

A importância do data mining na descoberta de padrões e tendências nos dados

Pedro Bastos

ESA-Instituto Politecnico de Braganca, CIMO - Centro de Investigacao de Montanha, Campus de Sta Apolonia, Braganca, Portugal

Hoje em dia, a quantidade de dados gerados e armazenados no decurso de uma qualquer actividade excede a capacidade de análise desses mesmos dados sem o uso de técnicas de análise automatizadas. Como consequência deste aumento efectivo de informação, o processamento através de métodos tradicionais tornou-se mais difícil e complexo [1]. As ferramentas convencionais de análises de dados têm capacidades limitadas em detectar padrões e descobrir o conhecimento inserido nos dados, pois só utilizam métodos estatísticos [2]. Surgiu então a urgente necessidade de uma nova geração de técnicas e ferramentas computacionais de forma a assistir o ser humano na extracção de informação útil, isto é, conhecimento. Assim, no final dos anos 80 emergiu a área da descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases* – KDD), recorrendo a modelos e técnicas de *Data Mining* para extrair conhecimento útil, padrões e tendências, previamente desconhecidas, de forma autónoma e semi-automática [3]. O *Data Mining* e a descoberta de conhecimento em bases de dados tornaram-se proeminentes campos de pesquisa em constante e rápido crescimento. Actualmente muitas técnicas e algoritmos têm sido desenvolvidos e utilizados em *Data Mining*. O *Data Mining*, por outro lado, através do uso de algoritmos específicos ou de mecanismos de pesquisa, tenta descobrir padrões discerníveis e tendências nos dados, inferindo regras para os mesmos [4]. O processo de *Data Mining* torna-se mais fácil de compreender, implementar e desenvolver se for enquadrado no contexto de uma metodologia [5]. A ferramenta de *Data Mining* IBM SPSS Modeler consiste numa ferramenta de Data Mining capaz de desenvolver rapidamente modelos preditivos especializados em vários ramos de negócio e implementá-los em operações de forma a melhorar o processo de toma de decisão. Concebido em torno da metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*), o Modeler suporta todo o processo de *Data Mining*, partindo dos dados e atingindo os melhores resultados para o negócio [6]. A ferramenta integra um conjunto de algoritmos de modelação e aprendizagem, com uma interface de programação visual que torna compreensível todo o processo de extracção de conhecimento a partir de diversas fontes de dados. O utilizador elabora projectos ou *streams*, onde define as sequências de operações que deseja realizar sobre os dados (processo de extracção de conhecimento). Estas operações são representadas por ícones que estão agrupados de acordo com o seu tipo de funcionalidade.

1. Goebel, M. and L. Gruenwald, A survey of Data Mining and Knowledge Discovery Software Tools, in *SIGKDD*. 1999. p. 20-33.
2. Santos, M.F. and C. Azevedo, *Data Mining - Descoberta de conhecimento de bases de dados*. 2005: FCA - Editora de Informática. 194.
3. Apte, C., et al., *Business Applications of Data Mining*. ACM, 2002. **45**(8): p. 49-53.
4. Han, J. and M. Kamber, *Data Mining: Concepts and Techniques*. 2001, USA: Morgan Kaufmann Publishers.
5. Groth, R., *Data Mining: Building Competitive Advantage*. 2000, USA: Prentice Hall PTR.
6. Clementine, S., *SPSS Clementine Applications Guide*. 2007, SPSS Inc.: Chicago.