

Uma arquitetura estendida para *data mining* integrado ao processo decisório

Miguel A. Feldens*

Altino Pavan**

Resumo

O termo *data mining* é empregado no meio acadêmico há quase uma década como “aprendizado a partir de bases de dados”. A área surgiu com *background* de anos de pesquisas em aprendizado de máquina, redes neurais e estatística. Dentro de um processo mais amplo, chamado descoberta de conhecimento em bases de dados [FAY 96], seu objetivo é extrair conhecimento novo, útil e interessante a partir de bases de dados. Em termos das organizações onde o *data mining* é aplicado [BER 97], tais ferramentas apóiam a extração de valor a partir das bases de dados, ativo comumente sub-utilizado. Apenas recentemente a mineração de dados tem conquistado espaço nas organizações onde, ao aplicar *data mining* a grandes volumes de dados, sujeitos a redundâncias, inconsistências, distribuídos e armazenados em formatos distintos, foi encontrado o primeiro “gargalo” para se obter sucesso na aplicação da tecnologia: a preparação e qualificação de dados. Integrar esquemas, deduplicar entidades [REG 00], padronizar a representação de diversos atributos e disponibilizar repositórios de dados para aplicações analíticas foram alguns dos desafios onde *data warehousing* [KIM 96] e *data mining* se encontraram. Atualmente, novos desafios são encontrados em outras etapas deste processo, tal como a aplicação de resultados, que é o tema deste trabalho.

A descoberta de conhecimento em bases de dados é um processo cíclico e interativo, formado pela preparação de dados, *data mining*, aplicação dos modelos e monitoração de resultados. Considerando que a preparação de dados é ponto de convergência com ferramentas de higienização de dados que formam a infra-estrutura para *data warehousing* e *database marketing* [REG 00], e que as aplicações reais do *data mining* motivaram o aprimoramento dos algoritmos de aprendizado, o principal gargalo do processo encontra-se atualmente na aplicação dos resultados. Em aplicações típicas de *data mining*, mais que algumas regras interessantes – os chamados *nuggets* -, geralmente são encontradas centenas ou milhares de regras, entre as quais há o óbvio, o que não interessa ao usuário e algum conhecimento interessante. Preocupados com esta limitação, diversos autores trabalharam sobre a questão do *interestingness*, definindo métricas dependentes ou não de cada domínio, criando técnicas e *templates* para filtragem de conhecimento [KLE 97].

* miguel@godigital.com.br

** altino@godigital.com.br

A filtragem do conhecimento é orientada ao usuário que elabora estratégias, buscando informação que suporte ações e decisões bem informadas. Independentemente disso, o que se propõe aqui é que o conhecimento descoberto pode ter aplicação direta no nível operacional, sugerindo ações, disparando alertas, personalizando a interação com o cliente dinamicamente, etc. Em database marketing, por exemplo, todas as ações e tomadas de decisão sobre o relacionamento com o cliente passam pelos chamados pontos de contato (*touch points*), tais como *call center*, *website* e frente de loja, etc. É nestes pontos de contato que devem ser exploradas as oportunidades de venda de produtos associados (*cross-sales*), oferta de produtos mais rentáveis, adequados ao perfil de cada cliente (*up-sales*), e deveriam ser percebidos comportamentos que demandam ações para não perder o relacionamento com cada cliente (*client retention*).

Para levar o valor extraído das bases de dados até estes pontos de contato, foi criada a arquitetura HyconesXDA (*eXtended Data mining Architecture*), baseada em componentes distribuídos. Tais componentes implementam tanto as partes do processo de *data mining*, quanto das atividades pós-*data mining*, tais como: configuração de *data mining queries*, aprendizado, filtragem de conhecimento, inferência baseada nas regras descobertas através de *data mining* (dedução e abdução com tratamento de incertezas), comparação de bases de conhecimento, etc. A arquitetura HyconesXDA pode ser integrada a aplicativos existentes, seja para torná-los capazes de aprender a partir de dados, seja para utilizar uma base de conhecimento anterior, inferir a partir da interação e sugerir alguma ação ou conteúdo. Exemplificando, alguns dos produtos desenvolvidos, ou em desenvolvimento, baseados nesta arquitetura:

- **InfoZoom:** Catálogo eletrônico para navegação em bases de dados, para usuários não especializados (*everyday business users*). Ferramenta produzida pela empresa alemã humanIT (<http://www.humanit.de>), utiliza componentes de *data mining* HyconesXDA para descobrir regras em *background*, enquanto o usuário explora suas bases de dados.
- **Empathy Call Center:** Baseado em regras previamente descobertas, e utilizando um componente capaz de realizar inferência, os *scripts* apresentados à operadora de um sistema de atendimento são escolhidos dinamicamente, conforme a situação e o perfil do cliente. Conforme o perfil do cliente sendo atendido, utilizando as regras conhecidas, o sistema otimiza o aproveitamento de oportunidades de *cross-sales*, oferece os produtos mais adequados, alerta possibilidade de atrito.
- **Empathy Web Personalization:** Integração de bases de regras incorporadas a agentes para personalização dinâmica de conteúdos na *web*. Um *website* pode oferecer uma experiência personalizada para cada internauta, utilizando regras baseadas no seu perfil declarado (dados demográficos, etc.) e observado (páginas que visitou, produtos que comprou, etc.).
- **Watchdogs:** Agentes para monitoração e alerta. Cada “cão de guarda” (*watchdog*) monitora uma regra do negócio, que pode estar relacionada à cadeia produtiva,

logística, etc. Ao verificar tal regra, o sistema pode disparar um alerta na mídia preferida de seu usuário. O responsável pela análise de custos de uma indústria petroquímica, por exemplo, poderia receber um alerta através de uma mensagem via celular, quando houvesse uma alta de determinados insumos que deverão repercutir nos custos de produção. O componente de *data mining* apóia a construção do sistema, realizando aquisição automática do conhecimento nos *watchdogs* a partir de bases de dados.

Na medida em que os mercados mais competitivos criam demanda por mecanismos que apóiem a tomada de decisão, e tecnologias como o *data mining* atingem a “maturidade”, estas podem estar disponíveis direta ou indiretamente a usuários em diferentes níveis da organização. O resultado do *data mining* pode ser utilizado não apenas por quem elabora estratégias, mas estar diretamente integrado no momento em que estas estratégias são operacionalizadas. Esta é a proposta implementada na arquitetura HyconesXDA.

Referências

- [BER 97] BERRY, M. J. A. & LINOFF, G. Data Mining Techniques: For Marketing, Sales, and Customer Support. John Wiley & Sons, 1997.
- [FAY 96] FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P. & UTHURASAMY, R. (eds.), Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996.
- [KIM 96] KIMBALL, R. The data warehouse toolkit. John Wiley & Sons, 1996.
- [KLE 97] KLEMENTTINEN, M.; MANNILA, H. & TOIVONEN, H. A data mining methodology and its application to semi-automatic knowledge acquisition. In: Proc. of the 8th Intl. Conf. and Workshop on Database and Expert Systems Applications (DEXA'97), 670 - 677, Toulouse, France, September 1997.
- [REG 00] REGO, R. Data Cleansing Technology. Conferência Internacional de Database Marketing. São Paulo: IDBM, Maio, 2000.

