Review of Machine Learning Approaches for International Trade Flow Prediction

Prateek Rao
Dept. of Computer Science
PES University
Bengaluru, India
xrprateek@gmail.com

Rahul Ramesh
Dept. of Computer Science
PES University
Bengaluru, India
ynotrr@gmail.com

Pooshpal Baheti
Dept. of Computer Science
PES University
Bengaluru, India
pooshpalbaheti@gmail.com

***Abstract*—The concept of analyzing previous trends and making future predictions has become indispensable in every field since the advent of machine learning techniques. By scrutinizing economic trends and probing into several economic indicators, we can attempt to predict future economic events and capital flow between countries. Given that bilateral trade flow is an important factor in establishing trade policies, it is of paramount importance that we are able to forecast the flow as accurately as possible. On investigating various papers, we find that there is a growing interest in employing machine learning approaches for economic forecasting over traditional statistical methods.**

I. Introduction

International trade refers to a country's economic activity with respect to its relationship with another country. International trade flow refers to the exports and imports of a country. Bilateral trade flow alludes to international flow between two countries. In order to formulate and revise economic policies and reforms, a country's export and import values are primary indicators that are used to make the decisions. Export from a country creates employment, thereby positively effecting the GDP of a country. This makes prediction of export trends a priority for decision-making. Bilateral flow represents the value of goods and services that have been exported from one country to another. As such, it strongly influences both international trade policy and domestic economic policy in both countries. A higher rate of flow can also suggest developing relationships between the countries, which can also affect foreign policy, and vice versa. Prediction of international trade trends are crucial as they can signify economic stability, development and growth. Bilateral flow represents the value of goods and services that have been exported from one country to another. As such, it strongly influences both international trade policy and domestic economic policy in both countries. A higher rate of flow can also suggest developing relationships between the countries, which can also affect foreign policy, and vice versa. Thus, accurate prediction of bilateral flow is of prime importance. Generally, trade policies are made using a wide variety of conventional econometric techniques for forecasting flow. However, in contemporary uses, artificial intelligence and machine learning techniques are being phased into this process, as they have demonstrated supremacy over traditional approaches in respect of forecasting reliability. The aim of this study is to reflect on latest techniques used to predict trade flow, their advantages over traditional forecasting methods, and the best approaches to be taken to solve this problem.

II. Dataset

We intend to make use of the ”TradHist” dataset by CEPII [1], which provides around 1.9 million bilateral trade observations for 1827-2014, for 200+ different countries and regions. It has both geographical dimension and time dimension. Due to the huge time-frame of this data, there are a lot of missing values due to difference in data collection techniques. We need to avoid these anomalies in data, which can have a negative effect during training and prediction. We remove incomplete data from the dataset, i.e. data with missing feature values. We split data at 2009, as we intend to predict the values from 2009 - 2014, which are relatively recent trade flows. We also intend to use some advanced methods of feature engineering to select the most influential features for model training.

III. Literature Review

We first focus on the revolutionary work of Walter Isard in 1954 [2], which gave rise to the gravity model of international trade. This model predicts bilateral trade flows based on the economic sizes and distance between two units. Between two countries i and j, it takes the form of

$$F_{ij} = G.\frac{M_i^{\beta_1} M_j^{\beta_2}}{D_{ij}^{\beta_3}}$$

Here, G is a constant, F stands for trade flow, D stands for distance and M stands for economic dimensions of the countries that are being measured. The gravity model has been used throughout history to predict some important events, and it provides us with a suitable baseline for evaluation. Isaac Wohl and Jim Kennedy [3] used artificial neural networks to analyze international trade data. Making use of data between United States and it's main trade partners, they trained the model on the set from 1986 to 2006, and made predictions from 2007 to 2016. Results were reasonably close to the real values. Jingwen Sun et al. [4] analyzed Bilateral Trade Flow, but with a different target variable in mind. They used yearly import/export data from 217 countries for the 1960-2017 period, and predicted the GDP using several statistical factors. They concluded that RBF Regressor was the better model. S Circlaeys et al. [5] tested several machine learning models to predict pattern of bilateral trade flows. They used a variety of linear regression and neural network models to train on economic and geographic factors. Due to the ability of neural networks in capturing non-linear interaction among features, it performed better than any regression-based model. This paper will serve as our main point of reference. Feras Batarseh et al. [6] used machine learning algorithms to select the best economic variables that influence trade between specific commodities, and train models using these features. This type of feature engineering was found to be more accurate than statistical methods in predicting future trend patterns. Jin-kyu Jung et al [7] employed a number of ML methods as a new approach to forecast. They applied the Elastic Net, SuperLearner, and Recurring Neural Network on macro data of seven advanced and emerging economies. Results show that these ML algorithms outperformed conventional statistical approaches and offering great potential in the area of economic forecasting. Lastly, Kottou et al. [8] made use of wavelet transforms and many deep learning techniques to predict flow, with the best accuracy being given by Decision Trees and Ensemble Learning.

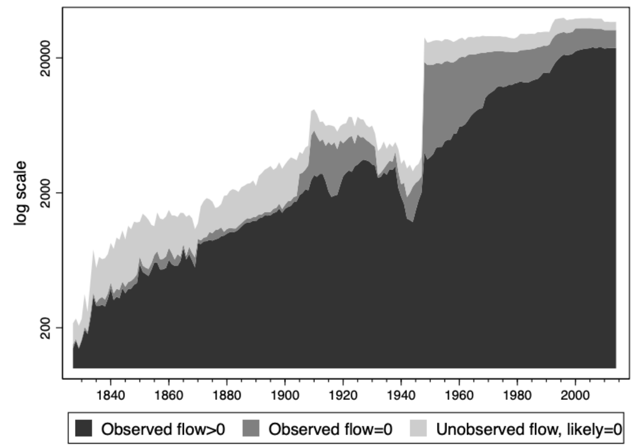

Fig. 1. Number of bilateral trade observations

IV. Exploratory Data Analysis

In this section, we will explore the data briefly in order to understand trends and uncover some patterns. First, let us examine the number of bilateral trade observations for different values of `FLOW`, which serves as our target variable.

In Fig. 1., we immediately notice that the number of observations are increasing per year, which can be attributed to three factors: First, it reflects the increasing number of existing countries, which mechanically increases the number of potential international trade flows. Second, the increasing number of observed flows reflects the increasing number of existing country pairs that are actually engaged in bilateral trade. Third, it can be a consequence of the easier access to primary sources for recent years, due to conservation issues and the difficulties in locating historical statistics for more ancient times.

Fig. 2., shows the number of null values present in the data for some of the common variables. We can

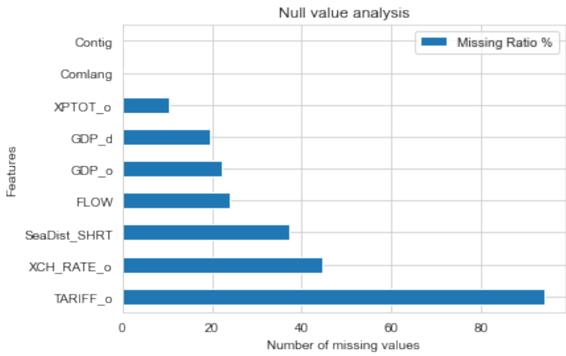

Fig. 2. Null Value Analysis

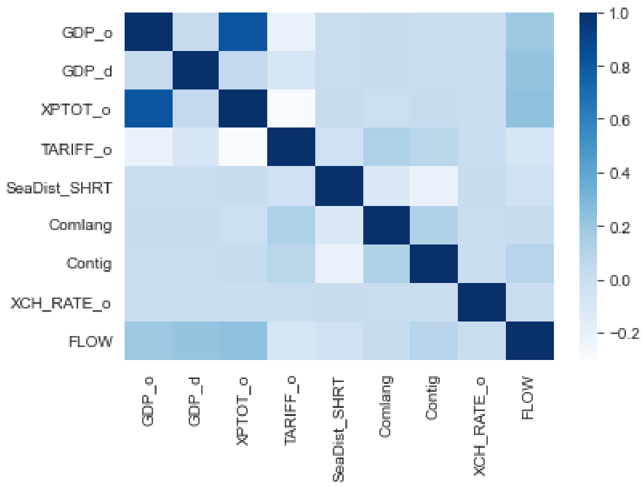

Fig. 3. Correlation Matrix for some common features

see that almost 20% of the values are missing for `GDP_o` and `GDP_d`, which tells us that we will have to remove these values from our data before model training. Similarly, other variables can be inferred.

Fig. 3. shows a correlation matrix between some common features that are used to predict bilateral trade flow. We can observe that there is a significant linear positive correlation between `GDP_o`, which is the GDP of the origin country, `GDP_d` (which is the GDP of the destination country) and `XPTOT_o` (which is the total exports of origin country), with respect to the target variable `FLOW`. A slight positive correlation can be seen between `FLOW` and the categorical variables `Comlang` and `Contig`, which suggest that there would exist greater trade flow if the participating countries are contiguous and share a common language.

We can thus infer that some of the positive correlations present in the correlation matrix can be used as primary features to train our model.

References

[1] M. Fouquin and J. Hugot, "Two centuries of bilateral trade and gravity data: 1827-2014," Working Papers 2016-14, CEPII, May 2016.

[2] W. Isard, "Location theory and trade theory: Short-run analysis," *The Quarterly Journal of Economics*, vol. 68, no. 2, pp. 305–320, 1954.

[3] I. Wohl and J. Kennedy, "Neural network analysis of international trade," *US International Trade Commission: Washington, DC, USA*, 2018.

[4] J. Sun, Y. Suo, S. Park, T. Xu, Y. Liu, and W. Wang, "Analysis of bilateral trade flow and machine learning algorithms for gdp forecasting," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3432–3438, 2018.

[5] S. Circlaeys, C. Kanitkar, and D. Kumazawa, "Bilateral trade flow prediction," *Unpublished manuscript, available for download at http://cs229. stanford. edu/proj2017/final-reports/5240224. pdf*, 2017.

[6] F. Batarseh, M. Gopinath, G. Nalluru, and J. Beckman, "Application of machine learning in forecasting international trade trends," *arXiv preprint arXiv:1910.03112*, 2019.

[7] J.-K. Jung, M. Patnam, and A. Ter-Martirosyan, *An algorithmic crystal ball: Forecasts-based on machine learning*. International Monetary Fund, 2018.

[8] E. M. Kottou, T. A. Grubelich, and X. Wang, "Bilateral trade flow prediction models enhanced by wavelet and machine learning algorithms," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1510–1516, 2020.