

5 Conclusões e trabalhos futuros

Em resposta a necessidade de classificação de grandes quantidades de informação, introduzimos neste trabalho uma abordagem para organização de dados semi-estruturados, representados por frames, sem uma estrutura de classes prévia.

O agrupamento de dados é uma técnica muito aplicada em diversas áreas como na biologia, estatística, psicologia, engenharia, medicina, marketing (análise de mercado), arquivologia, informática, negócios, entre muitas outras. Contudo, achar a melhor solução para este problema não é fácil.

A utilização de algoritmos supervisionados tem um ponto diferencial conveniente que é possuir uma massa de dados para treinamento já classificados. Porém, em determinados casos, necessitamos uma forma de classificação automática, já que o ato de classificar dados para treinamento pode ser uma tarefa árdua ou, até mesmo, impossível. Deste modo, algoritmos de agrupamento não supervisionados fazem um papel fundamental na classificação de dados similares.

Discutimos inicialmente os métodos mais utilizados para classificação. Em seguida, resumimos os principais algoritmos e técnicas utilizadas no processo de agrupamento. Por fim, apresentamos a implementação de um sistema de classificação automática de dados baseada numa estrutura de dados chamada *frame* e nas teorias de classificação, nos algoritmos e nas técnicas de agrupamento apresentadas ao longo deste trabalho.

O processo de classificação proposto neste trabalho obteve sucesso nas massas de dados sintéticas e reais. Uma contribuição importante, além da classificação automática de dados e da geração de classes, advém da análise realizado sobre a massa de dados do SIAE, onde foi possível analisar a classificação original e recomendar alterações de acordo com a classificação gerada pelo processo.

Os conhecimentos produzidos a partir deste trabalho representam um ponto de partida para contribuições através da implementação de outros algoritmos de agrupamento supervisionados e não-supervisionados, validação de classificações realizadas manualmente e a comparação deste método de classificação com métodos já existentes.