Conflict or Cooperation?: Predicting Future Tendency of International Relations

Yoshikawa & Ma Lab
M2 Chen Peng (陳 鵬)
26th Sep. 2019

Supervisor: Adam Jatowt
Advisor: 林 冬惠，森 信介

Introduction

Motivation

- International relations analysis is crucial to policy makers and certain groups of practitioners or scientists (e.g., journalists, economists).

- Predicting tendency helps them prepare or make decisions **in advance** for events that are likely to happen.

Introduction

- Generally, recent events between two countries directly reflect the international relation between them.

Idea

- Predicting future tendency of international relations by predicting future **conflict** / **cooperation** events.

Introduction

Data sources

- Historical data is the basis for prediction.
- Many news sources move online, making it convenient to access global **historical events**.

Related Work

Predicting future events using GDELT

- Predicting Social Unrest Using GDELT [1]. (models: Random Forest, Ada Boost, LSTM)
- Predicting conflict events in Afghanistan with RNN [2].

Prediction task under different settings

- Stock Price Prediction Using Attention-based Multi-Input LSTM [3].

[1] Divyanshi Galla and James Burke. 2018. Predicting Social Unrest Using GDELT.InInternational Conference on Machine Learning and Data Mining in PatternRecognition. Springer, 103–116.
[2] Smith, Emmanuel M., et al. ”Predicting the occurrence ofworld news events using recurrent neural networks and auto-regressive moving average models.”
[3] Li, Hao, Yanyan Shen, and Yanmin Zhu. "Stock Price Prediction Using Attention-based Multi-Input LSTM." Asian Conference on Machine Learning. 2018.

Introduction

Dataset

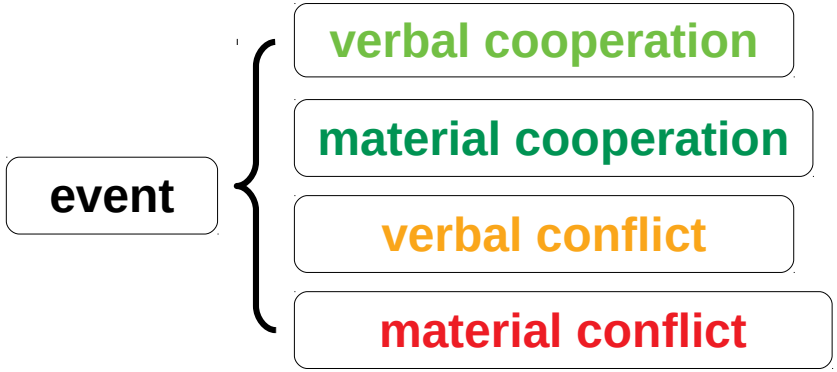
- GDELT: Global Data on Events, Location, and Tone
- GDELT is available online, providing auto-coded event records from news sources all over the world from 1979 to now.
- Total size: 222.5GB

  (up to 25th Sep. 2019)

Introduction

GDELT event table

- One line per event record
- Event Details: Date, Two Actors, Event Code, NumMentions, Quad Class ….

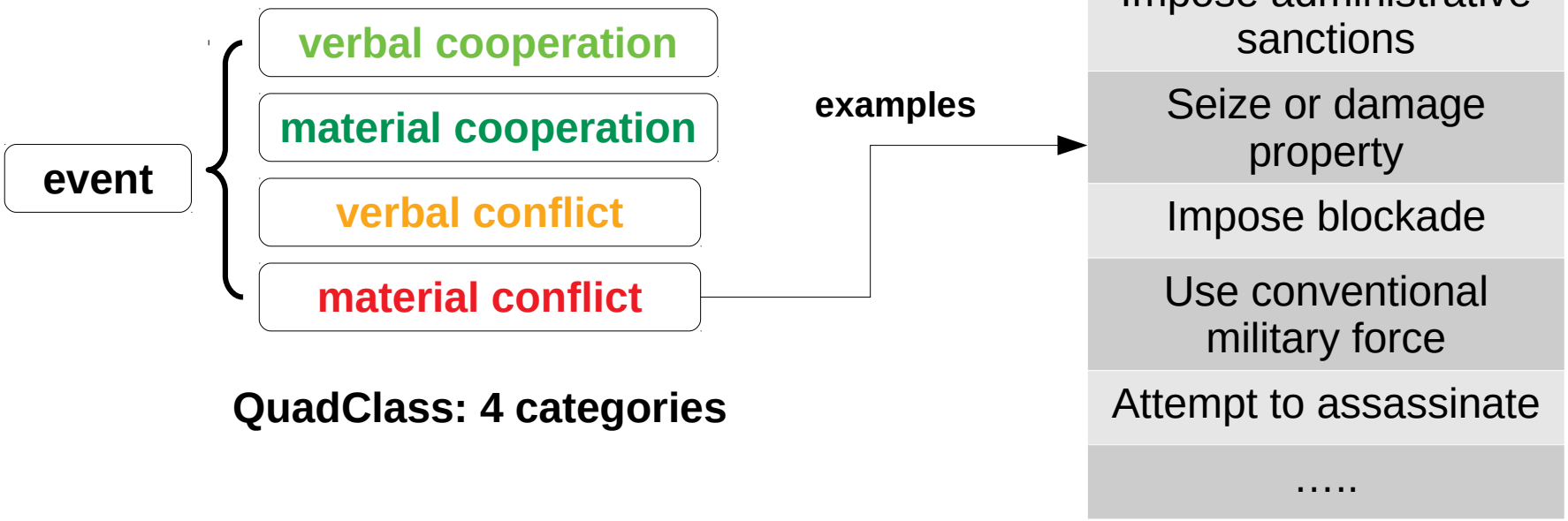


QuadClass: 4 categories

SQLDATE	Actor1Code	Actor2Code	EventCode	QuadClass	GoldsteinScale	NumMentions
20150313	USA	JPN	43	1	2.8	4
20160919	USA	JPN	20	1	3	6
20170519	USA	JPN	40	1	1	4
20170519	USAGOV	JPN	61	2	6.4	8
20180517	USA	JPN	20	1	3	1
20181231	USA	JPN	193	4	-10	7
20181231	USA	JPN	80	2	5	2

Introduction

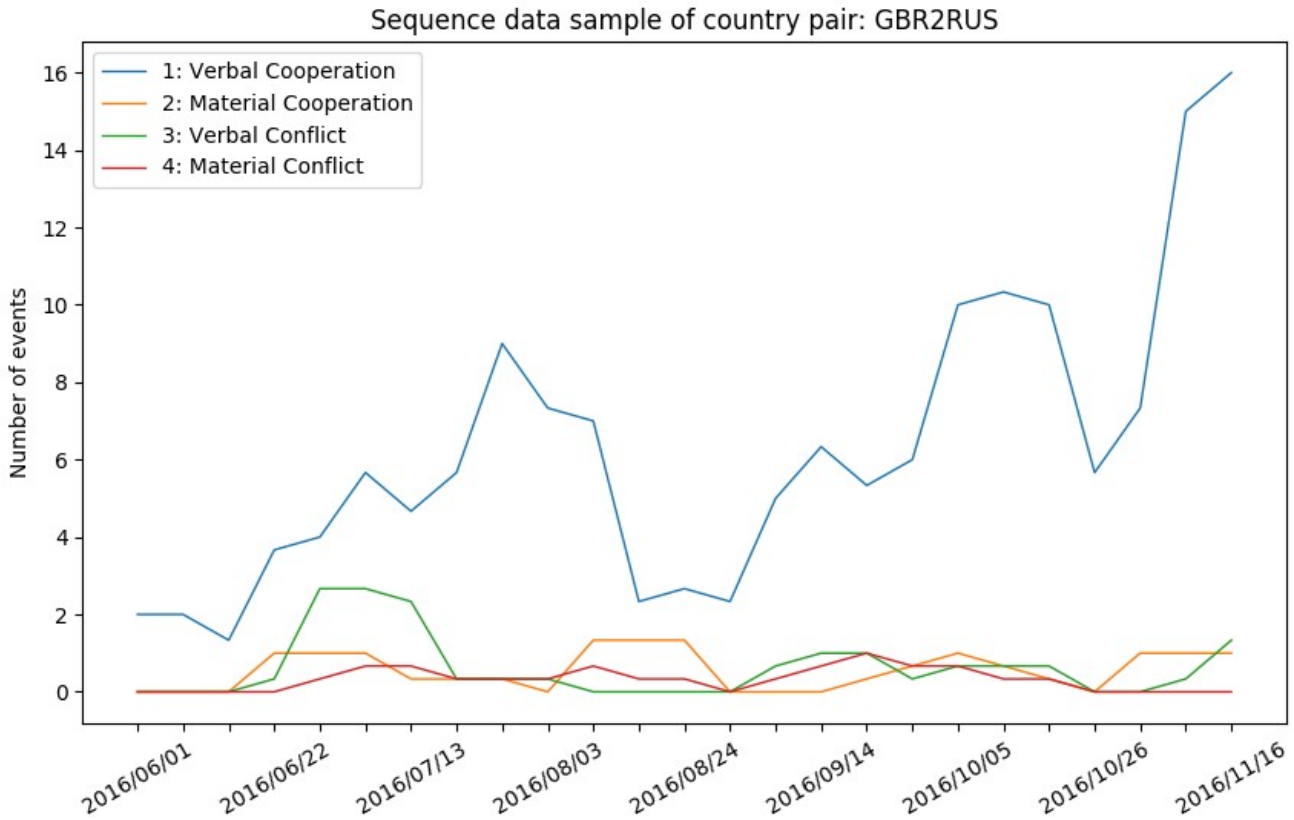
GDELT event table



QuadClass: 4 categories

Introduction

- Events happened from UK to Russia
- **Week** as a time unit, counting the number of events in each week.
- **4 categories** of events



Problem Statement

- Focus on **country pairs**.
- Use week as a time unit, counting the number of events in each week. (4 categories of events)
- Given data for consecutive M weeks, predict the number of **conflict** and **cooperation** events in the next week between two countries.

Methodology

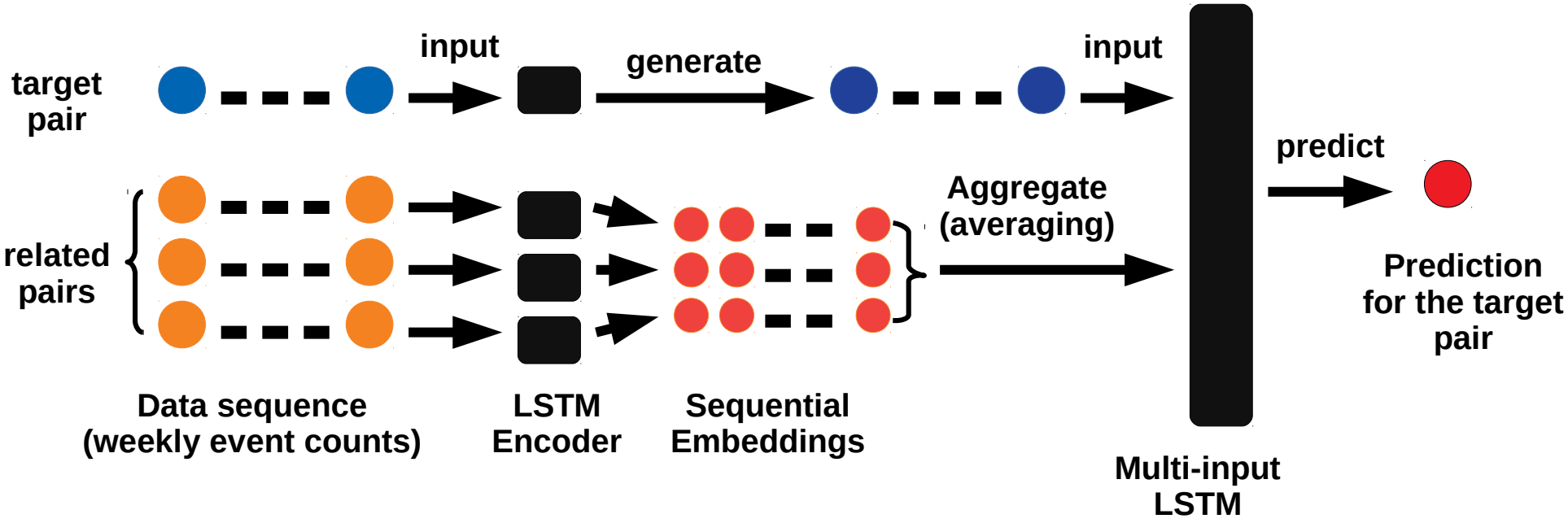
- Explore the potential usage of **related country pairs**. ***Hypothesis:*** *Relations between a specific pair of countries are likely to be affected by other country pairs.*

  For example, (RUS, USA) may be related to (CHN, USA)

- The Model we use is **Multi-input** LSTM, considering relations between different country pairs.

Methodology

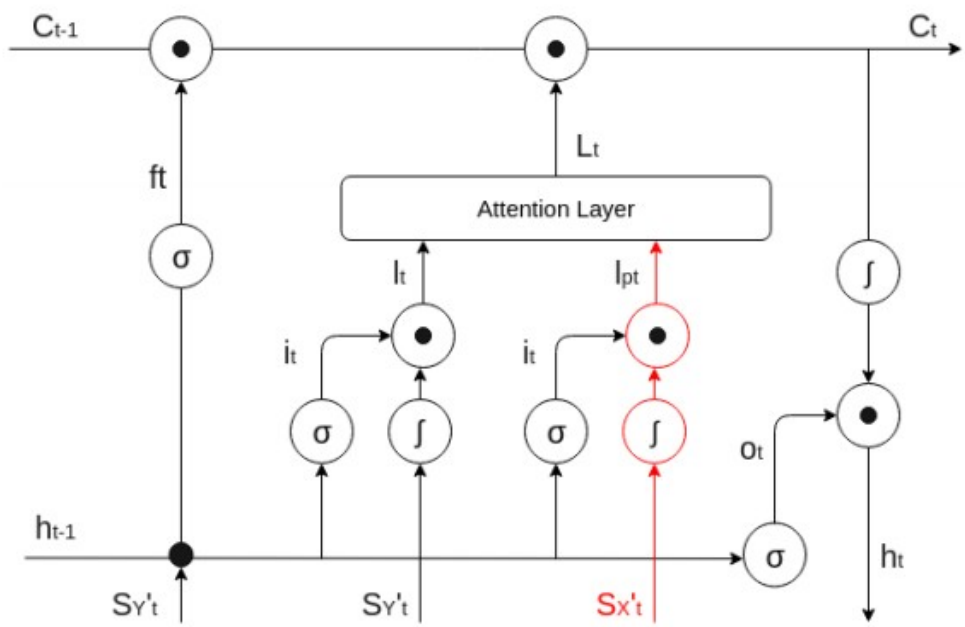
- Model: LSTM + Multi-input LSTM



Methodology

- One unit of Multi-input LSTM
- The structure is similar to vanilla LSTM but one more input and additional attention layer.
- **Intuition:**

  Use additional information extracted from related pairs to update cell states.



Methodology

Related country pairs selection

- **1. Pearson Correlation Coefficient (PCC)**

  Define the data sequence of countries Y and Xi as S_Y and S_{xi}

$$Cor(S_Y, S_{X_i}) = \frac{Cov(S_Y, S_{X_i})}{\sqrt{Var(S_Y) * Var(S_{X_i})}}$$

where Cov means covariance and Var means variance.

country pair	1st related	2nd related	3rd related
(JPN, USA)	(USA, JPN)	(USA, GBR)	(CAN, USA)
(CHN, USA)	(USA, CHN)	(GBR, USA)	(CAN, GBR)
(RUS, USA)	(USA, RUS)	(DEU, RUS	(CHN, USA)

Example of top 3 related pairs by PCC

Methodology

Related country pairs selection: Two other methods

- **2. Geographical Distance**

  Use geographical distance between **country capitals** as distance between different countries. For a target pair (JPN, USA) and another pair (CHN, USA), the correlation between them is inversely proportional to the distance between Tokyo and Beijing.

	1st related	2nd related	3rd related
(JPN, USA)	(USA, JPN)	(JPN, CAN)	(USA, KOR)
(CHN, USA)	(USA, CHN)	(CHN, CAN)	(USA, PRK)
(IND, GBR)	(GBR, IND)	(IND, FRA)	(IND, BEL)

KOR: South Korea

PRK: North Korea

GBR: United Kindom

BEL: Belgium

Methodology

Related country pairs selection

- **3. Semantic Similarity**

  Get the difference between two semantic vectors (word2vec[1]) of a country pair. For example, for (JPN, USA), the difference vectors is $\overrightarrow{JPN,USA} = \overrightarrow{USA} - \overrightarrow{JPN}$. Use the cosine similarity of two difference vectors as the correlation between two country pairs.

	1st related	2nd related	3rd related
(JPN, USA)	(USA, JPN)	(USA, CHN)	(USA, DEU)
(CHN, USA)	(USA, CHN)	(USA, IND)	(USA, JPN)

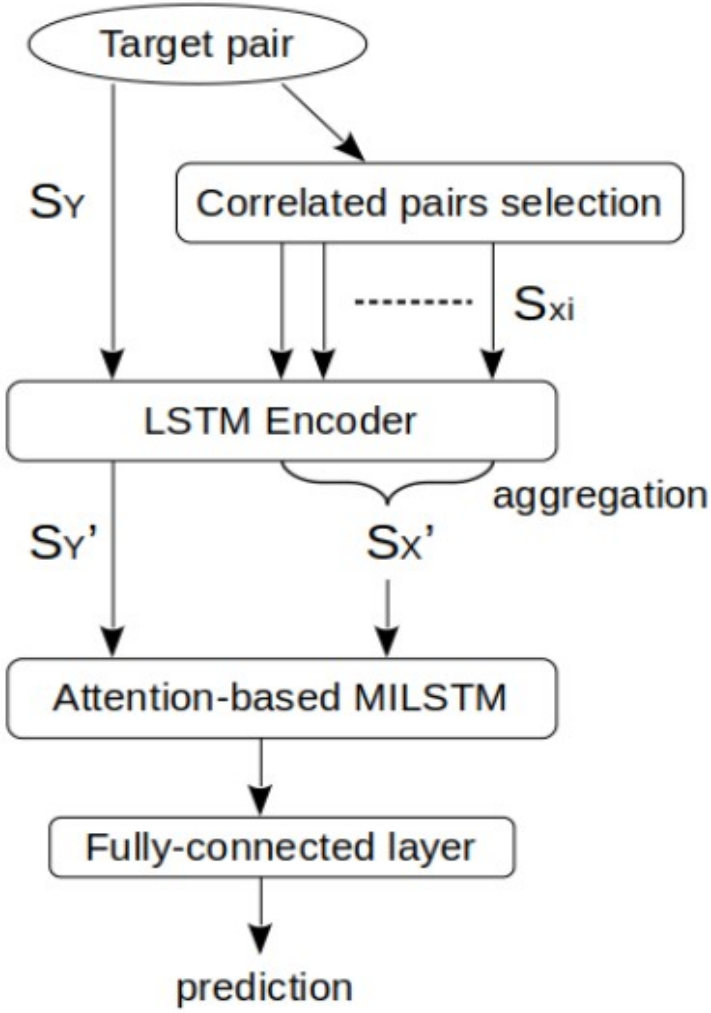
DEU: Germany

[1] Pretrained word2vec on Google News: https://code.google.com/archive/p/word2vec/

Methodology

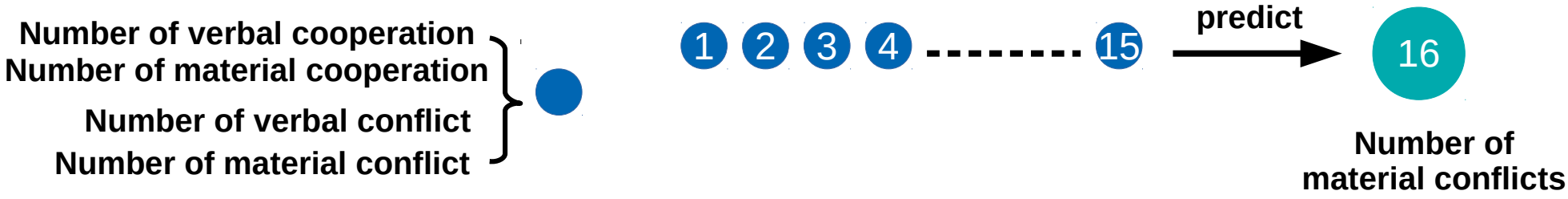
Workflow

- Given a target pair
- Related pairs selection
- LSTM Encoder
- Multi-input LSTM



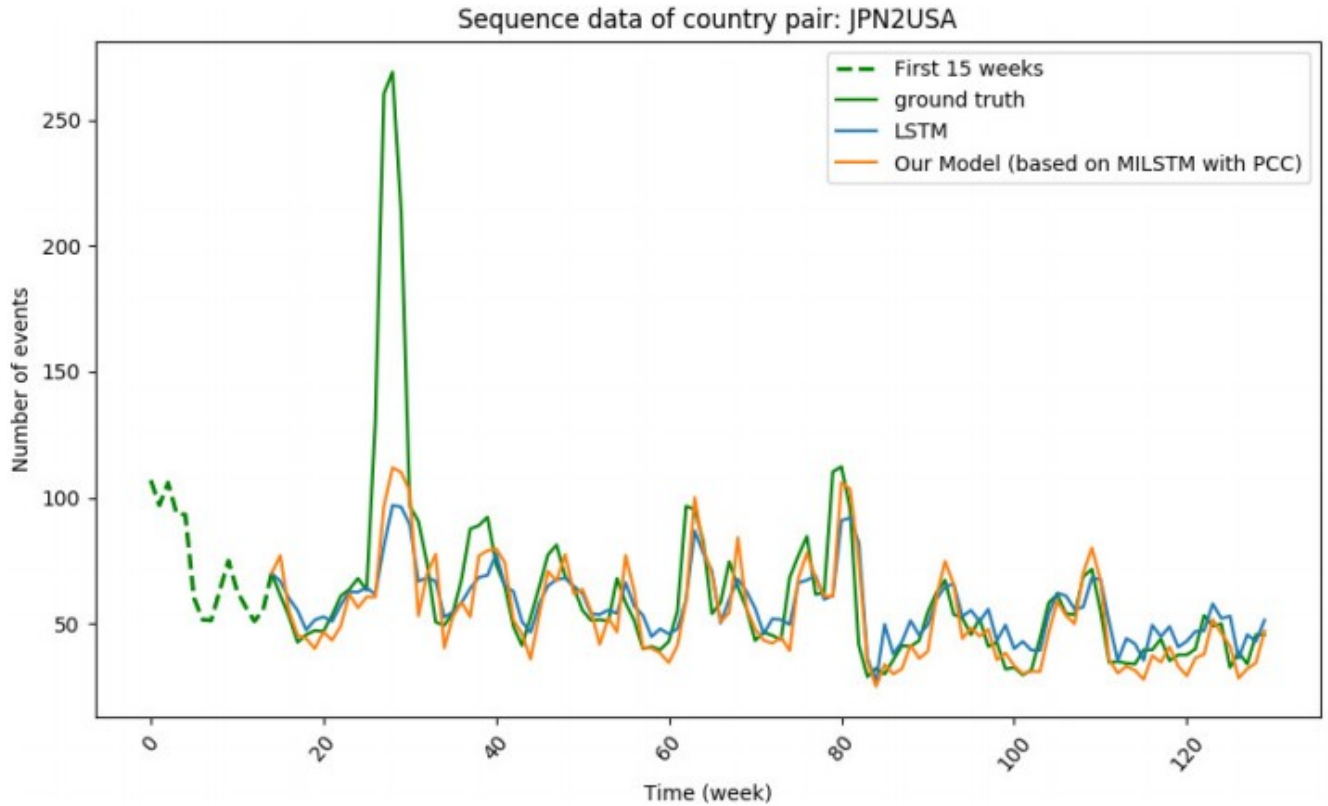
Experiment

- Settings: M = 15, thus the data of 16th week is what we want to predict. Currently, we predict the number material conflicts.
- 16 countries of interest, which form 240 country pairs
- Baseline: LSTM
- Training set: 2005/01/01 to 2016/06/01 (595 weeks)
- Test set: 2016/06/01 to 2018/12/01 (130 weeks)



Experiment

- Example of prediction of country pair (JPN, USA) on the whole test set.
- Prediction of the number of material conflict events
- Compare our model and LSTM



Result

- Use MSE (Mean Squared Error) Loss, evaluated on the whole test set.
- Training epochs: 25
- The average improvement over LSTM:

  MILSTM with PCC: **9.06%**

  MILSTM with dist. : **9.50%**

  MILSTM with w2v : **7.02%**

	LSTM	MILSTM with PCC	MILSTM with distance	MILST with word2vec
(JPN, USA)	945.0	882.1	899.3	**873.8**
(USA, JPN)	226.3	191.0	**186.8**	212.8
(JPN, CHN)	27.6	29.5	**26.9**	33.3
(CHN, JPN)	**25.2**	25.5	27.2	27.0
(USA, CHN)	290.6	272.3	248.7	**241.2**
(USA, CAN)	1377.1	995.3	**939.1**	963.2
(CAN, USA)	437.6	343.5	346.6	**337.7**
(GBR, JPN)	499.2	**425.9**	461.7	449.8
(FRA, GBR)	192.4	179.5	**169.0**	174.3
(GBR, FRA)	127.6	123.7	**114.7**	130.1
(USA, DEU)	71.1	69.1	76.2	**68.8**
(GBR, JPN)	16.1	**14.2**	14.9	14.8

Future Work

- Run the model on the whole set of country pairs. (in progress)
- Predict more categories, e.g. material cooperation.
- Explore other kinds of models, e.g. GCN (graph convolutional network), etc.

References

[1] Qiao, Fengcai, et al. ”Predicting social unrest events withhidden Markov models using GDELT.”

[2] Smith, Emmanuel M., et al. ”Predicting the occurrence ofworld news events using recurrent neural networks and auto-regressive moving average models.”

[3] Li, Hao, Yanyan Shen, and Yanmin Zhu. "Stock Price Prediction Using Attention-based Multi-Input LSTM." Asian Conference on Machine Learning. 2018.

[4] Divyanshi Galla and James Burke. 2018. Predicting Social Unrest Using GDELT.InInternational Conference on Machine Learning and Data Mining in PatternRecognition. Springer, 103–116.

Thank you

Questions and comments please!