

Redução de Cardinalidade de Dimensões em
Bancos de Dados Multidimensionais a partir de
agrupamento de Séries Temporais

Aluno: André Henrique Serafim Casimiro
Orientador: João Eduardo Ferreira

1 de Dezembro de 2011

Contents

1	Resumo	3
2	Introdução	3
2.1	Contextualização	3
2.2	Organização do trabalho	3
3	Fundamentos	3
3.1	Banco de Dados Multidimensional	3
3.2	Séries Temporais	4
3.3	Hematócrito e hemoglobina	4
3.4	Redução dimensional	4
4	Experimentos	4
4.1	Análise global	4
5	O Algoritmo de classificação	5
5.1	Descrição	5
5.2	Implementação	5
6	Parte Subjetiva	6

1 Resumo

2 Introdução

2.1 Contextualização

O Grupo de Banco de Dados do IME-USP (Data-IME) participa desde 2006 do REDS-II, um projeto americano sobre segurança em transfusões de sangue. O Data-IME trabalha com 3 hemocentros brasileiros: o de São Paulo (Fundação Pró-Sangue), o de Minas Gerais (Hemominas) e o de Pernambuco (Hemope). Para trabalhar com todas as informações agregadas em um único lugar foi necessário criar-se um Data Warehouse, trabalho este realizado pelo Data-IME em 2009. Agora é possível fazer análises sobre os dados de forma a tirar conclusões dos mesmos. Este trabalho tem como objetivo encontrar métricas que permitam classificar os doadores com relação ao desenvolvimento de anemia, a saber utilizando níveis de hemoglobina e hematócrito.

2.2 Organização do trabalho

Aqui vai a organização do trabalho...

3 Fundamentos

3.1 Banco de Dados Multidimensional

Bancos de dados estão por toda parte, das pequenas as grandes empresas e instituições. Eles são parte fundamental da maioria dos sistemas de gerenciamento de informação. Este trabalho, no entanto, lida com um tipo específico de banco de dados, os bancos de dados multidimensionais. Também conhecidos como data warehouses (depósito de dados), os bancos de dados multidimensionais são construídos para integrar várias bases de dados e permitem que se faça uma análise global dos dados sob diversas dimensões.

Segundo [1], Warehousing é uma técnica utilizada para recuperação e integração de dados a partir de fontes distribuídas, autônomas e, possivelmente, heterogêneas.

A principal diferença entre um data warehouse e os bancos de dados convencionais presentes na maioria dos sistemas de informação é o propósito de uso. Estes sistemas de informação tem como principal objetivo o armazenamento e recuperação de dados de forma pontual, tais como: inserir uma nova venda, atualizar o cadastro de um determinado cliente, remoção de uma cobrança, etc. Esse tipo de comportamento é conhecido como OLTP (On-Line Transaction Processing) e é responsável pelo processamento de operações críticas para o bom funcionamento de qualquer empresa. Os data warehouses são bases de dados construídas a partir de bases OLTP, e são projetadas para prover aplicações OLAP (On-Line Analytical Processing). Tais aplicações provêm uma interface de consultas analíticas a uma base de dados unificada e tem como função auxiliar a tomada de decisões a nível gerencial.

Mesma tecnologia mas modelagem diferente. Falar sobre os esquemas (snowflake?).

Integrar bases de dados grandes e frequentemente diferentes não é uma tarefa fácil. Os dados provenientes de diversos bancos transacionais podem ser completamente diferentes, da tecnologia à semântica dos campos e tabelas. Assim, é necessária a elaboração do processo de inserção dos dados no data warehouse, a este é o processo de carga dos dados. Trata-se de um conjunto de operações pelas quais os dados devem passar, de forma a obter-se uma base de dados integrada, limpa e confiável. Dessa forma, todas as cargas são feitas de forma incremental ao longo do tempo.

Falar sobre a dimensão tempo, quase obrigatória.

- Bancos que contém grande quantidade de dados
- Foco em consultas analíticas
- Criados a partir de cargas de bancos operacionais
- Mesma tecnologia mas modelagem diferente (fato e dimensões)
- Dimensão obrigatória: tempo

3.2 Séries Temporais

3.3 Hematócrito e hemoglobina

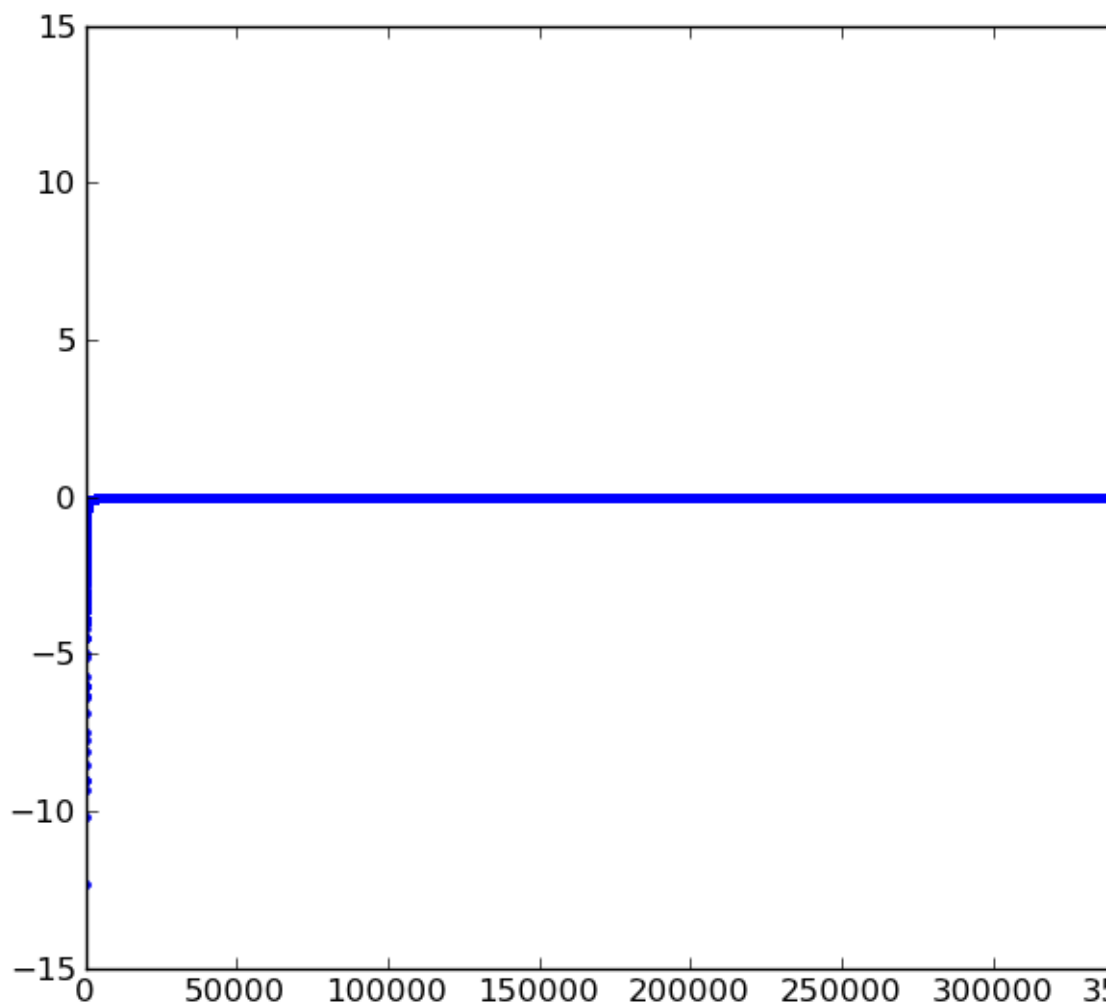
3.4 Redução dimensional

4 Experimentos

4.1 Análise global

Descrição Usar todos os doadores com no mínimo 2 doações e contruir alphas.

Colocar e comentar as fotos dos gráficos



Conclusões Nenhuma classe pode ser encontrada.

5 O Algoritmo de classificação

5.1 Descrição

???

5.2 Implementação

???

6 Parte Subjetiva

References

- [1] João Eduardo Ferreira, Introdução a banco de dados. IME-USP, 2005.