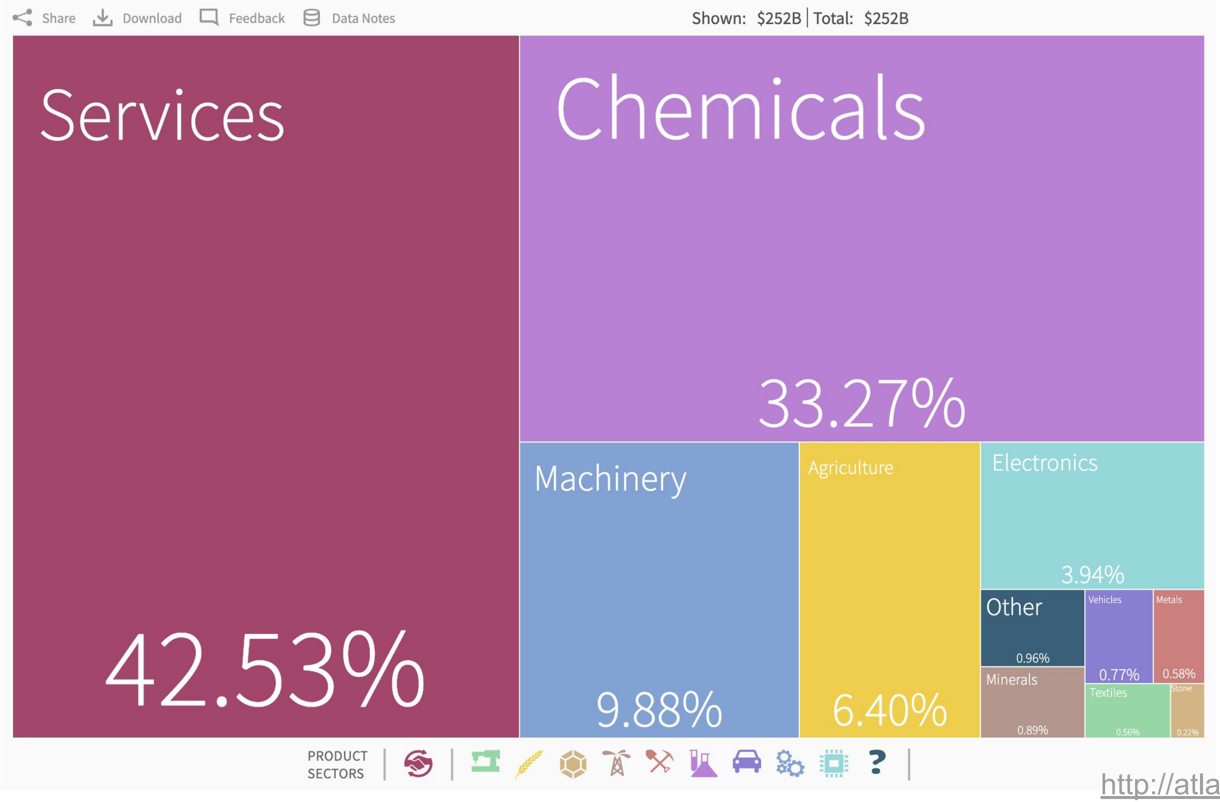
Time-Series Recommender:

Recommending export growth across time, countries, and products

Faran Sikandar
Insight AI Fellow
August 2019

What did Ireland export in 2012?



http://atlas.cid.harvard.edu/

Recommend exports across time

Help investors identify potential in untapped product and service areas.

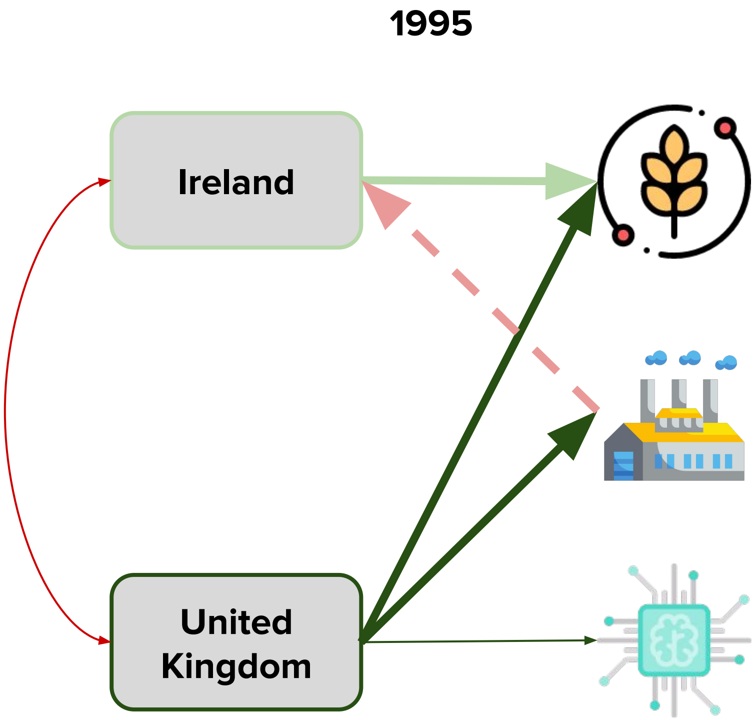
Help countries identify growth opportunity.

Why Is It Hard to Make Recommendations Over Time?

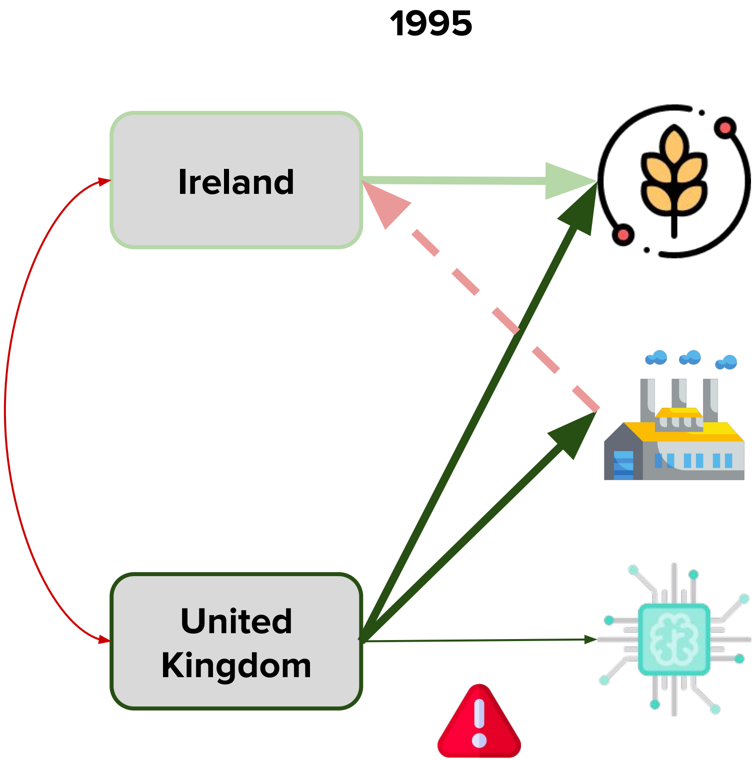
- **Recommenders:** User-item recommendations
  - Across multiple categories
  - **BUT** time-static

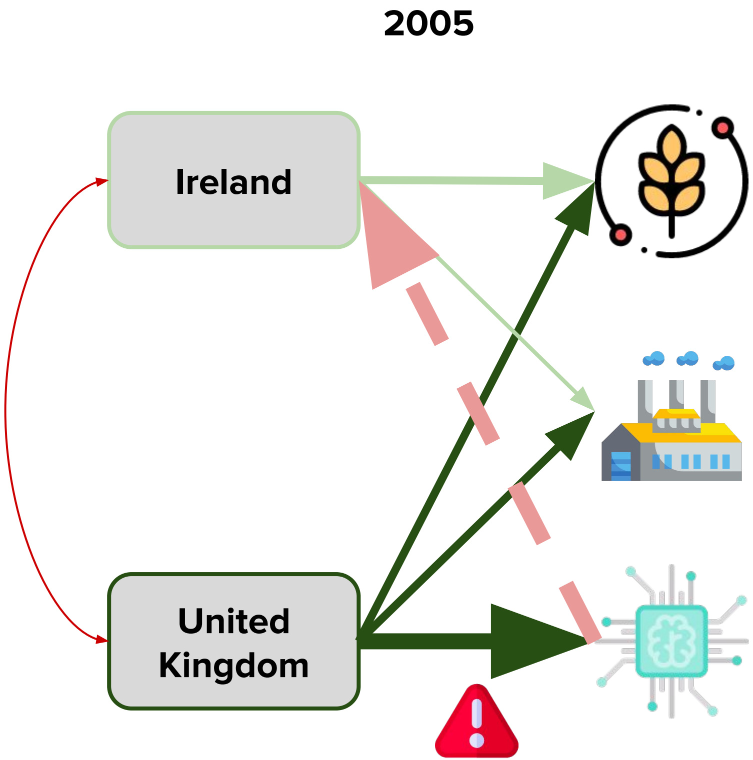
- **Time-Series Forecasting:** Predictions
  - Time-dynamic
  - **BUT** only single category

Collaborative Filtering Over a Single Period of Time



Collaborative Filtering Over Multiple Time Periods





https://www.flaticon.com/

ATLAS DATA: Reducing Search Costs for Complexity Gain

250
COUNTRIES /
TERRITORIES

×

5,000+
PRODUCTS

×

20 YEARS

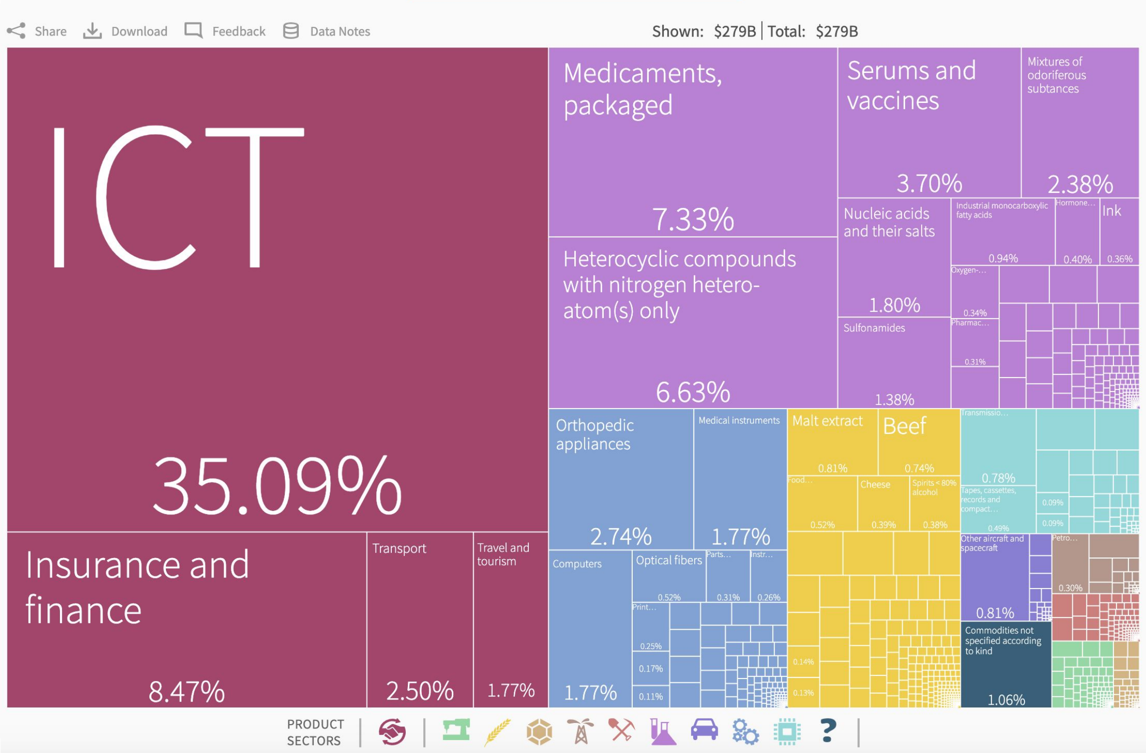
25 MILLION RECORDS

Results: Top Growth Export Areas, by % Change

Recommendations for Ireland for 2005-2014

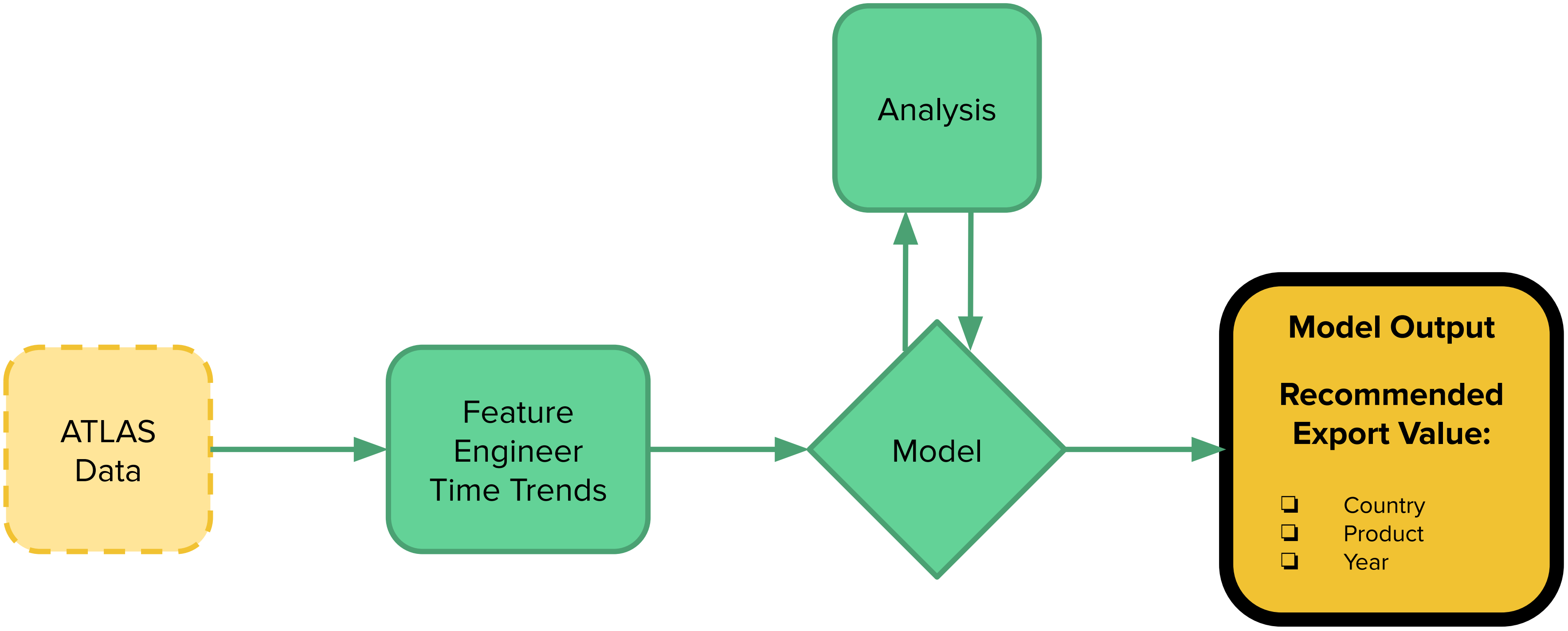
	index	product_id	name	export_pct_change	pct_change_rank	predictions	pred_pct_change
0	5397159	11003	Information and communications technology	6.272000e+11	50.0	2.926768e+09	2.926768e+09
1	5397179	11002	Transport services	4.814000e+10	48.0	2.550333e+09	2.550333e+09
2	5397169	11004	Insurance and financial services	2.013000e+11	49.0	3.547189e+08	3.547189e+08
3	5379089	8233	Cold rolled iron or non-alloy steel, coil, wid...	1.092420e+06	25.0	8.422073e+07	8.422073e+07
4	5366509	6975	Chem wood pulp, soda/sulphate, non-conifer, bl...	3.664330e+05	19.0	7.250805e+07	7.250805e+07
5	5378899	8214	Hot rolled iron or non-alloy steel, coil,width...	2.238962e+06	32.0	4.185072e+07	4.185072e+07

What Did Ireland Export in 2015?

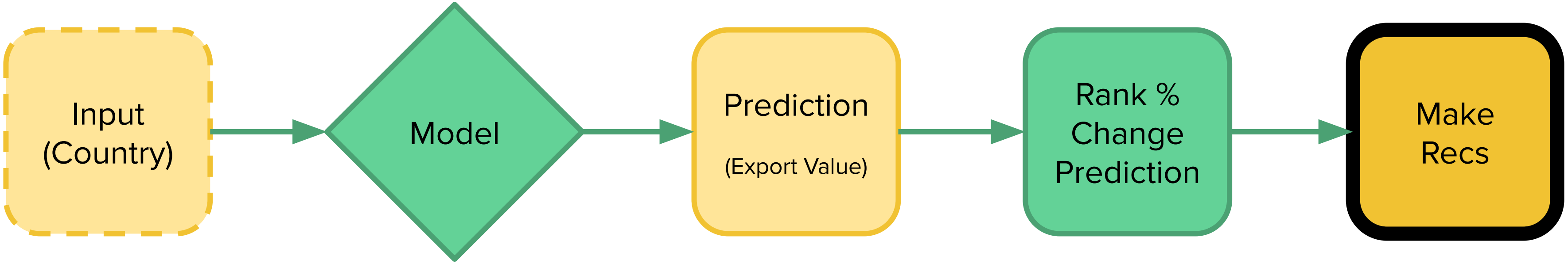


http://atlas.cid.harvard.edu/

Training Pipeline



Inference Pipeline



Approaches - Tested		Ease	Accuracy	Time Dimension
1	**Normalization of Target Variable** (Value of Exports)	✓	✓	✗
2	**Shallow Model** (Dot Product)	✓	✗	✗
3	**Fully-Connected Deep Learning Model**	✓	✓	✗
★ 4	**DL + Time-Series Model** (Using Trend Inputs)	✗	✓	✓

Validation: Comparing Ranks

Model	**Cosine Similarity** Export Values vs Predictions	**Rank Cosine Similarity** Export % Change Rank vs Predictions
KNN	0.024	0.825
1-Layer Dot Product Neural Net	0.624	0.664
3-Layer Neural Net	0.543	0.673
3-Layer Neural Net + Time Features	0.531	0.661
5-Layer Neural Net	0.514	**0.854**
5-Layer Neural Net + Time Features	0.524	**0.854**

Faran Sikandar

Harvard University
MPA/ID Quantitative Economics

University of California at Berkeley
BS Business Administration + BA Philosophy

NIH

J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

MIT D-Lab

SEAL OF THE CITY AND COUNTY OF SAN FRANCISCO
ORO EN PAZ FIERRO EN GUERRA

NUMIDA

APPENDIX

	Approaches - To Do	Ease	Accuracy	Time Dimension	Notes
1	**LSTM / GRU Architectures**	?	?	✓	• Prediction across **SINGLE categories** • **Collaborative filtering DILUTED?**
2	**Oversampling / Undersampling** (Minority Class)	✓	✓	✗	• **EASY to filter** to dense dataset, but **LIMITS generalizability**
3	**Increasing Frequency of Time Features**	✗	✓	✓	• May lead to **overfitting**
4	**Graph Alignment** (DL Training + Feature Engineering)	✗	✓	?	• **STRONGER country comparisons** • **Human INTERPRETABLE** • May be **redundant**

Raw Data Structure

	location_id	product_id	year	export_value	export_total_loc_year	export_period1	export_period2
0	0	100	1995	20473.0	1.292147e+09	37003.0	711479.0
1	0	100	1996	0.0	1.546077e+09	37003.0	711479.0
2	0	100	1997	0.0	1.798469e+09	37003.0	711479.0
3	0	100	1998	0.0	1.541506e+09	37003.0	711479.0
4	0	100	1999	16530.0	1.864731e+09	37003.0	711479.0
5	0	100	2000	14033.0	3.437210e+09	37003.0	711479.0
6	0	100	2001	11856.0	3.144972e+09	37003.0	711479.0
7	0	100	2002	9776.0	2.227290e+09	37003.0	711479.0
8	0	100	2003	576713.0	2.811004e+09	37003.0	711479.0
9	0	100	2004	99101.0	3.646478e+09	37003.0	711479.0
10	0	101	1995	30728.0	1.292147e+09	57424.0	77439.0
11	0	101	1996	0.0	1.546077e+09	57424.0	77439.0
12	0	101	1997	0.0	1.798469e+09	57424.0	77439.0

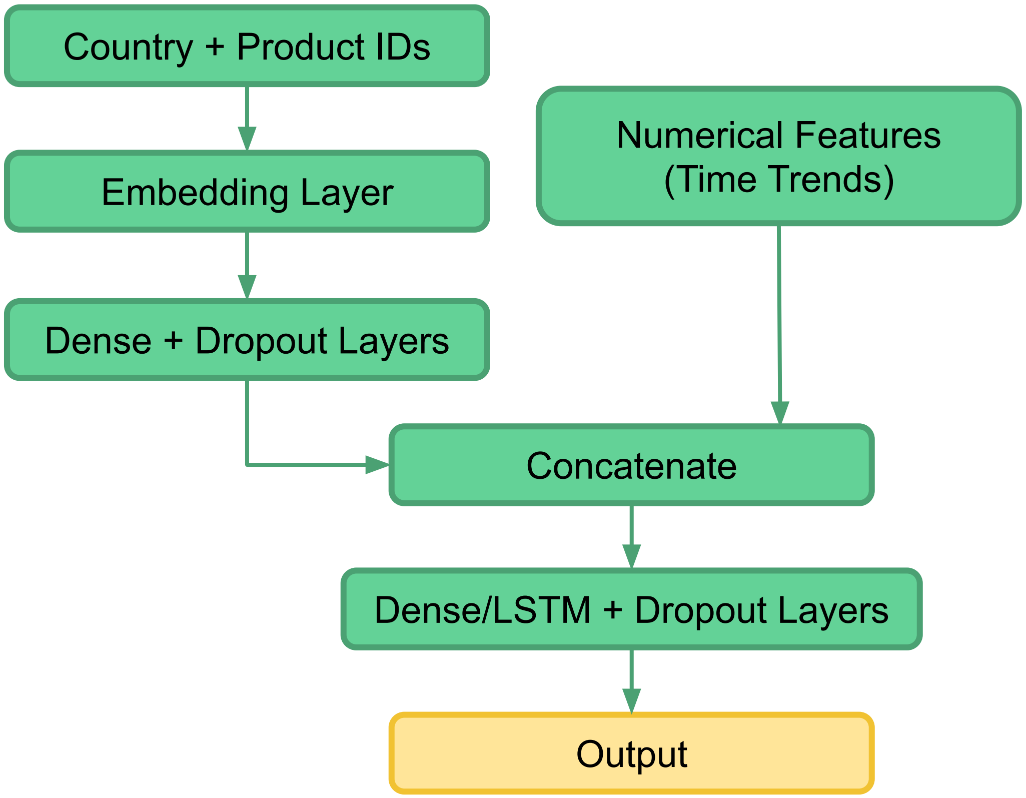
Time Series: Feature Engineering

10 Years Training Data: Export Value per Product Area

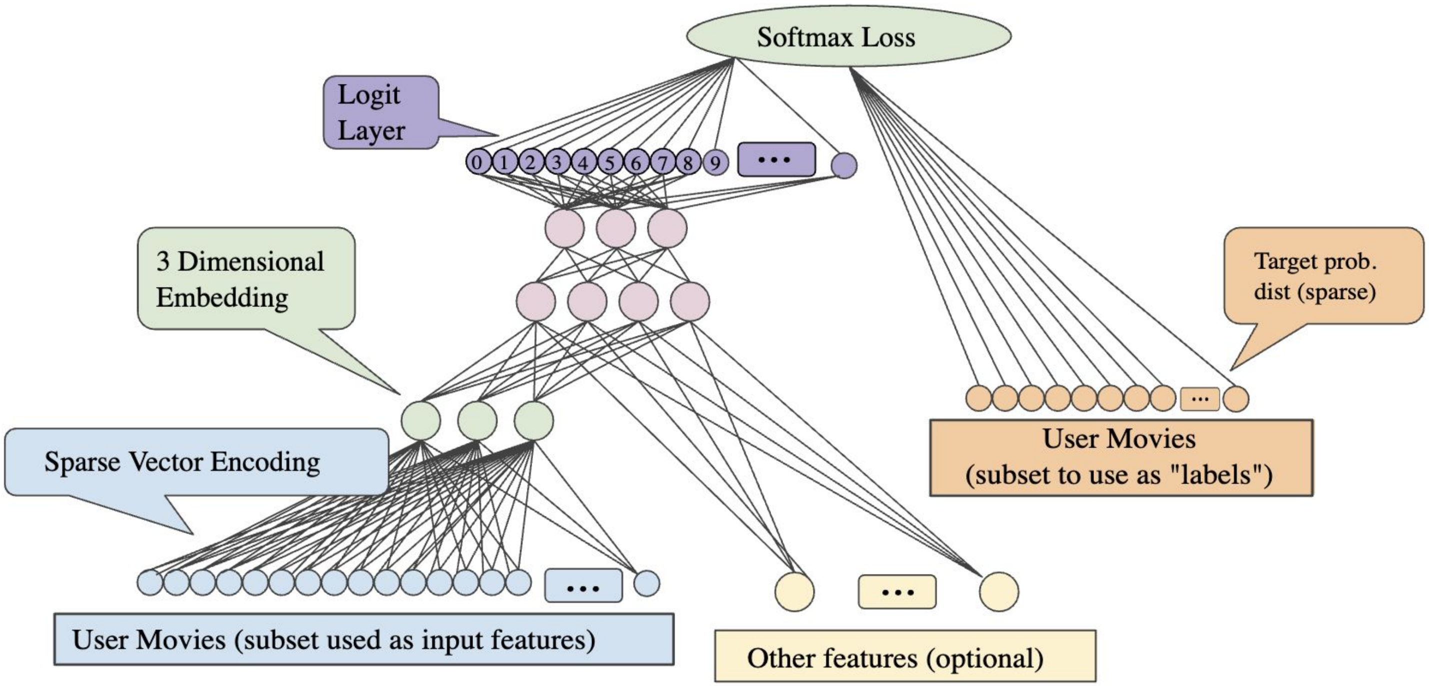
Second 5 Years: Sum Exports per Product Area − First 5 Years: Sum Exports per Product Area

= Overall Trend: Straight Line Growth (Like a Simple Derivative)

Neural Net Model Architecture



How Embeddings Work - User-Movie Collaborative Filtering



How Embeddings Work - User-Movie Collaborative Filtering

Deep Network

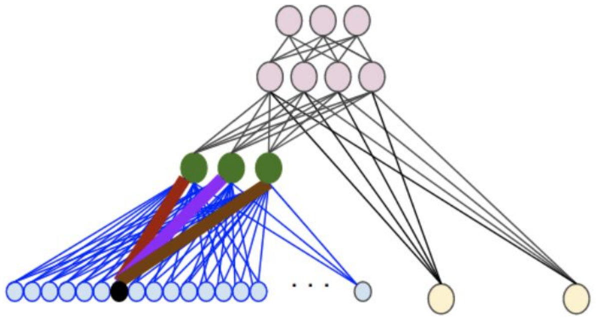
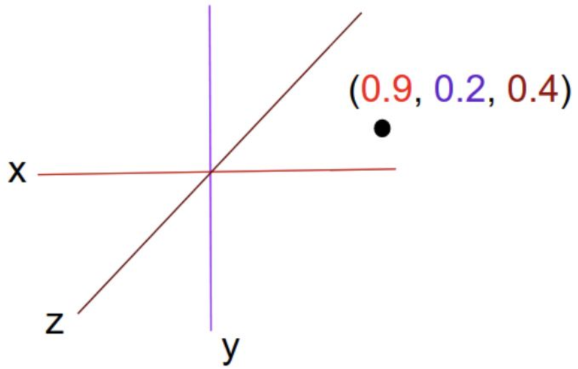
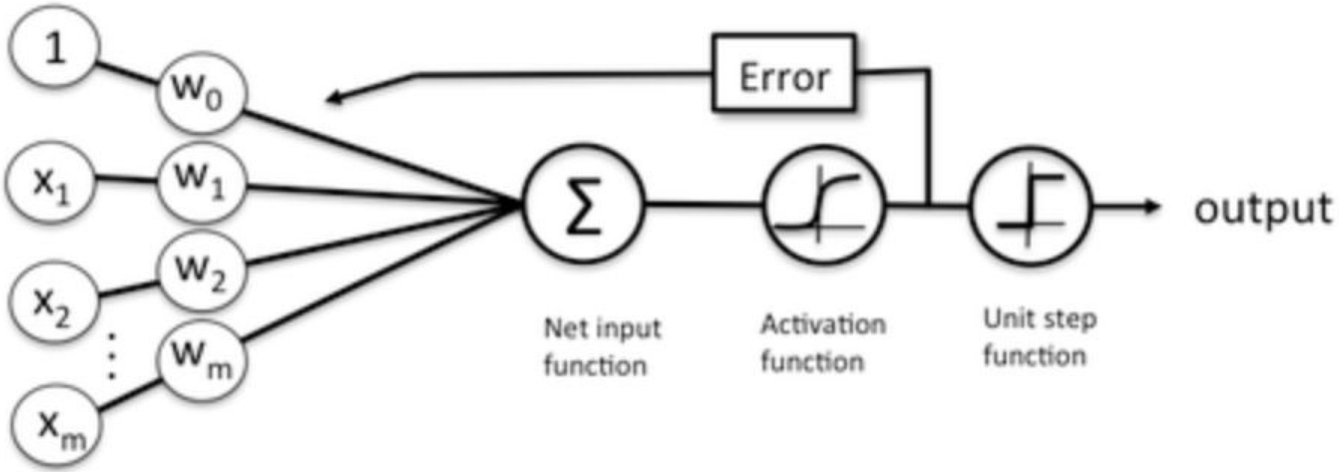


Figure 6. A geometric view of the embedding layer weights.

Geometric view of a single movie embedding

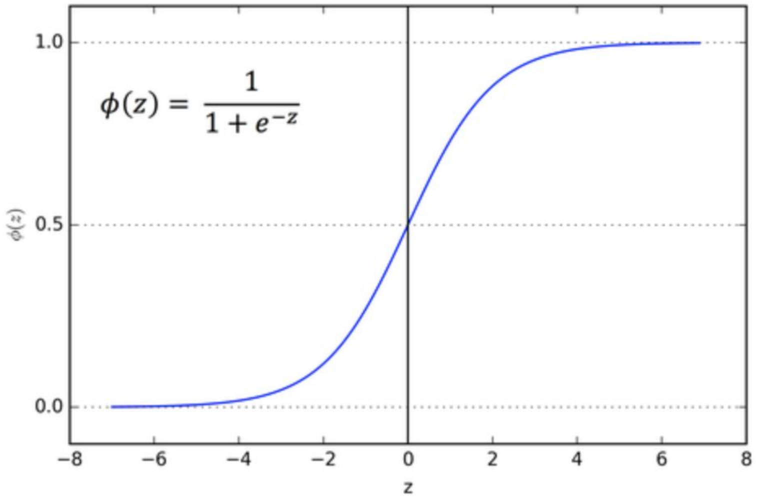


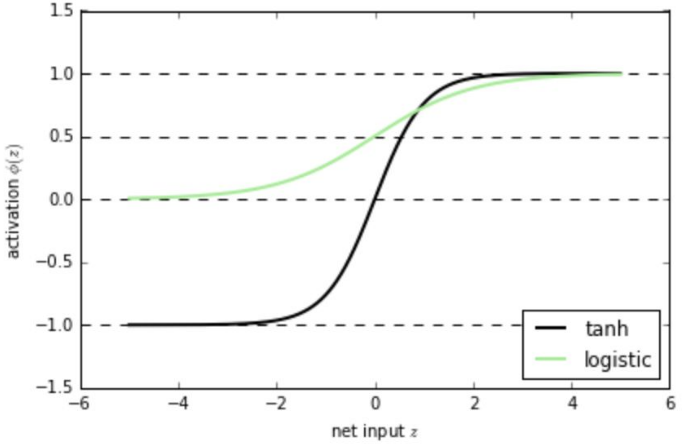
Neural Networks As Stacked Regression Nodes



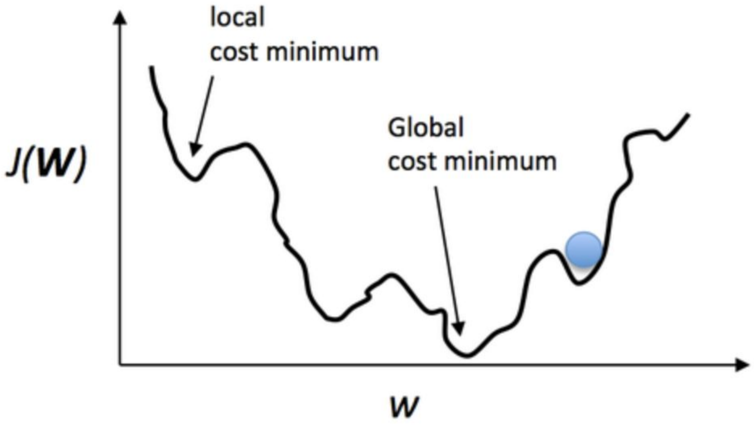
Schematic of a logistic regression classifier.

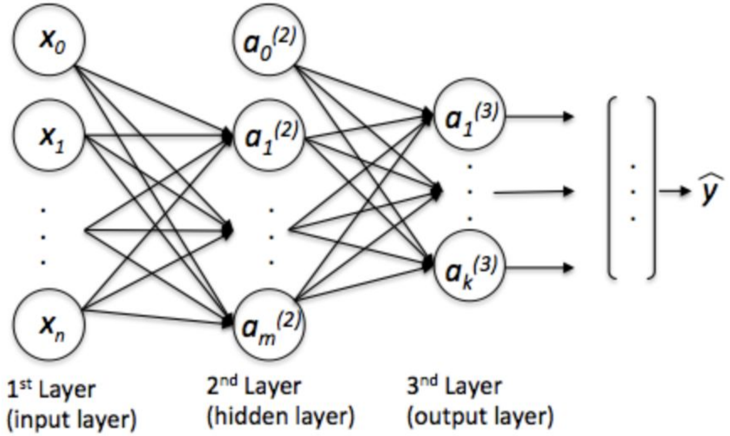
Sigmoid Activation + Tanh Activation





Local Minima + Multi-Layer Perceptron





Schematic of a multi-layer perceptron.

Validation: Potential Methods

- ❏ **Cosine Similarity** - Comparing predictions with what actually happened
- ❏ **Comparing Actual Ranks with Predicted Ranks** - Eyeball estimate
- ❏ **Develop a Country Similarity Score** - E.g. Graph alignment or using various World Bank Development Indicators
- ❏ **Forecast Individual Product Areas** - Compare forecasts for individual product areas with the predicted export values

Validation Metrics

Metric	Implemented	Tradeoffs
Mean Squared Error	Yes	Large values, difficult to interpret
Cosine Similarity	Yes	Fixed [-1, 1] range; Sensitive to small perturbations
Rank Cosine Similarity	Yes	Accounts for order, but not recall or coverage
MAP@K (Mean Average Precision @ Cutoff K)	No	Unclear if it technically applies to econ problem; Classification-based
MAR@K (Mean Average Recall @ Cutoff K)	No	Unclear if it technically applies to econ problem; Classification-based

Shortcomings to Collaborative Filtering

	Reference: Bangladesh	Comparison: Pakistan	Will Bangladesh Break In?
Agriculture	Yes	Yes	X
Electronic Circuits	No	No	?
Garments/Light Manufacturing	Yes	No	?
Information Technology	No	Yes	?
Soccer Balls	No	Yes	?

Mean Squared Error

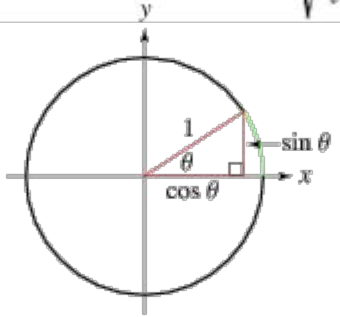
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2$$

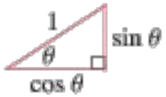
where N is the number of data points, f_i the value returned by the model and y_i the actual value for data point i.

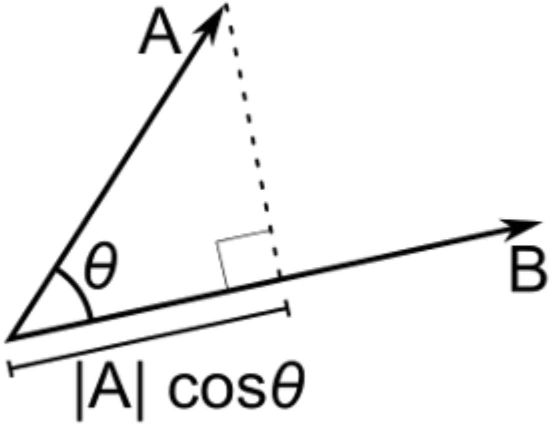
https://www.researchgate.net/figure/Mean-Squared-Error-formula-used-to-evaluate-the-user-model_fig1_221515860

Cosine Similarity - Between Vectors

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$







http://www.ahmetyildirim.com.tr/2018/03/simple-cosine-similarity-for-php.html

MAP (Mean Average Precision) - How Relevant Recs Are

TU WIEN

MAP

= relevant documents for query 1

Ranking #1

Recall	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.8	1.0
Precision	1.0	0.5	0.67	0.5	0.4	0.5	0.43	0.38	0.44	0.5

= relevant documents for query 2

Ranking #2

Recall	0.0	0.33	0.33	0.33	0.67	0.67	1.0	1.0	1.0	1.0
Precision	0.0	0.5	0.33	0.25	0.4	0.33	0.43	0.38	0.33	0.3

$$\textit{average precision query 1} = (1.0 + 0.67 + 0.5 + 0.44 + 0.5)/5 = 0.62$$
$$\textit{average precision query 2} = (0.5 + 0.4 + 0.43)/3 = 0.44$$

$$\textit{mean average precision} = (0.62 + 0.44)/2 = 0.53$$

44

FACULTY OF !NFORMATICS

https://www.slideshare.net/JosRamnRosViqueira/information-retrieval-evaluation-64158263

Data: The Atlas of Economic Complexity

Meta-Data

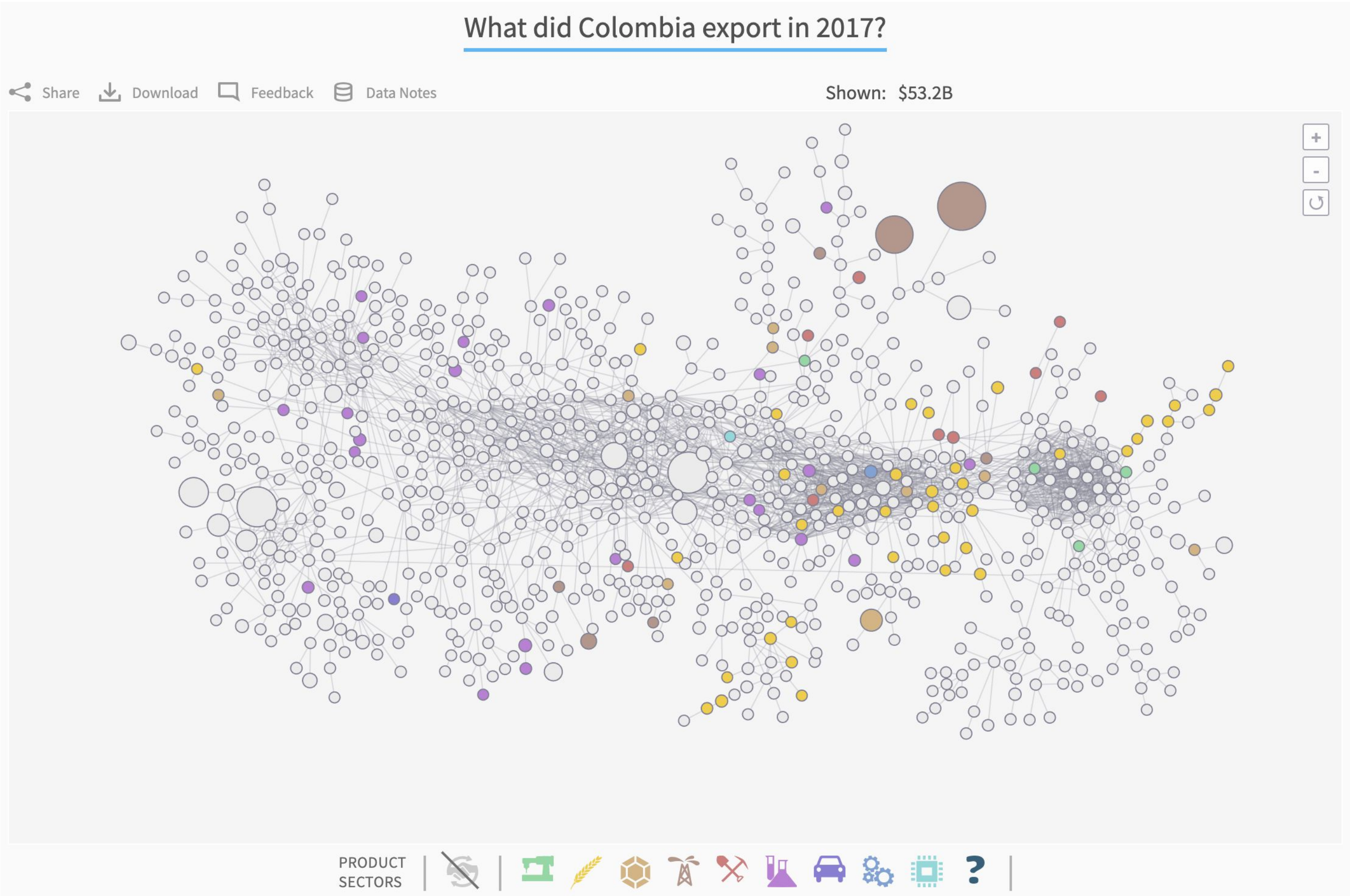
- 250 countries
- 20 categories of goods and 5 categories of services
- 6,000 products
- 1 MM + edges between product nodes in a year
- Data from 1995-2017

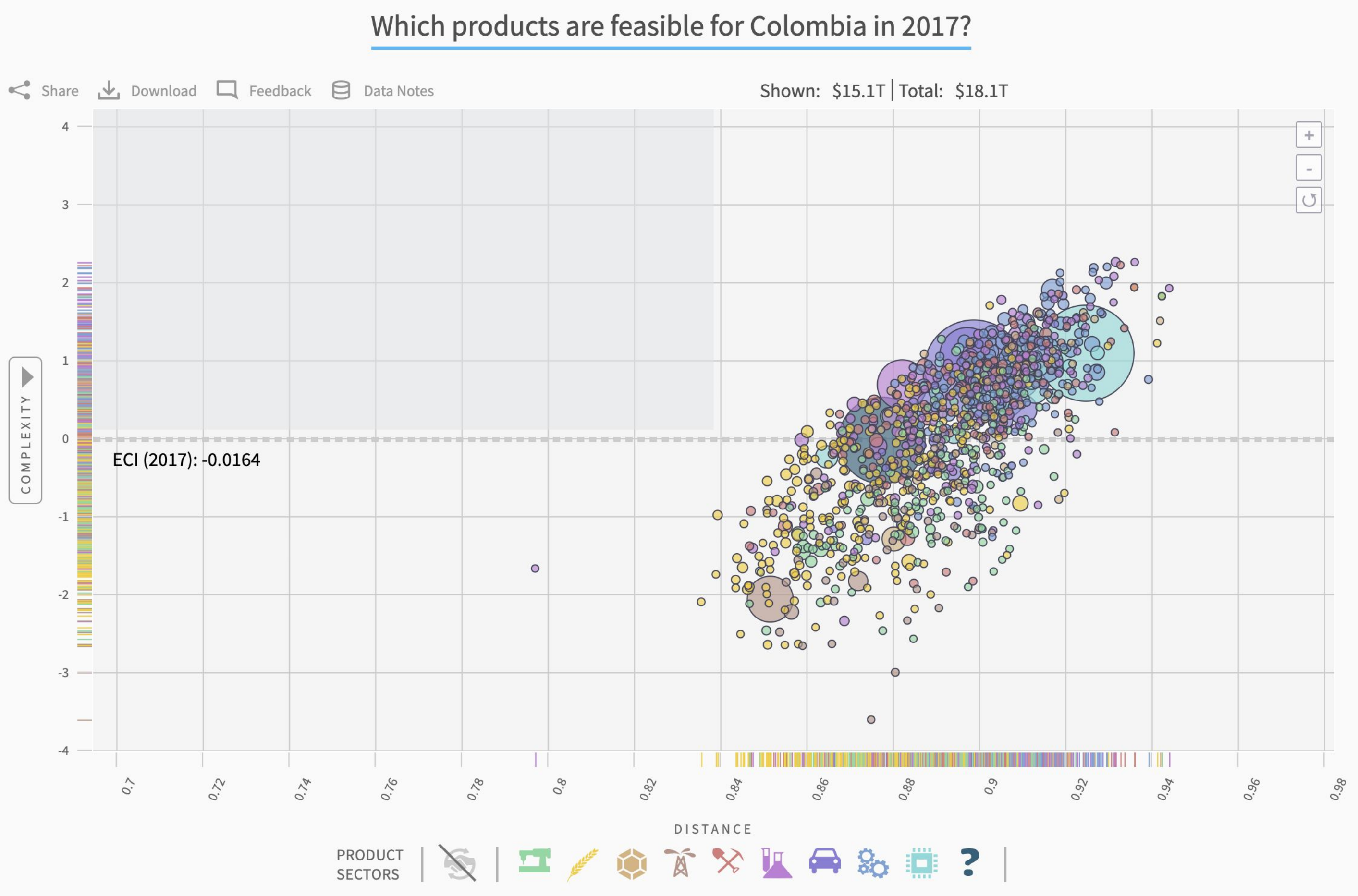
Core Metrics for Recommendation

- Proximity between product nodes (edge weights)
- Import/export volumes (user ratings)

Metrics for Model Tuning

- Complexity metrics (Atlas)
- Comparative advantage metrics (Atlas)
- Network similarity score (graph alignment)
- World Bank Development Indicator data


What did Colombia export in 2017?
Share
Download
Feedback
Data Notes
Shown: $53.2B
+
-
PRODUCT
SECTORS
?



Which products are feasible for Colombia in 2017?
Share
Download
Feedback
Data Notes
Shown: $15.1T | Total: $18.1T
COMPLEXITY
4
3
2
1
0
-1
-2
-3
-4
ECI (2017): -0.0164
0.7
0.72
0.74
0.76
0.78
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
DISTANCE
PRODUCT SECTORS
?

Data Issues - Class Imbalance

Imputing 0s for NaNs

	export_value	import_value	location_id	product_id	year
count	1.266075e+06	1.266075e+06	1.266075e+06	1.266075e+06	1266075.0
mean	4.450776e-05	4.737319e-05	1.249981e+02	7.522347e+03	2008.0
std	1.850794e-03	1.845512e-03	7.245527e+01	1.457730e+03	0.0
min	0.000000e+00	0.000000e+00	0.000000e+00	5.000000e+03	2008.0
25%	0.000000e+00	0.000000e+00	6.200000e+01	6.261000e+03	2008.0
50%	0.000000e+00	3.740213e-08	1.250000e+02	7.522000e+03	2008.0
75%	9.520578e-08	1.851461e-06	1.880000e+02	8.783000e+03	2008.0
max	1.000000e+00	1.000000e+00	2.500000e+02	1.100400e+04	2008.0

Dropping NaNs

	export_value	import_value	location_id	product_id	year
count	4.963030e+05	4.963030e+05	496303.000000	496303.000000	496303.0
mean	3.645034e+07	3.563083e+07	123.226773	7688.407447	2007.0
std	9.064218e+08	8.159947e+08	70.998969	1464.348734	0.0
min	1.700000e-01	-1.612358e+07	0.000000	5000.000000	2007.0
25%	9.507500e+03	1.128100e+05	64.000000	6497.000000	2007.0
50%	1.358000e+05	9.158600e+05	120.000000	7792.000000	2007.0
75%	2.157591e+06	6.031491e+06	181.000000	8966.000000	2007.0
max	2.516812e+11	2.451209e+11	250.000000	11004.000000	2007.0

Validation: Comparing Ranks

Model	**Cosine Similarity** Export Values vs Predictions	**Cosine Similarity** Export % Change vs Predictions	**Rank Cosine Similarity** Export Values Rank vs Predictions
KNN	0.024	0.008	0.825
1-Layer Dot Product Neural Net	0.624	0.733	0.664
3-Layer Dot Neural Net			
3-Layer Neural Net	0.543	0.730	0.673
3-Layer Neural Net + Time Features	0.531	0.730	0.661
3-Layer Dot Neural Net + Time Features	0.198	0.446	0.682
5-Layer Neural Net	0.514	0.709	0.854
5-Layer Neural Net + Time Feat (Class OR Rank)			
5-Layer Neural Net + Time Features	0.524	0.691	0.854
5-Layer Neural Net + Time Features v2	0.541	0.645	0.846

Validation: Comparing Ranks

Model	**Cosine Similarity** Export % Change vs Predictions	**Rank Cosine Similarity** Export % Change Rank vs Predictions
KNN	0.008	0.825
1-Layer Dot Product Neural Net	0.733	0.664
3-Layer Neural Net	0.730	0.673
3-Layer Neural Net + Time Features	0.730	0.661
5-Layer Neural Net	0.709	**0.854**
5-Layer Neural Net + Time Features	0.691	**0.854**

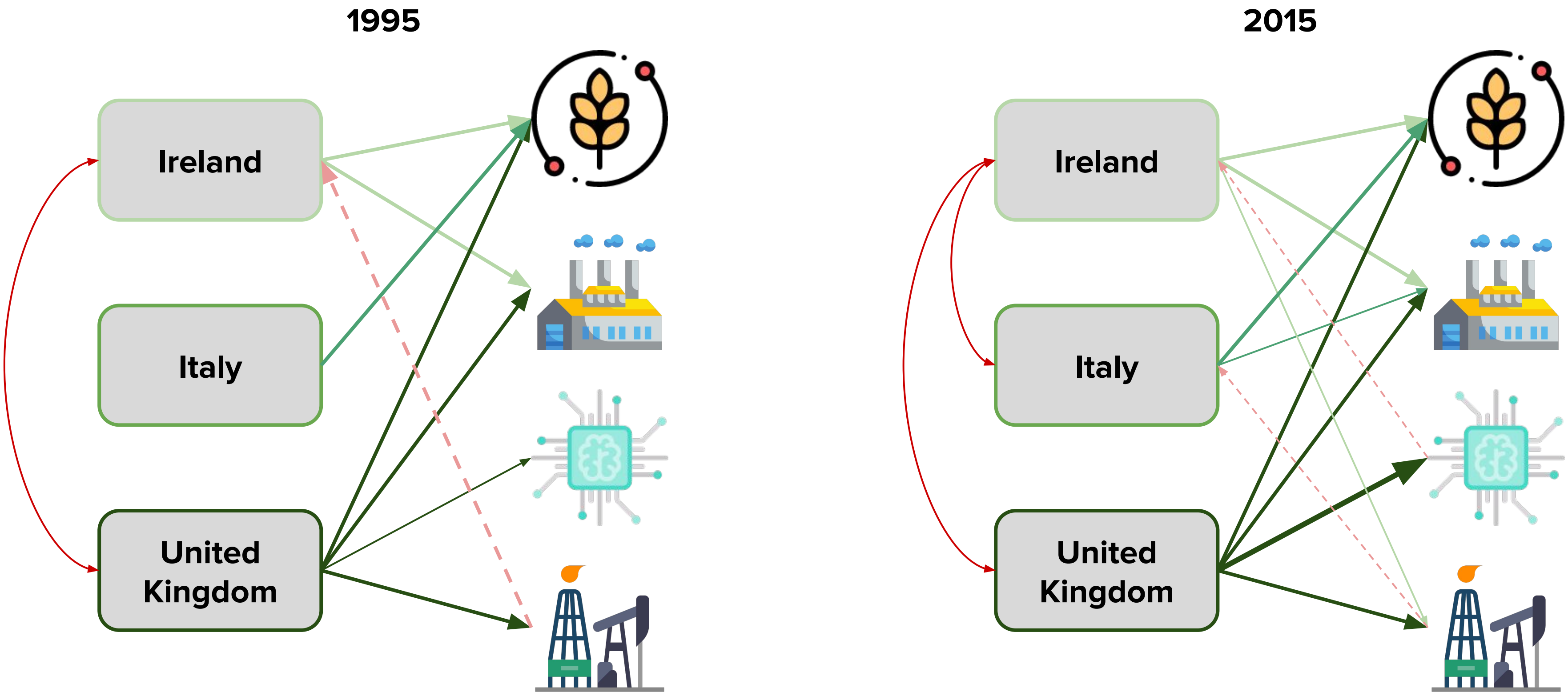
Results: Top N Growth Export Areas, by % Change

Recommendations for Ireland for 2005-2014

	product_id	name	export_period_test	export_pct_change	predictions	pred_pct_change
5397159	11003	Information and communications technology	6.272000e+11	6.272000e+11	967637504.0	967637504.0
5397179	11002	Transport services	4.814000e+10	4.814000e+10	967637504.0	967637504.0
5397169	11004	Insurance and financial services	2.013000e+11	2.013000e+11	634600512.0	634600512.0
5379089	8233	Cold rolled iron or non-alloy steel, coil, wid...	1.092420e+06	1.092420e+06	37569908.0	37569908.0
5397149	10039	Trade data discrepancies	1.047290e+07	1.047290e+07	32325480.0	32325480.0
5378899	8214	Hot rolled iron or non-alloy steel, coil,width...	2.238962e+06	2.238962e+06	32185160.0	32185160.0
5366509	6975	Chem wood pulp, soda/sulphate, non-conifer, bl...	3.664330e+05	3.664330e+05	31242950.0	31242950.0
5349919	5316	Coffee, roasted, not decaffeinated	1.186287e+08	1.186287e+08	27821750.0	27821750.0
5397119	10036	Antiques older than one hundred years	5.220432e+07	5.220432e+07	27130038.0	27130038.0
5362689	6593	Finishing agents, dye carriers, dressing, mord...	5.791601e+06	5.791601e+06	26423296.0	26423296.0
5355559	5880	Silicon, <99.99% pure	3.592090e+05	3.592090e+05	23307810.0	23307810.0

SUB-APPENDIX

Collaborative Filtering Over Time

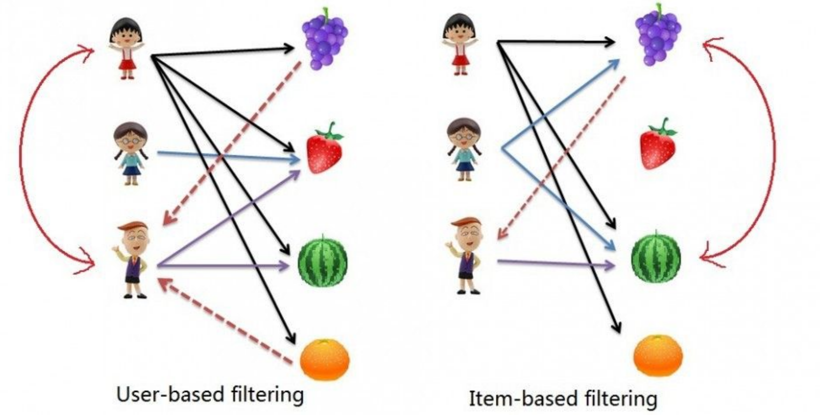


https://www.flaticon.com/

Model: Looking Across Time

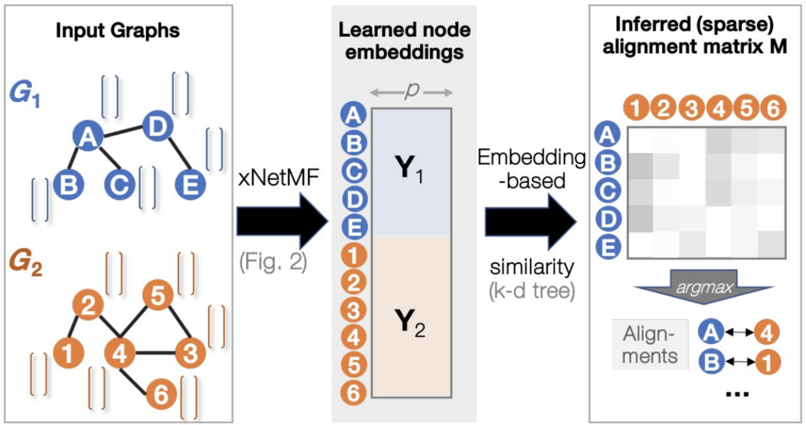
Making User-Item Based Recommendations

Collaborative filtering model to create a recommender system

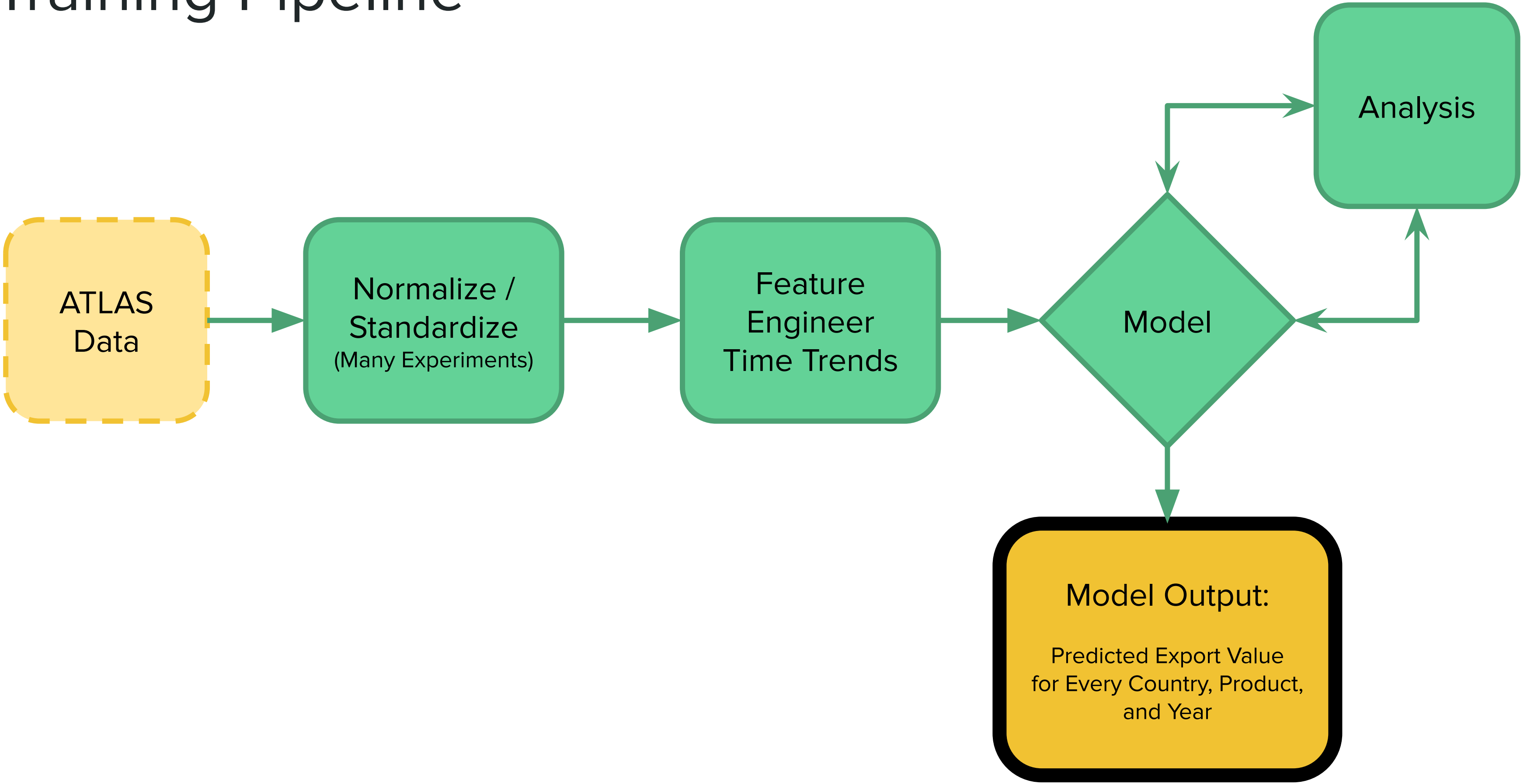


Comparing Networks at a Macro Level

Representation learning based graph alignment (i.e. creating a network similarity score)



Training Pipeline



Challenges

- **Economic Theory** - Deciding on targets: Value of exports or percent of exports?
- **Imbalanced Data / Normalization** - Class imbalance + high variance
- **Time Series** - Incorporating an RNN-type structure into the data across multiple categories difficult
- **Validation** - No validation set and want to train the data on the full structure of the economy.
- **Graph Embeddings** - Develop a graph alignment score

Solutions

- **Economic Theory** - Compare similar export percentages - possibly yields more information between differently sized economies
- **Imbalanced Data / Normalization** - Standardization by country did the trick!
- **Time Series** - Engineer a feature representing the 5-year trend for a country-product class
- **Validation** - Train on one 10-year slice and test on next 10-year slice.
    - Compare predictions to actual values: subtract from actual values ➜ Ranking.
- **Graph Embeddings** - Create a graph database for each country

Challenges and Next Steps

Making a Meaningful Model

- Time series analysis to improve model performance across time and predict future trends - start with simply multiplying the target vector (export percent) by the export trend vector
- Combine all predictions into one table for fast lookup + ability to compare to actual values
- De-normalize predictions to make them easier to interpret
- Improve specificity of model by resampling, changing loss measure, or otherwise

Making a Dynamic Model

- Figuring out how to use graph databases!
- Incorporating improved metrics into the model, e.g. graph alignment similarity score and/or other easily human-interpretable metrics
- Creating a tunable model and app interface (with either Atlas or additional metrics)