

Projeto de Mineração de Dados – CAP359

Nome do Aluno

6 de outubro de 2016

Resumo

Neste experimento vamos explicar o que fizemos como projeto na disciplina CAP-359 (Princípios e Aplicações de Mineração de Dados) no Programa de Pós-Graduação em Computação Aplicada no Instituto Nacional de Pesquisas Espaciais. O resto do resumo deve descrever os experimentos que foram feitos, os dados usados, a importância do experimento em si e algumas conclusões, de forma resumida, se possível sem passar de 15 linhas. Ao elaborar o relatório do projeto considere que o mesmo pode vir a ser submetido para uma conferência ou revista: inclua todas as informações que achar que seriam necessárias para os revisores apreciarem seu relatório!

1 Introdução

Na introdução devemos descrever o problema real que estamos tentando resolver com mineração de dados. É importante contextualizar e explicar que importância o problema tem e que benefícios esperamos obter com a mineração destes dados.

Para relatórios de projetos que não sejam estritamente sobre mineração de dados (ex. implementação de algoritmos) podemos descrever o problema que deve ser resolvido com o algoritmo e a importância desta solução.

As seções neste documento são uma **sugestão** de estruturação do mesmo, não precisam ser seguidas à risca. Uma das métricas usadas para avaliar os experimentos através deste relatório é a de *reprodutibilidade*: com os mesmos dados (ou dados da mesma natureza) e as instruções neste relatório conseguimos reproduzir os experimentos? Idealmente sim!

2 Dados

Pode ser interessante explicar como os dados foram coletados e como/onde são armazenados. Isto dará ao leitor uma visão do que seria necessário fazer para reproduzir os experimentos descritos neste relatório com uma base similar de dados ou até mesmo com a mesma base atualizada.

É interessante também nesta seção apresentar:

- O volume total dos dados (número de registros, espaço em disco, etc.) para dar uma ideia a mais sobre a complexidade/importância de minerar estes dados;
- Tabelas que sumariam os dados (descrição dos campos com exemplos dos dados, se factível); no caso de dados mais estruturados mostrar gráficos (imagens, séries temporais, etc.). Veja, por exemplo, a tabela fictícia 1.

ID	x	y	z
A113	8.2	7.3	9.4
A298	7.2	7.2	9.1
A337	8.2	7.3	9.3
B048	9.2	7.5	8.4
C107	9.8	7.5	8.8
...

Tabela 1: Descrição resumida dos dados.

3 Pré-processamento

Explique aqui as fases de pré-processamento que foram aplicada a estes dados para permitir a execução dos algoritmos. Comente sobre cada uma delas em subseções, mantendo somente as que forem pertinentes e adicionando outras se for o caso.

Pense em como deixar claro os passos tomados para garantir a reprodutibilidade do experimento! Use diagramas (Figura 1) e listas para facilitar a compreensão.

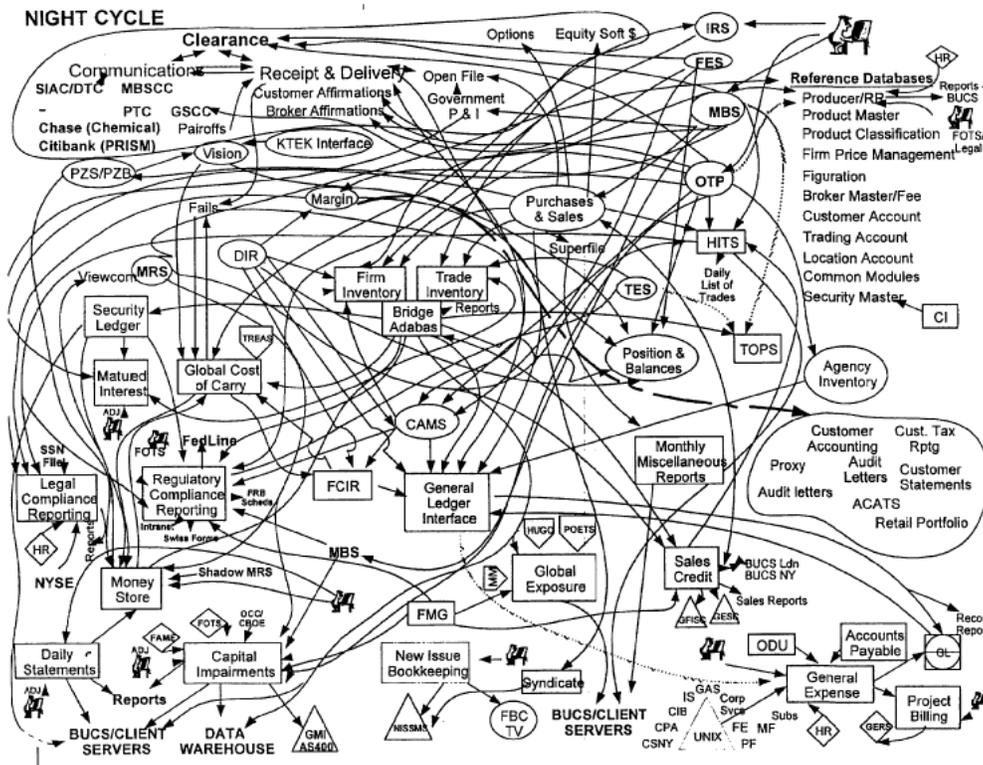


Figura 1: Este diagrama representa de forma simplificada o pré-processamento realizado.

3.1 Seleção

Justifique por que alguns dados foram selecionados para mineração e outros não (se este passo for necessário).

3.2 Transformação

Justifique por que alguns dados foram transformados para mineração e outros não, e que tipos de transformação foram realizados (se este passo for necessário).

3.3 Outros passos

Descreva que outros passos foram tomados para transformar os dados brutos nos que vão ser minerados.

4 Experimentos

Nesta seção você deve, baseado no que foi explicado sobre os dados e a expectativa de resultado da mineração dos mesmos, descrever sumariamente o algoritmo que usará para mineração dos mesmos.

Se o algoritmo for conhecido, ao invés de descrever os passos do mesmo, use referências. Por exemplo, em vez de descrever detalhadamente o algoritmo *Self Organizing Map*, use a referência do Kohonen [1].

Descreva os parâmetros usados, a implementação (R? Python? Weka? RapidMiner? Outro software específico?) e todos os detalhes que permita a reprodutibilidade do experimento.

Se fizer mais de um experimento repita a informação para todos, e se possível justifique por que fez mais de um experimento. Por exemplo, *”fiz vários testes variando o número de camadas da rede neural...”* ou *”Para comparação usamos duas árvores de decisão com critérios de poda diferentes...”*.

Para cada experimento descreva os resultados obtidos, de forma sumariada. Pode ser uma boa ideia organizar cada experimento (ainda mais se houver pré-processamento específico para cada) em subseções.

5 Resultados e Considerações Finais

Descreva os resultados do experimento **como um todo**. Descreva o que poderia ser feito se tivesse mais tempo, poder computacional, dados, etc. – o relatório pode servir de inspiração para que você ou outro leitor prossiga na exploração destes dados!

Use figuras (veja a Figura 2), tabelas e gráficos para deixar os resultados bem claros. Se for o caso referencie figuras que estejam na seção sobre os experimentos (seção 4)

Referências

[1] Teuvo Kohonen. *Self-Organizing Maps*. Springer, 2nd edition, 1997.



Figura 2: As conclusões são óbvias.