

TRABALHO Nro 2

Disciplina: SCC0532 - Tópicos Avançados em Inteligência Artificial

Professor: Alneu de Andrade Lopes

PAE: Jorge Valverde Rebaza

Instruções gerais

O objetivo deste trabalho é desenvolver a tarefa de mineração de regras de associação e identificar as regras mais significativas. Para isso, você deve:

1. Usar um conjunto de dados do tipo “market basket data”.
2. Preparar e pre-processar os dados.
3. Encontrar regras de associação usando diferentes configurações de parâmetros. Determinar qual dessas configurações é a que considera a mais apropriada e justifique.
4. Determinar quais das regras de associação encontradas são as mais interessantes, justifique e mostre o possível uso de tais regras.

Para este trabalho deverá ser usado o conjunto de dados “dotto”, o qual é disponibilizado no site do Tidia (arquivo dotto.data). No conjunto de dados “dotto”, cada linha representa uma transação, e cada item em uma transação é separado por um espaço em branco.

Para o desenvolvimento deste trabalho você tem a liberdade de usar qualquer software para o processamento das regras de associação. Algumas ferramentas disponíveis são: Weka, programa R, entre outros. O principal algoritmo para descoberta de regras de associação é o algoritmo Apriori, e nessas ferramentas você deverá usar tal algoritmo. Porém, fique a vontade de explorar algum outro algoritmo para realizar uma comparação de resultados. Sendo assim, para o uso dos dados do dotto nessas ferramentas, você deverá preparar o conjunto de dados para o formato aceito pelo software escolhido.

A determinação de regras de associação mais interessantes fica dependendo da sua capacidade de análise. Fique a vontade de se auxiliar, se acredita necessário. De alguma ferramenta software. Caso você use alguma ferramenta, deverá fazer uma descrição da mesma, bem como dos parâmetros e configurações usadas.

Reporte

Para a entrega do reporte você deve seguir o seguinte formato:

1. Capa: título, identificação dos alunos e professor, disciplina e data
2. Introdução: o que está sendo abordado neste trabalho, objetivos, motivação e organização da monografia. Deixar bem claro o domínio e potenciais benefícios do uso das regras de associação.
3. Descrição dos dados. O que se tem no conjunto de dados? É necessário fazer um pre-processamento de dados? Descreva o raciocínio que te levou a escolher certas técnicas de pré-processamento (discretização, ignorar certos atributos, ignorar certas transações, etc) e quais parâmetros usou para isso.
4. Técnicas utilizadas: uma breve descrição do algoritmo Apriori. Se você escolhe usar também outra técnica deverá descrevê-la e mostrar os motivos do uso (é melhor que o

Apriori?, É mais rápido que o Apriori?, etc). Também deverá descrever os parâmetros usados.

5. Resultados: Mostrar um resumo das melhores regras encontradas avaliando-las pelas medidas de desempenho (suporte e confiança) considerando os diferentes parâmetros que considerou no uso do Apriori (e dos outros algoritmos que você considerou). Fique a vontade de usar ferramentas visuais para mostrar a qualidade das regras de associação encontradas (O weka oferece diferentes gráficos no seu painel de visualização, o programa R tem um pacote especial disponível aqui: <https://cran.r-project.org/web/packages/arulesViz/vignettes/arulesViz.pdf>). Mostre também as regras que você considere as mais importantes para um possível cliente seu e justifique o motivo da sua escolha.
6. Referências: referências utilizadas e padronizadas

Se você usou alguma ferramenta de mineração de regras de associação (Weka, R, etc), deverá anexar no seu escaninho o arquivo preprocessado no formato aceitado por tais ferramentas e que você usou para a descoberta das regras de associação, por exemplo, se você usou o Weka deverá anexar o arquivo no formato ARFF que usou.

O trabalho deve ser desenvolvido em duplas.

Data de apresentação: 16 de outubro.