# Predicting Earthquakes with Hierarchical Neural Network Models

Matheus Junqueira[a], Yoshito Hirata

[a] Degree Programs in Systems and Information Engineering
Graduate School of Science and Technology
University of Tsukuba – Japan

October 2024

スライドの日本語版!
tinyurl.com/mrx3b28y

理工情報生命学術院
システム情報工学研究群
Degree Programs in Systems and Information Engineering
Graduate School of Science and Technology

# Introduction

https://tinyurl.com/mrx3b28y

- Applying machine learning in earthquake forecasting faces many obstacles
- The major one being the amount and quality of data
  - Missing data near the lower detection threshold
  - Catalogs don’t reach far enough back in time
  - Changes in data collection methods over time

# Introduction

https://tinyurl.com/mrx3b28y

- Applying machine learning in earthquake forecasting faces many obstacles
- The major one being the amount and quality of data
  - Missing data near the lower detection threshold
  - Catalogs don’t reach far enough back in time
  - Changes in data collection methods over time
- We try to circumvent this problem by:
  - extracting as much information as possible from the catalogs
  - devising a prediction model that can process that information

# Introduction

- We limit our analysis to fault-independent information
  - ↪ Catalogs with only basic earthquake characteristics: timestamp, magnitude, depth, latitude, longitude
  - ↪ Widely available for various seismic regions around the globe
  - ↪ $\Rightarrow$ More data to work with
  - ↪ $\Rightarrow$ Applicability to various regions

# Introduction

- We limit our analysis to fault-independent information
  - Catalogs with only basic earthquake characteristics: timestamp, magnitude, depth, latitude, longitude
  - Widely available for various seismic regions around the globe
  - $\Rightarrow$ More data to work with
  - $\Rightarrow$ Applicability to various regions
- Existing methods do not seem to leverage all the information provided by these datasets
- Improvements can be achieved in:
  - modeling of the problem
  - feature engineering

# Feature Engineering

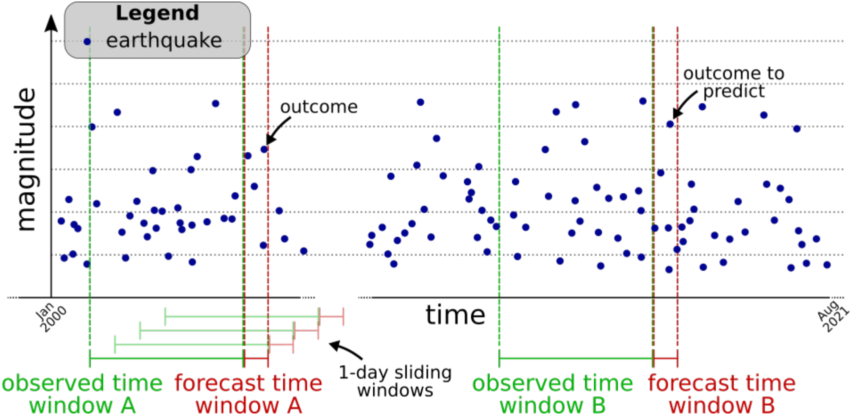
- Based on the earthquake catalog information, we derive two kinds of features:
  - Distances between earthquake point patterns
  - Seismicity indicators
- Our prediction model, however, can be extended to other features

# Edit distances

- We pre-process the earthquake catalog to obtain the sets of earthquakes in time windows of a certain length
  - 7 days for next-day forecasting
- Each set is a series of points scattered over a tridimensional space, with extra information associated to each point:
  - magnitude
  - timestamp
- These sets can be compared by using statistical tools
  - We use **edit distances**
- Main idea is that similar earthquake patterns will result in similar outcomes[1]

[1] M. H. Junqueira Saldanha and Y. Hirata (2022). "Solar activity facilitates daily forecasts of large earthquakes". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32.6, p. 061107.

# Feature Engineering



# Edit distances

- Let $P_1$ and $P_2$ be two marked point processes:

$$P_1 = \{(t_i, \boldsymbol{u}_i) \mid 1 \leq i \leq N_1\} \qquad P_2 = \{(s_j, \boldsymbol{v}_j) \mid 1 \leq j \leq N_2\}$$

- The idea is to transform $P_1$ into $P_2$ using primitive operations, each incurring a cost.
- The primitive operations are:
  - **Insertion**: insert point $(s_j, \boldsymbol{v}_j)$ from $P_2$ into $P_1$, paying a cost of $1$;
  - **Deletion**: remove point $(t_i, \boldsymbol{u}_i)$ from $P_1$, paying a cost of $1$;
  - **Shifting**: replace point $(t_i, \boldsymbol{u}_i)$ in $P_1$ by the point $(s_j, \boldsymbol{v}_j)$ from $P_2$, paying a cost that depends on how different these two points are.
- The **edit distance**[2][3] is defined as the lowest possible cost necessary to transform $P_1$ to $P_2$.

[2] J. D. Victor and K. P. Purpura (1997). "Metric-space analysis of spike trains: theory, algorithms and application". In: *Network: computation in neural systems* 8.2, pp. 127–164.

[3] S. Suzuki, Y. Hirata, and K. Aihara (2010). "Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies". In: *International Journal of Bifurcation and Chaos* 20.11, pp. 3699–3708.

# Seismicity indicators

- Now consider the sets of earthquakes over time-windows of size $T$
  - $T = 7, 15, 30, 60, 90, 180, 360$ days
- Seismicity indicators can be calculated for each of these sets
  - mean and maximum magnitudes;
  - rate of release of seismic energy, given by

$$\frac{\sum_{i=1}^{k}\left(10^{10.8+1.5e_i^{\mathrm{mag}}}\right)}{\Delta t},$$

    where $\Delta t$ is the length of the time window;
  - time elapsed between all earthquakes that exceed a certain magnitude threshold $\tau$;

# Seismicity indicators

- ↪ $a$, $b$ values of the Gutenberg–Richter's law fit;
  ↪ sum of square errors of the magnitudes to the regression line of the GR law;
  ↪ magnitude deficit, defined as the difference between the expected maximum magnitude (according to the GR law) and the observed maximum magnitude; and
  ↪ coefficient of variation of inter-event times, after removing earthquakes with magnitude below a certain threshold $\tau$.

# Seismicity indicators

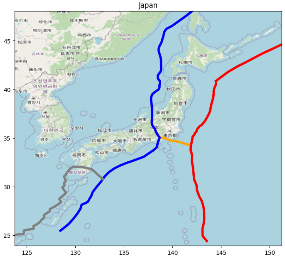
- ↪ $a$, $b$ values of the Gutenberg–Richter's law fit;

  ↪ sum of square errors of the magnitudes to the regression line of the GR law;

  ↪ magnitude deficit, defined as the difference between the expected maximum magnitude (according to the GR law) and the observed maximum magnitude; and

  ↪ coefficient of variation of inter-event times, after removing earthquakes with magnitude below a certain threshold $\tau$.

- Distance to each fault line (undergoing improvements)



# Forecasting Framework

- If we calculate these features for different sizes of time-windows, we are left with a large volume of features to work with
- However, more features $\Rightarrow$ bigger models $\Rightarrow$ more data needed for good fit (Bengio, Goodfellow, and Courville 2017)[4]

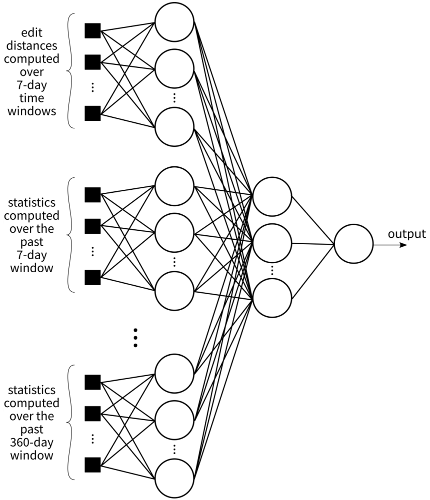
[4] Y. Bengio, I. Goodfellow, and A. Courville (2017). *Deep learning*. Vol. 1. MIT press Cambridge, MA, USA.

# Forecasting Framework

- If we calculate these features for different sizes of time-windows, we are left with a large volume of features to work with
- However, more features $\Rightarrow$ bigger models $\Rightarrow$ more data needed for good fit (Bengio, Goodfellow, and Courville 2017)[4]
- To tackle this, we design a neural network architecture particular for the problem at hand
- Objectives:
  - Features are mathematically processed in a logical manner
  - Minimal number of trainable weights
  - Extensibility of the model to use different kinds of features

[4] Y. Bengio, I. Goodfellow, and A. Courville (2017). *Deep learning*. Vol. 1. MIT press Cambridge, MA, USA.

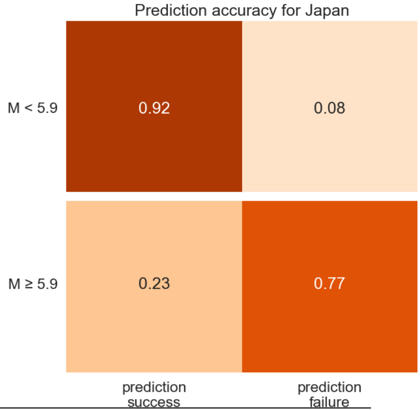
# Forecasting Framework




- The features associated with time-window $T_i$ are fed to one single neural sub-network
  - ↪ There is no need to evaluate linear-combinations of features of different time window lengths
  - ↪ Nor linear combinations between seismicity indicators and edit distances
- The outputs of the sub-network are then fed to a neural network that will then derive the predictions

# Results

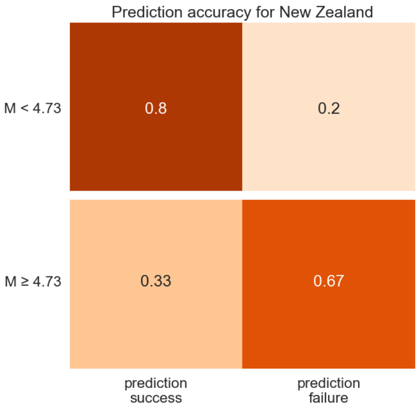
- We apply our models on Japan, New Zealand and Balkan Catalogs[5]
- Binary classification using the 99% quantile of magnitudes as threshold




- Overall accuracy: 91.31%
- Class weighted accuracy: 57.50%

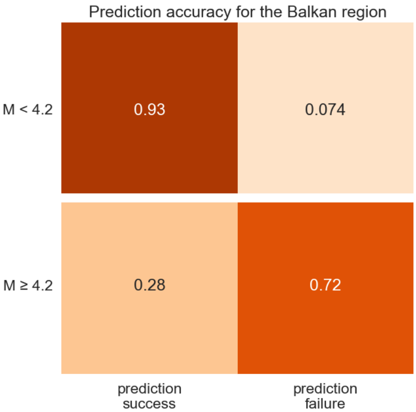
[5] Catalogs from the Japanese Meteorological Agency, the New Zealand GeoNet project and the University of Athens. Minimum magnitude threshold: 2.5

# Results




- Overall accuracy: 79.53%
- Class weighted accuracy: 56.50%

# Results




- Overall accuracy: 92.35%
- Class weighted accuracy: 60.50%

# Conclusions

- We found that earthquake catalogs without fault-dependent information can still provide vast amounts of features
- Processing becomes difficult as the number of features increases

# Conclusions

- We found that earthquake catalogs without fault-dependent information can still provide vast amounts of features
- Processing becomes difficult as the number of features increases
- We present a neural network architecture that
  - Processes features in a logical manner
  - Minimizes number of trainable weights
  - Is flexible and easy to extend for different features and use in different regions

# Conclusions

- Results so far are promising, bringing improvements to our past results (Junqueira Saldanha and Hirata 2022)[6]
- Possible improvements relative to recent work: (Yavas et al. 2024)[7]
  - 30-day forecasting
  - Overall accuracy: 55.97%
  - Large magnitudes accuracy: 11.9% – 27.0% (estimated)

[6] M. H. Junqueira Saldanha and Y. Hirata (2022). "Solar activity facilitates daily forecasts of large earthquakes". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32.6, p. 061107.

[7] C. E. Yavas et al. (2024). "Predictive modeling of earthquakes in los angeles with machine learning and neural networks". In: *IEEE Access*.

Thank You!

# References I


- Bengio, Y., I. Goodfellow, and A. Courville (2017). *Deep learning*. Vol. 1. MIT press Cambridge, MA, USA.
- Junqueira Saldanha, M. H. and Y. Hirata (2022). "Solar activity facilitates daily forecasts of large earthquakes". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32.6, p. 061107.
- Suzuki, S., Y. Hirata, and K. Aihara (2010). "Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies". In: *International Journal of Bifurcation and Chaos* 20.11, pp. 3699–3708.
- Victor, J. D. and K. P. Purpura (1997). "Metric-space analysis of spike trains: theory, algorithms and application". In: *Network: computation in neural systems* 8.2, pp. 127–164.
- Yavas, C. E. et al. (2024). "Predictive modeling of earthquakes in los angeles with machine learning and neural networks". In: *IEEE Access*.