Jordan Drexler DSC 540: Machine Learning for Data Science

November 3, 2021

Machine Learning Readiness

Part 3- Technical Report

In an effort to use “machine learning to translate applicant work history into predictors of performance and turnover” (Sajjadiani), the Journal of Applied Psychology article details a few key forms of learning/ tasks, but it seems they commonly stayed in the supervised learning realm of machine learning. Within the supervised learning category, one task they performed in multiple categories is text processing. “Text Processing is one of the most common task in many ML applications” (Nabi) One examples of text processing from the Nabi reference, that seems the most similar to the one used in this article and in the example I will describe below is “Sentiment Analysis: To determine, from a text corpus, whether the sentiment towards any topic or product etc. is positive, negative, or neutral” (Nabi).

One specific example where they use supervised learning text processing machine learning is in order to “identify words or phrases that signal applicants’ relatively stable psychological characteristics based on a priori categories” (Sajjadiani). As described on page 1209, they are looking at the open-ended responses from applicants about why they left a previous role. And then within these responses, the article describes using the text processing to pick up on key wording then categorizing the wording into pre-defined groupings.

One of the key pieces that makes this example supervised learning is because of the work done prior to run this learning “We trained a small sample of data (3% of the data) and manually categorized applicants’ reported reasons for leaving a past position into four categories: (a) involuntary, (b) avoiding bad jobs, (c) approaching better jobs, and (d) other reasons” (Sajjadiani). So in summary, because they pre-trained the model, and because of the manual categorization setup initially, this would be considered supervised learning.

Another example described in the article that is similar in terms of supervised learning is their use of Naïve Bayes Classifier to classify job titles and descriptions. “For the first step, we used supervised machine learning techniques to develop an algorithm that classified self-reported job titles and job descriptions into an O\_NET standardized occupation code” (Sajjadiani). So similarity here is the goal of classifying, but the difference being the use of Naïve bayes, and that the classifications are already standardized into groupings, reducing some of the preliminary work.

In an article from the GCU library, they aim to use “use physical parameters for phase prediction of multi-component alloys by the help of TensorFlow Machine Learning with Limited Data” (Sarlar). Within this article they describe their use of “gradient boosted trees machine learning”, this method can be further explained as “Gradient Boosted Regression Trees (GBRT) or shorter Gradient Boosting is a flexible non-parametric statistical learning technique for classification and regression” (Steadman). So relating this form of learning back to the original article on applicants data, is that they are both aiming to classify the data.

References

Nabi, J. Machine Learning- Text Processing. Towards Data Science. Retreived from: <https://towardsdatascience.com/machine-learning-text-processing-1d5a2d638958>.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. Journal of Applied Psychology, 104(10), 1207–1225. <https://doi-org.lopes.idm.oclc.org/10.1037/apl0000405>

ŞARLAR, K. (2021). Using Physical Parameters for Phase Prediction of Multi-Component Alloys by the Help of TensorFlow Machine Learning with Limited Data. Sakarya University Journal of Science, 25(3), 766–773. <https://doi-org.lopes.idm.oclc.org/10.16984/saufenbilder.840548>

Steadman, M. Gradient Boosted Regression Trees. Data Robot. Retrieved from: https://www.datarobot.com/blog/gradient-boosted-regression-trees/