contém informações relevantes para este trabalho e uma vez que se deseja que o

arquivo ARFF seja gerado com informações do cluster em que cada ocorrência pertence então deve-se clicar no botão *Save*, como mostra a Figura 2.7

O novo arquivo ARFF gerado ficou como exemplificado abaixo:

```
@relation preciptation0-clustered     # Rótulo da execução

@attribute Instance_number numeric   # Início  atrib
@attribute 197401 numeric
@attribute 197402 numeric
...
@attribute 198011 numeric
@attribute 198012 numeric
@attribute Cluster {cluster0,cluster1,...,cluster19}

@data    # Início dos dados
0,264.8,72.4,439,121.8,17.1,..,95.2,249.3,cluster8
1,230.3,100.1,342.4,91.6,...7,60.4,69.2,172.3,cluster4
...
1339,182.5,163,585,131.4,...,172.8,291.4,349,cluster3
1340,182.5,163,585,131.4,...172.8,291.4,19,cluster17
```

Algumas diferenças criadas pelo WEKA podem ser notadas deste arquivo em relação ao original:

- A palavra *clustered* foi adicionada ao nome
- Foi incluído um atributo de ordenação na primeira posição
- Foi incluído um atributo que indica a qual cluster a ocorrência pertence como último atributo.

2.7.4. Merge

Como o interesse deste trabalho é descobrir similaridades entre estações meteorológicas e estas possuem uma localização espacial pontual no plano cartesiano então a visualização estaria adequadamente representada usando-se o mapa da região em questão, o estado de São Paulo, e nele fossem

plotados os pontos das estações meteorológicas identificados de acordo com o cluster em que foi categorizado.

Para construir este cenário foi criada uma aplicação em Java que lê o arquivo ARFF original (com latitude e longitude) e o arquivo clusterizado. Para cada linha de dado ele consegue recuperar as latitudes e longitudes do arquivo original e o cluster em que a estação foi classificada do arquivo gerado pelo WEKA, pois a ordem de saída foi mantida quando a execução do processo foi configurada (ver item 2.7.1).

Para armazenar estes dados foi criada uma tabela no banco PostgreSQL com a seguinte estrutura:

```
qtd_cluster numeric  → quant. de clusters da execução
id_cluster numeric   → identif. cluster da estação
sp_latlon geometry   → localização espacial da estação
```

A quantidade de clusters da execução é necessária, pois, como veremos nos resultados, foram feitos testes com 8, 10, 15, 20 e 25 agrupamentos (clusters) e, deste modo, consegue-se armazenar várias execuções numa mesma tabela. Para cada execução, foi gerado um arquivo ARFF que foi unido com o arquivo ARFF original.

2.7.5. A visualização dos resultados

Para plotar o mapa e os pontos das estações meteorológicas foi usada a tecnologia WMS (WMS, 2010) da OGC (OGC, 2010). Esta especificação permite mostrar tanto dados em vetoriais (como o mapa do contorno do estado de São Paulo) como dados pontuais (como as estações meteorológicas).

O contorno do estado de São Paulo foi mostrado a partir de um *shapefile* (ESRI, 1998) e as estações meteorológicas a partir da tabela gerada no sistema PostgreSQL, que contém suas localizações geográficas.

Para diferenciar a classificação de cada estação em forma de pontos coloridos foi utilizado um artifício de classes onde é informada a coluna da tabela que se deseja classificar e uma cor é associada a cada valor diferente encontrado nela.

As Figuras 2.8 a 2.12 mostram os diferentes agrupamentos encontrados.

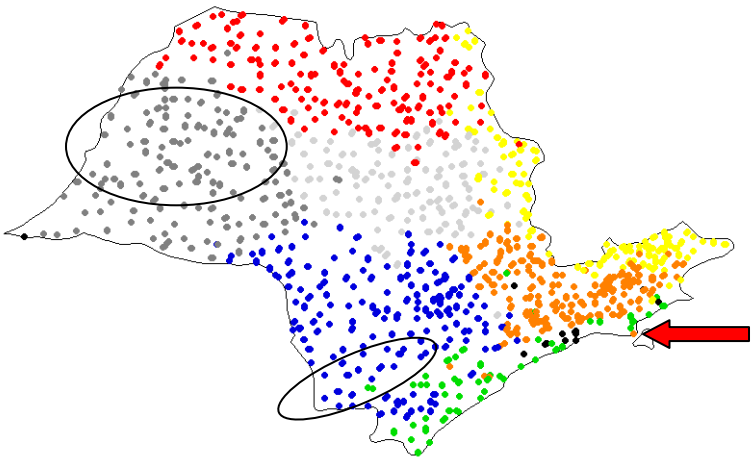

Figura 2.8 – Similaridade das estações meteorológicas considerando precipitação acumulada em 8 agrupamentos

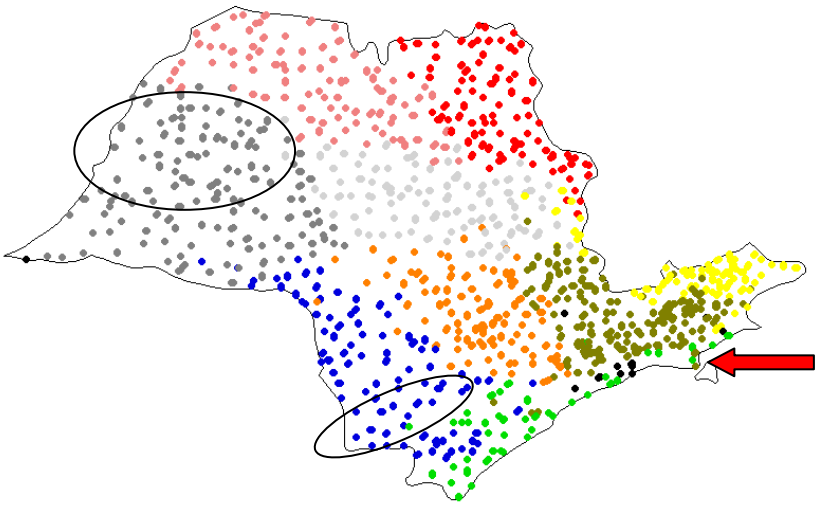

Figura 2.9 – Similaridade das estações meteorológicas considerando precipitação acumulada em 10 agrupamentos

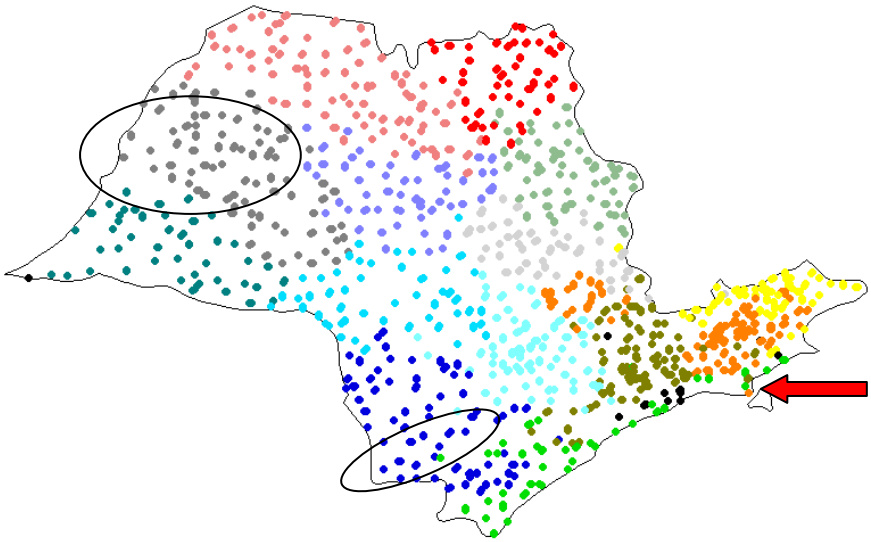

Figura 2.10 – Similaridade das estações meteorológicas considerando precipitação acumulada em 15 agrupamentos

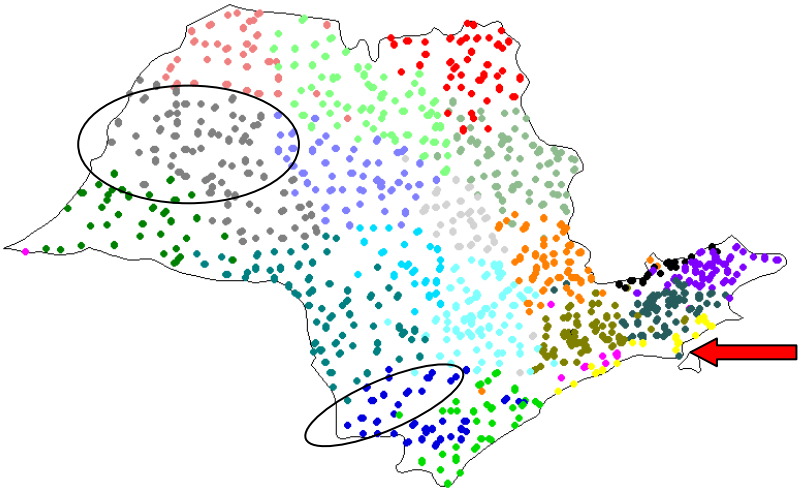

Figura 2.11 – Similaridade das estações meteorológicas considerando precipitação acumulada em 20 agrupamentos

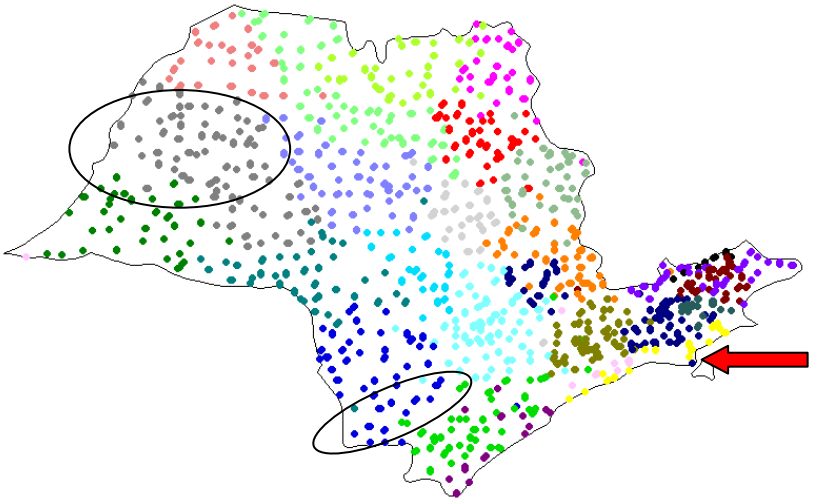

Figura 2.12 – Similaridade das estações meteorológicas considerando precipitação acumulada em 25 agrupamentos

2.8. Interpretação dos resultados

As elipses marcadas sobre o mapa das Figuras 2.8 a 2.12 mostram um exemplo de regiões geográficas que, não importando a quantidade de agrupamentos usada pelo algoritmo, sempre mantiveram as estações meteorológicas num mesmo grupo, que é considerado pelo autor como sendo um forte indício de que realmente estas estações são similares quanto a precipitação acumulada.

Outro ponto marcado nas figuras é a seta vermelha, onde há uma estação meteorológica que insiste em ser diferente das demais ao seu redor, o que indica que ela se comporta como *outlier*, ou seja, possui um valor discrepante em relação aos valores ao seu redor.

Estes apenas são alguns pontos que, a princípio, ficaram mais claros de serem identificados, o que sugere que uma observação mais apurada sobre as imagens seria bem vinda para tentar detectar outros pontos.

2.9. Consolidação do conhecimento obtido e trabalhos futuros

Para que o conhecimento adquirido possa ser utilizado na prática é preciso formar uma parceria com algum especialista em meteorologia que ajudaria na analise das aglomerações.

3 CONCLUSÃO

Como o objetivo do trabalho era melhorar o sistema de controle de qualidade o autor acredita que este trabalho funcionou como um protótipo, ou seja, cumpriu seu papel mostrando que a metodologia funciona mas tem muito a contribuir com implementações futuras, como:

- O envolvimento de outras variáveis poderia ser aplicado para verificar se as similaridades se mantêm;
- A topografia poderia ser incluída no mapa para verificar se há um relacionamento entre os agrupamentos e a topografia;
- O domínio espacial poderia ser ampliado para estudar outras regiões.

O aplicativo criado (PrecipExplorer) poderia evoluir nos seguintes itens:

- Testar outros algoritmos além do K-Médias;
- Implantar os algoritmos sem precisar compor arquivos usando o WEKA;
- Incorporar funções de geoprocessamento utilizando alguma biblioteca (ex. GeoTools) para mostrar os resultados da execução.

REFERÊNCIAS BIBLIOGRÁFICAS

ALAG, S. **Colective Intelligence in Action**. Greenwich: Manning Publications Co, 2009. 397p. ISBN. 1933988312

CARDOSO, A.O.; SILVA DIAS, P.L. Identificação de trimestres extremos no regime pluviométrico do Sul e Sudeste do Brasil em relação com a TSM. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, XII, 2002, Foz do Iguaçu/PR. SBMet, 2002.

ESRI. **ESRI Shapefile Technical Description** (PDF). USA: 1998. Disponível em: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf. Acesso em: 29 Nov 2010.

JAVA.SUN.COM, **Java 2D API**, 2010. Disponível em: http://java.sun.com/products/java-media/2D/index.jsp. Acesso em: 29 Nov 2010.

USAMA M. FAYYAD, GREGORY PIATETSKY-SHAPIRO, PADHRAIC SMYTH AND RAMASAMY UTHURUSAMY, **Advances in Knowledge Discovery and Data Mining**, MIT Press, 1996.

WWW.OGC.ORG, **Welcome**, 2010. Disponível em: http://www.opengeospatial.org/. Acesso em: 29 Nov 2010.

WWW.OPENGEOSPATIAL.ORG, **Overview**, 2010. Disponível em: http://www.opengeospatial.org/standards/wms. Acesso em: 29 Nov 2010.

WWW.ORACLE.COM, **Oracle Database**, 2010. Disponível em: http://www.oracle.com/us/products/database/index.html. Acesso em: 29 Nov 2010.

WWW.CS.WAIKATO.AC.NZ, **Weka3: Data Mining Software in Java**, 2010. Disponível em: http://www.cs.waikato.ac.nz/ml/weka. Acesso em: 27 Nov 2010.

WWW.POSTGIS.ORG, **What is PostGIS?**, 2010. Disponível em: http://postgis.org. Acesso em: 27 Nov 2010.

WWW.POSTGRES.ORG, **About**, 2010. Disponível em: http://postgres.org/about. Acesso em: 29 Nov 2010.

APÊNDICE A – RESUMO DAS TAREFAS PARA GERAR MAPA DE SIMILARIDADES DIVIDIDO POR TECNOLOGIA