BILATERAL TRADE FLOW PREDICTION

Sonia Circlaeys (sonia.circlaeys@stanford.edu), Chaitanya Kanitkar (kanitkar@stanford.edu), Daiki Kumazawa (dkumazaw@stanford.edu)

PREDICTING

Bilateral trade flow (the value of exports from one country to another) is an important economic indicator used by economists and policy makers. Traditional approaches in predicting trade flow have been (1) the Gravity Model, a theoretical economic model motivated by Newton's law of universal gravitation, and (2) traditional time series models, such as AR, MA, ARIMA type models. Our project applied Machine Learning techniques to improve trade flow prediction accuracy using a variety of economic indicators, and discovered that neural networks are a promising new approach.

DATA SET AND CLEANSING

a.) Data set

Our primary data source is the bilateral trade and economic variable data set "Tradhist" distributed by CEPII. It contains 1.9 million data points from the 1800s to 2014 for 200+ different countries.

b.) Cleansing Techniques Employed

1. Removed incomplete data (i.e. we did not work with data points that have missing values).
2. Removed all trade flow values that are below 100 GBP, as they are not economically significant and cause outlier problems.
3. Took the log of data to achieve a smoother distribution of the data (See Figure 1).
4. Removed data before year 2009 as our focus is to predict recent trade flows.

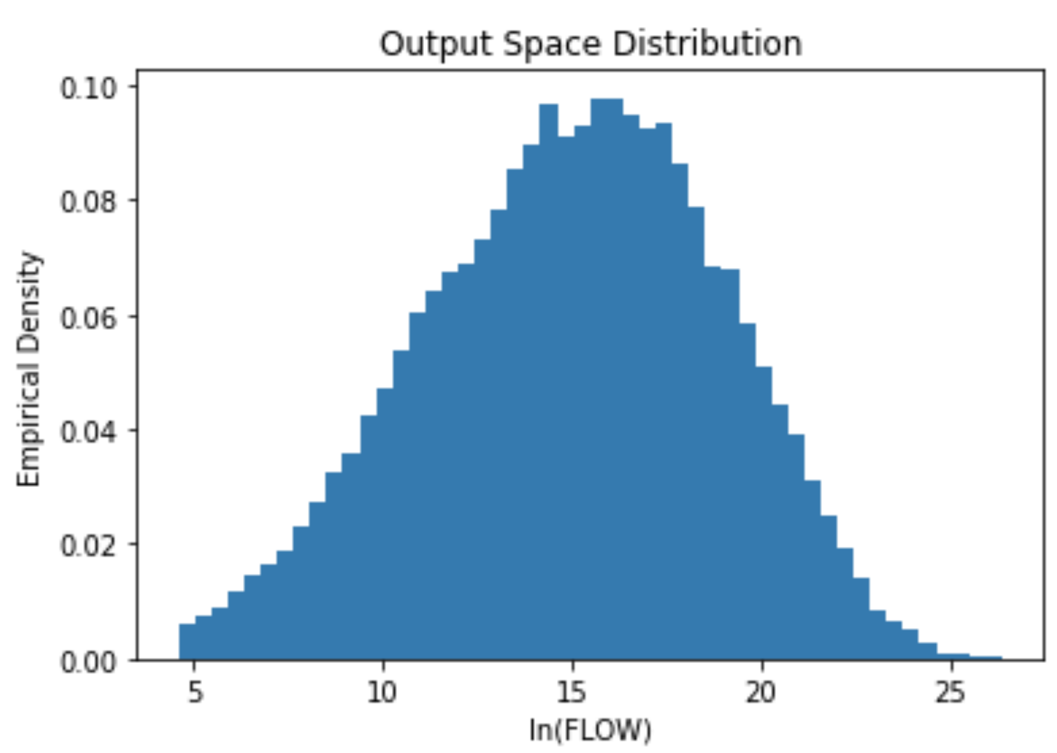

Figure 1: Histogram of the output space after cleansing.

FEATURES

Our training models use the following 12 economic features. All monetary values are in current GBP. We note that these features are well justified since they are commonly included in the Gravity Model.

GDP_o	GDP of origin country
GDP_d	GDP of destination country
Pop_o	Population of origin country
Pop_d	Population of destination country
Dist	"Great circle distance" between two countries
Comlang	1 if at least one language is spoken by more than 9% of the population of both countries
Contig	1 if two countries are contiguous
OECD_o	1 if the origin country is an OECD member
OECD_d	1 if the dest. country is an OECD member
GATT_o	1 if the origin country is a GATT member
GATT_d	1 if the dest. country is a GATT member
XPTOT_o	Total value of exports of the origin country

MODELS

1. **Ordinary Linear Regression with Regularization**
   This is our baseline model.
2. **Linear Regression based on Gravity Model**
   This model takes the following functional form:
   $$FLOW = \alpha \frac{GDP_o^{\beta} GDP_d^{\gamma}}{Dist^{\omega}}$$
   where 'FLOW' is our output. Taking the log yields
   $$ln(FLOW) = \alpha' + \beta' ln(GDP_o) + \gamma' ln(GDP_d) + \omega' ln(Dist)$$
   which is linear in parameters.
3. **Linear Regression with RBF and polynomial kernels**
4. **Neural Network using Gravity Model features**
   Since we assume no a priori knowledge on interrelations among features, we utilized a **fully connected** neural network. (See Figure 2).
5. **Neural Network using Gravity Model features + lagged values of bilateral trade**
   This last approach follows the spirit of AR models, which include the lagged values of the trade flow.

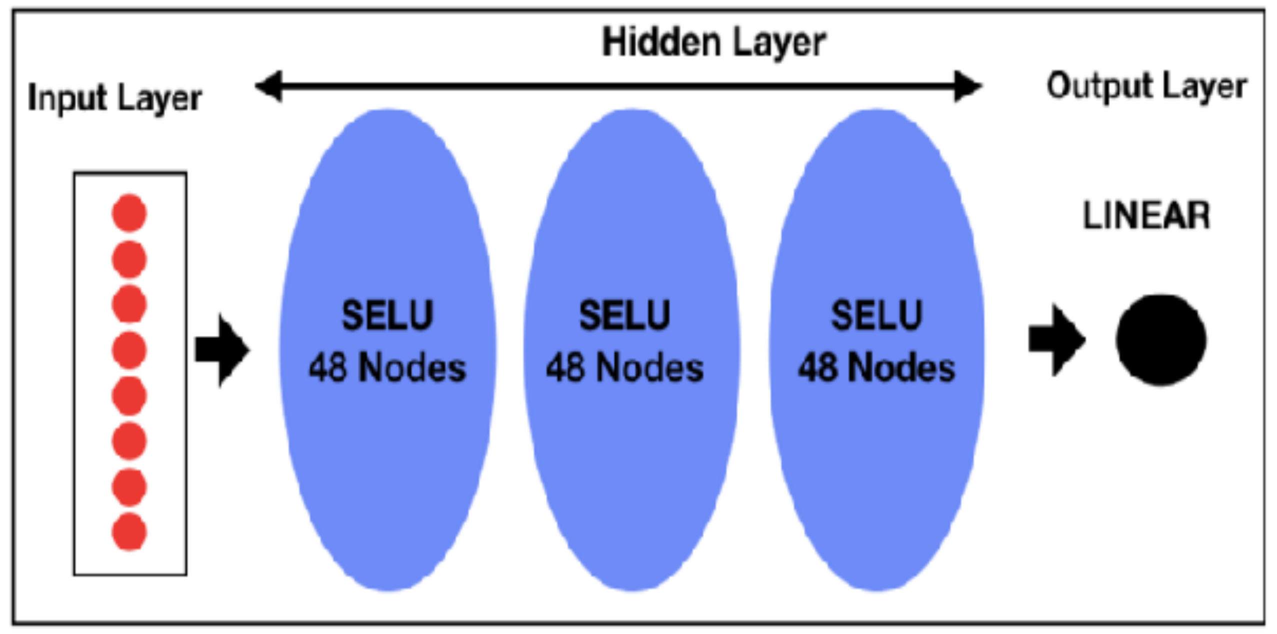

Figure 2: Model 4 (Fully connected neural network)

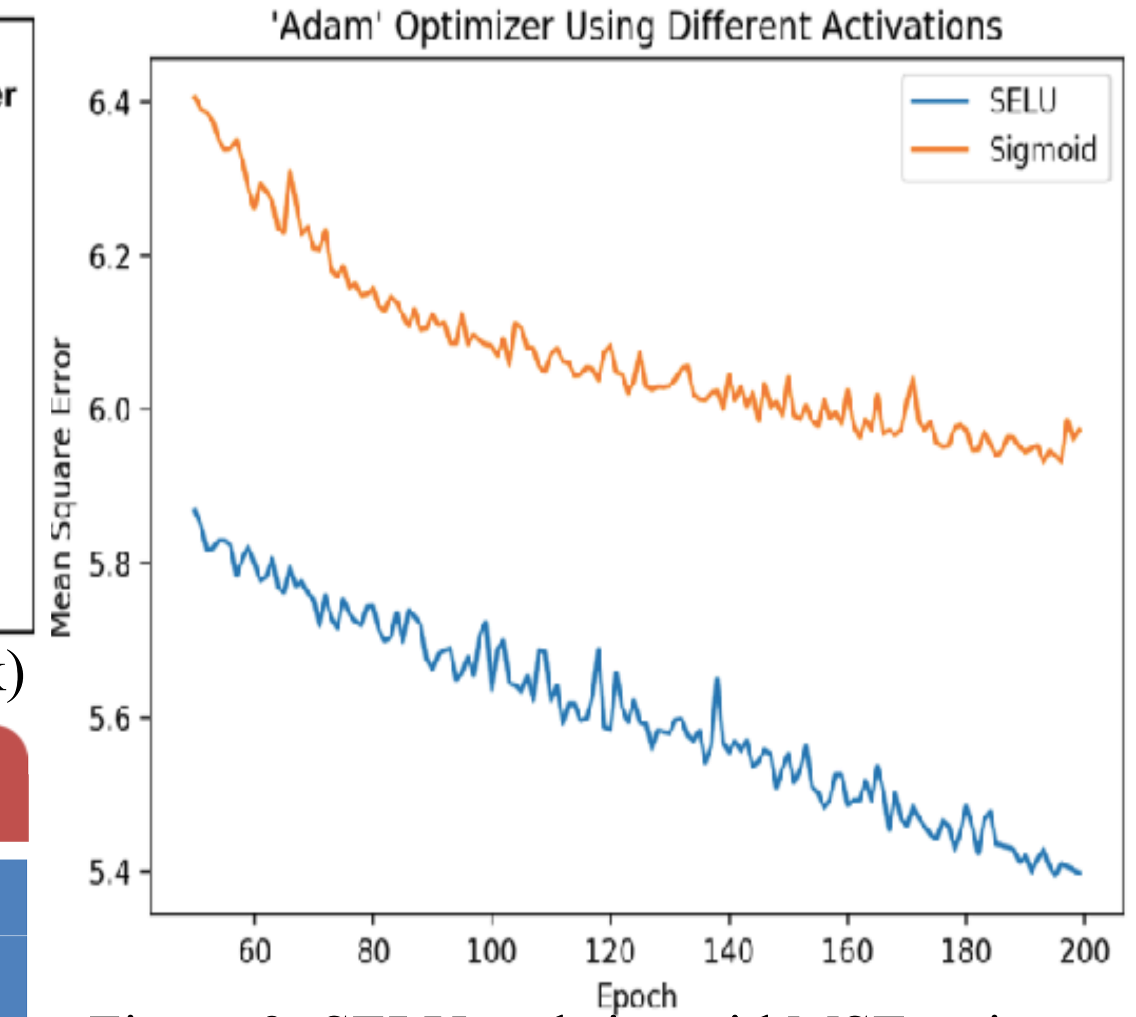

Figure 3: SELU and sigmoid MSE trajectory

RESULTS

	Train		Test	
Model	MSE	R^2	MSE	R^2
1	1.4x10e19	0.14	1.6x10e19	0.13
2	6.45	0.64	6.53	0.63
3	5.07	0.72	8.38	0.61
4	4.97	0.73	5.01	0.73
5	1.57	0.90	1.59	0.90

The original data set is separated into the train set of 91,747 examples (70%) and test set of 39,321 examples (30%). For the kernelized linear regression, we have constrained our analysis to a subset of the training set due to computational issue.

OBSERVATION & DISCUSSION

1. Kernel Regression did not perform well in practice. There were simply too many parameters to train given our large data set.
2. Wider neural network performed better than deeper ones. This was consistent with our initial hypothesis that casting the feature space into a higher dimension would capture non-linear interactions between features.
3. Regularizer successfully reduced the variance issue.
4. The neural net with Gravity Model features (Model 4) achieved a superior predictive performance than Gravity Model (Model 2), successfully capturing non-linear relations.
5. Optimization issue with neural network was addressed with the use of more effective activation functions (SELU as opposed to sigmoids) and optimizer (See Figure 3).
6. Model 5 achieved the best performance among all models.

FUTURE

- Since our data has both geographical and time dimensions, using the LSTM model extended to a panel setting (for example, the model developed by Bai, Zhang, and Gao (2017)) may be a fruitful avenue that we can pursue in the future.
- Discretizing our output space rather than using a continuous output space may produce more accurate results.

References

Bai, M., Zhang, B., & Gao, J. (2017). Tensorial Recurrent Neural Networks for Longitudinal Data Analysis. https://arxiv.org/pdf/1708.00185.pdf