

## Inferência de impacto causal de um sistema de recomendação na taxa de rejeição de páginas da Agência Embrapa

Flávio Margarito Martins de Barros<sup>1</sup>  
Stanley Robson de Medeiros Oliveira<sup>2</sup>

A Agência de Informação Embrapa é um sistema web com o objetivo de organizar, tratar, armazenar e divulgar informações técnicas e conhecimentos gerados pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Diariamente o site recebe milhares de acessos que são registrados em uma base de dados. A partir destes dados, por meio de regras de associação (HAN et al., 2011), foi criado e implantado um sistema de recomendação para páginas relativas à cana-de-açúcar (BARROS et al., 2013). O sistema foi avaliado por meio da métrica *bounce rate* (SCULLEY et al., 2009) ou taxa de rejeição, e obteve resultados parciais com relação ao impacto da rejeição de suas páginas para a cultura da cana-de-açúcar. O objetivo desse trabalho foi reavaliar o desempenho do sistema de recomendação por meio de modelos de séries temporais bayesianos, os quais podem revelar impacto causal do efeito da recomendação (BRODERSEN et al., 2014).

A metodologia utilizada na pesquisa incluiu a preparação de dados de 1.450.484 sessões de usuários, relativas ao período de novembro de 2010 a janeiro de 2013, onde a intervenção do sistema de recomendação ocorreu a partir do dia 01 de novembro de 2012. Foram calculadas séries temporais de taxas de rejeição para 20 páginas da Agência Embrapa cana-de-açúcar. Os modelos bayesianos foram ajustados para as séries entre o período de novembro de 2010 a outubro de 2012, tal que a projeção deste modelo foi comparada aos dados reais de taxa de rejeição para as mesmas páginas durante 60 dias após a intervenção. Pela comparação da diferença entre a

projeção e a série real, foi avaliada a probabilidade de obtenção dos valores efetivamente observados sob a hipótese do efeito nulo de intervenção.

Na Figura 1, observa-se o modelo ajustado, as previsões após a intervenção e a série observada após a intervenção, para a série de taxas de rejeição diárias da página de Cachaça. Nos 60 dias após intervenção do sistema de recomendação, observa-se uma ligeira queda na taxa de rejeição, mas somente observando a série não é possível verificar se a diferença é estatisticamente significativa. Assim, na Tabela 1, são apresentadas as probabilidades de ocorrência das séries efetivamente observadas, caso estas séries evoluíssem de acordo com o modelo ajustado antes da intervenção.

Da Tabela 1 observa-se que os resultados obtidos indicam que o sistema de recomendação teve impacto na diminuição da taxa de rejeição. Em 11

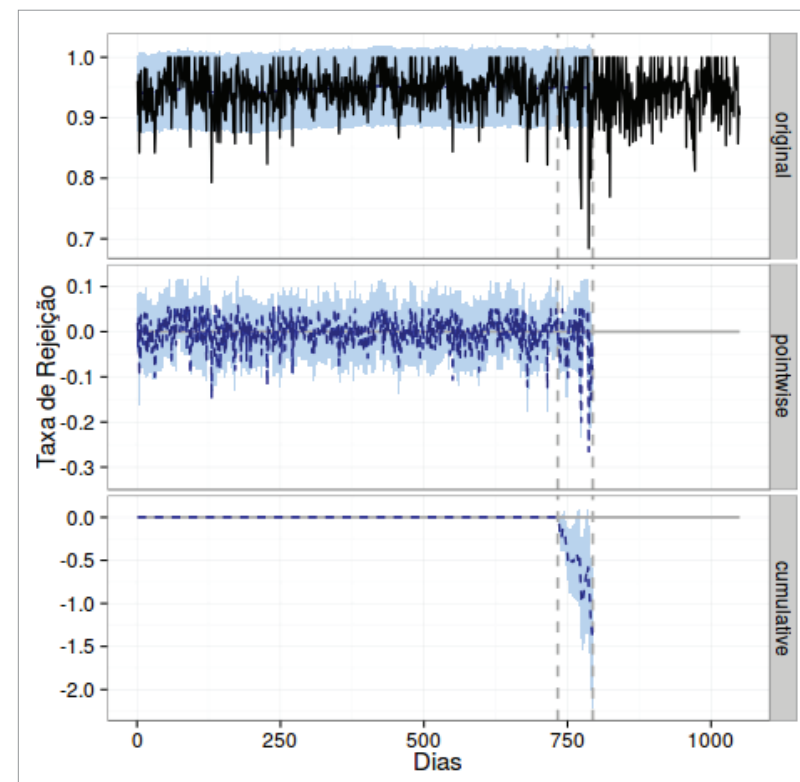


Figura 1. Modelo para a série da página de Cachaça.

<sup>1</sup> Universidade Estadual de Campinas (Unicamp)

<sup>2</sup> Embrapa Informática Agropecuária

**Tabela 1.** Páginas e as probabilidades de ocorrência da série observada sob hipótese de que o sistema de recomendação não teve efeito.

Página	Probabilidade
Cachaça	0,002
Fabricação do Açúcar	0,036
Qualidade da matéria-prima	0,006
Custos e rentabilidade	0,001
Variedades	0,059
Agricultura de precisão	0,030
Monitoramento ambiental	0,055
Análise de solo	0,071
Queima	0,082
Corte	0,052
Carregamento	0,022

das 20 páginas analisadas observou-se queda significativa nos valores da métrica, isto é, quando a probabilidade de observação da série, supondo um regime inalterado, foi inferior a 0,1, mostrando um impacto significativo. Salienta-se que a taxa de rejeição de uma página pode estar associada a outros fatores, além da falta de interesse do usuário, como no caso de conteúdo específico. Nesses casos o usuário pode se sentir satisfeito com a informação obtida e abandonar o website com uma única visualização. Algumas páginas apresentam conteúdos que podem demandar informações adicionais, tal que nesses casos a presença do sistema de recomendação pode ter um impacto positivo. De acordo com os resultados obtidos nesse trabalho, as conclusões de Barros et al. (2013), utilizando taxas de rejeição gerais, antes e após a implantação do sistema de recomendação, foram confirmadas. O sistema é eficaz na diminuição das taxas de rejeição, mas nem todas as páginas registram diminuição pelos motivos explicitados.

Palavras-chave: Taxa de rejeição, mineração de dados, regras de associação, informações tecnológicas agrícolas.

## Referências

- BARROS, F. M. M.; OLIVEIRA, S. R. M.; OLIVEIRA, L. H. M. Desenvolvimento e validação de um sistema de recomendação de informações tecnológicas agrícolas sobre cana-de-açúcar. *Bragantia*, Campinas, v. 72, n. 4, p. 387-395, dez. 2013.
- BRODERSEN, K. H.; GALLUSSER, F.; KOEHLER, N. R.; SCOTT, S. L. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, p. 1-33, 2014.
- HAN, J.; KAMBER, M.; PEI, J. **Data mining**: concepts and techniques, 3rd. San Francisco: Morgan Kaufmann Publishers, 2012. 740 p.
- SCULLEY, D.; MALKIN, R.; BASU, S.; BAYARDO, R. Predicting bounce rates in sponsored search advertisements. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 15., 2009. **Proceedings...** New York: ACM, 2009. p. 1325-1334.