

6 Conclusões e Trabalhos Futuros

6.1. Conclusões

Esta dissertação apresentou uma revisão detalhada de todo o processo de análise de agrupamentos, bem como o desenvolvimento de um aplicativo capaz de auxiliar em todo esse processo.

Como foi visto anteriormente, sabe-se que qualquer que seja o método de agrupamento de dados, eles sempre vão gerar agrupamentos, o que pode direcionar a análise para outra solução caso o resultado não tenha sido satisfatório. Por isso é importante conhecer as diferenças e semelhanças entre os principais métodos de agrupamento e saber quando aplicá-los.

Analisar os resultados obtidos é de fundamental importância para um bom conhecimento do próprio método utilizado, bem como conhecer a estrutura dos dados e, dessa forma, extrair informação útil sobre o conjunto analisado.

Nesse estudo foi apresentada uma revisão da teoria de agrupamento de dados e discutidos seus principais aspectos. Estudou-se primeiramente a seleção e tratamento dos dados, onde se puderam observar os diferentes tipos de atributos presentes em uma base de dados e as transformações necessárias sobre mesmos. Essas transformações têm bastante relevância, uma vez que elas são responsáveis pela eficiência do algoritmo, influenciando diretamente nos resultados gerados.

Estudou-se como ocorre o processo de classificação de um determinado dado a um determinado agrupamento e as medidas de proximidade envolvidas, definindo conceitos comumente usados como medidas de similaridade e dissimilaridade.

O estudo abrangeu também as características e funcionamento dos principais métodos de agrupamento de dados, mostrando, sobretudo, como entender e analisar a eficiência do resultado de um processo de agrupamento de dados como, por exemplo, o gráfico da silhueta e o dendograma para métodos hierárquicos.

É comum encontrar informações espalhadas dos diversos métodos de agrupamento de dados, porém se torna difícil encontrar toda essa informação estruturada de maneira adequada ao uso, por isso o estudo focou também no desenvolvimento de uma ferramenta capaz de auxiliar de maneira fácil e completa a análise de agrupamentos, apresentando uma interface de integração com duas ferramentas muito usadas hoje em dia, bem como incentivar o estudo desse assunto.

No estudo de caso realizado, a ferramenta se mostrou bastante prática, principalmente no apoio dado na análise dos resultados, influenciando de forma segura na escolha do método apropriado, bem como na determinação sobre o número de agrupamentos para uma determinada base de dados e na extração de informação sobre a estrutura formada.

Os métodos de agrupamento de dados de natureza fuzzy mostraram ter um grande diferencial em relação aos demais métodos, mostrando uma visão diferenciada sobre a estrutura dos dados e principalmente sobre os problemas abordados. Isso pode ser observado na análise feita no estudo de caso sobre a base de dados da Ilha do Governador.

6.2. Trabalhos Futuros

Como complemento do trabalho realizado e em continuidade ao estudo de análise de agrupamentos, seria relevante abordar:

- Técnicas automáticas de tratamento de dados, procurando demonstrar as técnicas mais comuns que sejam capazes de, por exemplo, eliminar dados duplicados ou corrompidos e tratar *outliers*;
- Técnicas automáticas de transformação de atributos categóricos em atributos quantitativos;
- Outras medidas de similaridade e dissimilaridade que possam ser úteis no processo de agrupamento de dados;
- Outros gráficos e medidas capazes de auxiliar na análise de agrupamentos;
- Outros métodos de agrupamento de dados, em especial métodos que utilizem técnicas fuzzy, redes neurais ou algoritmos genéticos;

- Criação de módulo que facilite o desenvolvimento de interfaces entre o aplicativo desenvolvido com outros aplicativos;
- Criação de módulo para parametrização do método de agrupamento de dados selecionado;
- Implementação de outras técnicas de tratamento de dados;
- Incorporação de novos gráficos e medidas capazes de auxiliar na análise de agrupamentos.