

Técnicas de aprendizado de máquina aplicadas ao projeto de formulações em um caso industrial

Autores: Elcio Fernando dos Santos Pereira^{1*}; Moisés Teles Dos Santos^{2*}

¹ INDOVINYA, P&D tecnologias de processo

² Departamento de Engenharia Química, Escola Politécnica, Universidade de São Paulo

Autores Correspondentes: *elcio.pereira@oxiteno.com, *moises.teles@usp.br

Resumo

A produção de especialidades químicas garante melhores margens de lucro em um mercado competitivo que pressiona o custo dos produtos *commodities*. Sua produção consiste no desenvolvimento de formulações, que são misturas que aliam as propriedades de seus constituintes para o atendimento de uma aplicação. A busca pela redução do custo e o tempo de desenvolvimento desses produtos é o objetivo deste trabalho, que apresenta a aplicação de aprendizado de máquina à uma base de dados de análises de propriedades químicas, físico-químicas e testes de desempenho de formulações uma especialidade química. O primeiro passo foi o gerenciamento dos dados organizados em um arquivo com 84 variáveis e mais de 35.500 linhas. Cerca de 32% das células estavam com dados faltantes o que inviabilizou seu uso imediato. A aplicação do método de máxima verossimilhança, pacote missMDA da linguagem R, completou os dados faltantes (HUSSON et al., 2019). A etapa seguinte foi o treinamento de redes neurais artificiais e máquinas de vetores suporte, uma para cada ensaio, totalizando 46 de cada tipo. Vale ressaltar que outras 38 variáveis resposta eram categóricas desbalanceadas e não foram modeladas. A topologia das redes neurais e das máquinas de vetores suporte usou uma estratégia chamada “greedy search” que implica em variar simultaneamente a taxa de aprendizado, o erro e a arquitetura da rede (número de neurônios e de camadas), além da quantidade de variáveis de entrada (% dos ingredientes nas fórmulas, propriedades físico químicas). Para as máquinas de vetor suporte, foi variado um parâmetro de suavização (“C”) das margens entre o vetor de dados inseparáveis e as observações do banco de dados. Para ambas as metodologias foram executados testes de validação cruzada ao se repetir por dez vezes a amostragem aleatória de dados para treinamento e teste na razão de 70:30. O desempenho foi avaliado pelo erro quadrático médio (MSE), sendo inferior a 10% da faixa de aceitação de qualquer uma das 46 variáveis resposta. A máquina de vetor suporte apresentou desempenho superior em 67% das variáveis resposta estimadas, com menores valores de erros quadráticos médios.

Palavras-chave

Redes neurais, projeto de produto, aprendizado de máquina, formulação, máquinas de vetores suporte.

Agradecimentos

Indovinya e ao Programa de Pós-Graduação em Engenharia Química da USP.

Referências

- CHANG, C.-C., & LIN, C.-J. (2001). LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- FRITSCH, S. et al. Package “neuralnet” Title Training of Neural Networks. 2019.