

## 5 Conclusão e Trabalhos Futuros

### 5.1. Conclusões

Esta dissertação abordou conceitos de *data mining* voltado para marketing de relacionamento, propondo uma implementação da técnica de *market basket analysis* baseada na estrutura de grafo.

Para contextualizar o problema, inicialmente foram discutidos tópicos importantes relacionados à mineração de dados, bem como o processo de busca de conhecimento, os tipos de predição de informação e as técnicas mais utilizadas atualmente para descobrimento de padrões em dados armazenados em bancos de dados.

Maior foco foi dado à técnica de *market basket analysis*, pois, ao fazer uso de regras de associação, esta técnica caracteriza-se como uma ótima escolha para solucionar problemas no escopo de marketing de relacionamento, tais como venda casada e *up-selling*. Neste cenário, foi proposto um algoritmo que encontra as relações mais recorrentes entre os produtos (variáveis) nas compras (registros) registradas na base de dados.

A fim de realizar testes que mensurassem o desempenho do algoritmo desenvolvido, um protótipo de aplicação foi criado, permitindo que o usuário escolha a base em que irá realizar os estudos e as variáveis que deseja cruzar a fim de descobrir as relações existentes entre as variáveis. Sendo assim, este trabalho ainda apresenta testes e resultados obtidos através da utilização do algoritmo proposto.

Como principal contribuição do trabalho, podemos apontar a implementação de um algoritmo para *data mining* capaz de descobrir relações entre atributos dos registros de uma base de dados. A principal barreira enfrentada por tais implementações está na dificuldade de encontrar informação relevante, e em volume suficientemente grande, para estudar a eficiência do algoritmo em descobrir regras de associação referentes aos dados. No caso deste trabalho, as bases de dados utilizadas são reais e com qualidade e tamanho satisfatórios para uma análise mais apurada do retorno da solução proposta.

Como segunda contribuição, podemos ressaltar a generalidade do protótipo desenvolvido, que pode ser aplicado a qualquer estrutura e tecnologia de base de dados, através do uso de conector ODBC para realizar as consultas aos dados disponíveis.

## 5.2. Trabalhos Futuros

Como sugestões de trabalhos futuros, podemos enumerar:

- Identificação dos atributos discretos

No processo de descoberta das relações entre os atributos de um registro, só faz sentido trabalhar com aqueles que possuem domínio previamente conhecido, ou seja, atributos que sejam discretos. Sendo assim, é interessante apresentar para o usuário apenas as variáveis que sejam desse tipo, o que não foi realizado neste trabalho devido à uma limitação do conector ODBC utilizado.

- Investir na visualização gráfica das informações descobertas

A visualização do relacionamento entre os atributos em termos de imagens facilita a compreensão por parte do usuário e permite análises mais rápidas e intuitivas que o estudo das regras de associação descobertas. É interessante que um trabalho futuro explore a parte de visualização, buscando outros componentes que permitam a visualização do grafo de maneira mais limpa e que suporte grafos com grandes quantidades de nós e relações.

- Trabalhar com outras estruturas de dados

Este trabalho fez uso apenas de bases de dados de propriedade da Stone Age Tech, que não trabalha com tabelas como os bancos de dados convencionais, mas com uma estrutura colunar. Ao trabalhar com uma base de dados convencional, seria possível identificar se a estrutura da base influencia o desempenho final do algoritmo.

- Estudo comparativo entre algoritmos

A proposta dessa dissertação foi uma implementação da técnica de *market basket analysis*, utilizando-se os conceitos de grafos ponderados a fim de encontrar as relações mais frequentes, ou seja, o clique máximo do grafo. Como trabalho futuro complementar a este, pode-se sugerir a construção de outros algoritmos que tenham o mesmo princípio de regras de associação, como por exemplo, o Apriori, com o objetivo de realizar comparações de tempo de resposta e de acuracidade das informações sobre as relações.

- Incluir medidas de interesse como suporte e confiança

A introdução de medidas, como suporte e confiança, permite que os padrões descobertos sejam ordenados ou classificados de acordo com o grau de interesse associado aos mesmos. Além disso, essas medidas podem ser usadas para guiar ou restringir o espaço de busca, melhorando a eficiência da mesma ao eliminar conjuntos de regras de associação que não satisfaçam as condições predeterminadas.

- Acrescentar outras abordagens para descoberta do clique máximo

Na literatura existem algumas abordagens diferentes para a solução problema do clique máximo. A título experimental, o estudo e a implementação de novas técnicas é sugerido como trabalho futuro complementar ao proposto por esta dissertação.