

Coluna do Taurion

Saiba como usar algoritmos preditivos

Vivemos em um oceano de dados que só tende a crescer e já está claro que o valor das iniciativas em big data aparecem quando aplicamos analítica ou matemática aos dados. Ou seja, valor = analítica + dados.

por Cezar Taurion

A cada dia vejo que o conceito de big data começa a se firmar. Embora ainda vejamos mais hype do que ações concretas, algumas primeiras experiências já mostram que o processo é irreversível. Vivemos em um oceano de dados que só tende a crescer e já está claro que o valor das iniciativas em big data aparecem quando aplicamos analítica ou matemática aos dados. Ou seja, valor = analítica + dados.

Mas é curioso que na maioria das conversas sobre o assunto o foco é o volume e a variedade de dados. Mas, sem analítica ou algoritmos iríamos nos afogar neste oceano de dados. Vamos então falar um pouco de algoritmos, principalmente os preditivos. Existem diversos outros algoritmos, inclusive o descritivo, muito usado nos BI tradicionais, que descrevem o passado e eventualmente nos ajudam a tomar decisões sobre o futuro. Um exemplo típico são perguntas como "quantos clientes perdemos nos últimos 3 meses?" ou "as fraudes aumentaram ou diminuíram no último ano?". E com base neste histórico, nas nossas percepções, intuições e experiências, tomamos decisões futuras. Mas claramente vemos que isso não é suficiente para atender à complexidade crescente do cenário de negócio atual. Observamos que a intuição e a experiência têm peso excessivo nestas decisões.

Em um cenário de negócios relativamente estável, é perfeitamente adequado; mas quando o cenário é dinâmico e mutante, provavelmente nossas experiências e intuições podem não ser suficientes.

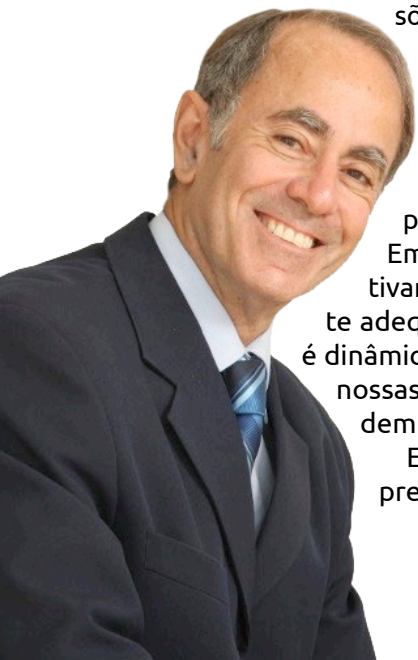
Entram em campo as análises preditivas, que não são novidade

e sempre estiveram presentes na academia e em setores específicos como meteorologia. Mas a crescente digitalização da sociedade e a disponibilidade facilitada de armazenamento e processamento de imensos volumes de dados tornou esta disciplina acessível a qualquer empresa.

Mas o que é um modelo preditivo? A resposta é simples: uma função matemática que aplicada à uma massa de dados, consegue identificar padrões ocultos e prever o que poderá ocorrer. Prever o futuro sempre foi um desafio e uma busca incessante – daí a leitura da palma das mãos, astrologia etc. Agora podemos pensar que sim: é possível fazer previsões bastante razoáveis.

Existem dois tipos de modelos preditivos, os supervisionados e os não supervisionados. No primeiro, em uma fase que chamamos de treinamento do modelo, os dados de entrada e a saída desejada são apresentados juntos. O treinamento dura até que o modelo aprenda a mapear os dados e identificar padrões entre as entradas e as saídas. Como exemplos destes modelos, temos as redes neurais e árvores de decisão. Os modelos não supervisionados só recebem os dados de entrada e sua função é descobrir os relacionamentos entre os dados apresentados. A técnica de clusterização é um bom exemplo deste modelo.

Vamos ver na prática como funcionam estes algoritmos preditivos. Suponhamos uma operadora de telefonia móvel. Um dos seus principais problemas é a taxa de desconexão ou churn rate. Diminuir esta perda de clientes é extremamente vantajoso para o negócio. Usando um modelo supervisionado entramos, entre outros, com dados como perfil de cada cliente, seu uso ao longo do tempo, número de reclamações efetuadas, e se ele se desconectou ou não. Agregando um grande volume de dados, o modelo consegue apren-



der e criar um padrão que reconhece o perfil de cliente que tende a permanecer ou se desconectar da operadora. Uma vez aprendido, validamos o modelo. Para isso, apresentamos um outro conjunto de dados e verificamos qual o nível de acerto em relação à taxa de desconexão. Se for um nível de acerto baixo, voltamos a treinar o modelo até que ele realmente seja considerado apto a fazer previsões.

Quando o modelo está validado, ele entra em operação, ou seja, começa a ser aplicado para identificar antes que aconteça a possibilidade de determinado cliente se desconectar. Agregando ao modelo regras de negócio, como agrupar clientes por rentabilidade, a operadora pode fazer ofertas diferenciadas para evitar a desconexão.

Entre as lições aprendidas nas experiências com modelos preditivos, dois fatores chamam atenção: primeiro é necessário dados em volume adequado para o modelo aprender e identificar um padrão, e em segundo, os dados têm que ser válidos. Se a base de dados contiver erros, não existe modelo que gere resultados adequados. A velha máxima de TI que diz que "se entra lixo, sai lixo" continua válida. Por outro lado, uma escolha errada de modelo pode jogar pela janela uma base de dados válida.

A escolha do modelo é importante. Diversas variáveis devem ser consideradas. Se, por exemplo, não houver condições de associar dado de entrada ao resultado, uma alternativa é usar um modelo não supervisionado. Quando é possível associar entrada com o resultado, escolhe-se modelos supervisionados. Mas qual? Se for necessário que o modelo consiga explicitar para o usuário o porque da decisão de indicar que determinado cliente se desconectará, devemos usar um modelo que consiga fazer o back tracking. Uma rede neural não permite isso, mas um modelo árvore de decisão sim.

Criar iniciativas de big data, como vemos, não é simplesmente adquirir tecnologias, mas demanda expertise para selecionar e validar dados, identificar os melhores modelos preditivos para cada caso e principalmente identificar com a maior exatidão possível o problema de negócio que se quer resolver. Caso a empresa queira começar com algumas experimentações, antes de mergulhar fundo, o uso de modelos open source baseados na linguagem estatística R pode ser uma alternativa. Quando escrevendo este texto, havia mais de 6100 pacotes de aplicações as mais diferente

Entre as lições aprendidas nas experiências com modelos preditivos, dois fatores chamam atenção: primeiro é necessário dados em volume adequado para o modelo aprender e identificar um padrão, e em segundo, os dados têm que ser válidos.

possíveis, em setores como biologia, finanças, genética, etc. Vejam [o link].

Vamos sumarizar algumas recomendações:

- ♦ Identifique com a maior precisão possível o problema de negócio. Quanto mais precisa a pergunta, mais precisa será a resposta e portanto maior o valor da resposta.
- ♦ Tenha dados em volume e qualidade adequados. Sem qualidade, volume não tem valor.
- ♦ Não superestime o valor da previsão. Mesmo em uma sociedade cada vez mais data-driven, a intuição muitas vezes é necessária. Uma velha piada entre os meteorologistas (um dos primeiros usuários de modelos preditivos) diz que existem modelos tão precisos que previram com exatidão sete dos últimos três furacões.;
- ♦ Não subestime o desafio da implementação. Não é só entupir a empresa de tecnologias, mas é necessário expertise (conhecimento do negócio, tecnologia, modelagem) para fazer a coisa acontecer.

No mais, big data e modelos preditivos estão deixando de ser novidade e vão se tornar, em um futuro não tão longínquo assim, algo tão comum quanto um ERP nas empresas. Na próxima década os automóveis sem motorista serão lugar comum nas ruas. E o que os dirige? Modelos preditivos! Um carro do Google por exemplo é um carro recheado de modelos preditivos, que tomam decisões a cada instante. ■

Cezar Taurion é profissional e estudioso de Tecnologia da Informação desde fins da década de 70, com educação formal diversificada, em Economia, mestrado em Ciência da Computação e MBA em Marketing de Serviços, além de possuir experiência profissional moldada pela passagem em empresas de porte mundial. É autor de seis livros que abordam assuntos como Software Livre, Grid Computing, Software Embarcado, Cloud Computing e Big data.
