
VLADIMIR SKAVYSH, Bank of Canada

Transformer NLP Models and Quantum Computing for Classification of Receipts Data

Automated categorization of goods and services is vital to data analysis of consumer behavior, as well as calculating strategic indices for policy-making. One of the key data sources is the scanner data coming directly from points of sale, where millions of entries containing product descriptions as well as prices are logged in a relatively short period of time. We manually annotate such a dataset, the USDA food product dataset with North American Product Classification System (NAPCS) codes. Then, we propose the use of state-of-the-art neural network transformer models with multi-class softmax layers as well as one-vs-all objectives to train efficient and accurate product description classifiers. Since the data is highly imbalanced, we also experiment in using a generative pretrained language model (GPT2) to perform data augmentation for underrepresented classes and show that this indeed improves weighted accuracy. We further investigate the use of quantum computing as means to improve results in the future.