Prediction of Geostrategic Events: International Conference on Machine Learning (ICML 2024)

Anonymous Authors[1]

Abstract

The problem of the prediction of historical events is with the meteorological prediction one of the most challenging one. We take the opportunity of the availaibility of a very rich dataset covering the last 50 years and the recent emergence of very recent geopolitical storm, to try to calibrate a model based on all the recent emergence of very powerful methods to attack prediction problems. We enlighten the methodology with information geometry tools and ground the calibration of our model with a schrodinger bridge based on a diffusion process. Willing to take into account the interaction with people using the model, we apply Hamiltonian representation of the method, that after a necessary quantification make place for the external interaction, we show the emergence of a new incertitude relationship concerning the information created by the model.

1. Introduction

1.1. Introductory reflexions

We want to create a concept of an incertitude relationship in the context of a perfect predictor of future events. We call this concept historical consistency relationship. It taps into several deep philosophical and theoretical physics considerations.

2. General Structure of the Predictive Chain

In order to Separate the difficulties, the prediction is split in two parts. One part generates a set of possible futures, and the second part analyse the set of possible futures to aswer questions that constitute the effective predictions of the system.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

%

% ici insertion d'un shema troi boites: generateur of futures events / futur events / measures : predictions

%

3. Structure of the generator of futur events

3.1. Clusturing of events

In order to take into account the path dependencies if events two meccanisms are implemented.

1. First the events are groupped into clusters that are associated with different time scales .

2. Secondly the different scales of events interact effectivelly during the forward pass and during the bacward pass. The interaction in designed to be modular, allowing to change the set of scales managed par the system generic, allowing the upgrade the type of interactions between the different scales

3. The historical sequences of clusters are overlapping as soon as their associated period is more than one day.

%

% ici insertion d'un shema N lignes , chaque lignes est une echelle de clusturing , avec comme exemple: 1jour, 3 jours, une semaine, 1 mois

%

3.2. Encoder / Decoder

3.3. Latent space diffusion

An essential caracteristics of the the event prediction is the the forward pass goes from the present to the past a contrario to the usual time. The past which is the most distant is the most blurred and at very long distance, it is completly random. Reversely, the backward pass , whic is the generative pass, acts like a deblurring machine, building the details of the events, strats afetr strats.

3.3.1. Forward pass

The diffusion process can be mathematically described using stochastic differential equations (SDEs). A basic SDE for a diffusion process is given by:

$$dx = \mu(x, t)\, dt + \sigma(x, t)\, dW_t \tag{1}$$

Where:

- x represents the state of the system (e.g., the data point in latent space).
- $\mu(x, t)$ is the drift coefficient, representing the deterministic part of the motion.
- $\sigma(x, t)$ is the diffusion coefficient, representing the stochastic or random part.
- dW_t is a Wiener process (or Brownian motion), representing the random fluctuation over an infinitesimal time interval dt.

3.3.2. Backward pass

The reverse diffusion process aims to reconstruct or predict the futur events from the past. This can be modeled as a reverse SDE:

$$dx = \left[\mu(x, t) - \sigma^2(x, t)\nabla_x \log p_t(x)\right] dt + \sigma(x, t)\, dW_t \tag{2}$$

4. Reservoir of formulas

4.0.1. Mathematical Formulation

The relationship $\Delta I_p \Delta I_o \leq L$ is intriguing. In mathematical terms, this inequality suggests a sort of conservation law for uncertainty and error. It's akin to the trade-offs seen in optimization problems, where improving one aspect leads to compromises in another. The constant L could be interpreted as the inherent limit of predictive accuracy in a system influenced by human knowledge and actions.

5. The latent space

$p_\theta\left(x_{t-1} | x_t\right) = \mathscr{N}\left(\mathrm{x_{t-1}}; \mu_\theta\left(\mathrm{x_t}, \mathrm{t}\right), \sigma_\mathrm{t}^2 \mathbf{I}\right)$

6. Diffusion Process for the the latent space

The diffusion process can be mathematically described using stochastic differential equations (SDEs). A basic SDE for a diffusion process is given by:

$$dx = \mu(x, t)\, dt + \sigma(x, t)\, dW_t \tag{3}$$

Where:

- x represents the state of the system (e.g., the data point in latent space).
- $\mu(x, t)$ is the drift coefficient, representing the deterministic part of the motion.
- $\sigma(x, t)$ is the diffusion coefficient, representing the stochastic or random part.
- dW_t is a Wiener process (or Brownian motion), representing the random fluctuation over an infinitesimal time interval dt.

6.1. Latent Space Representation

For a time series Y_t, we can transform it into a latent space representation Z_t using a function f:

$$Z_t = f(Y_t) \tag{4}$$

The function f could be a neural network trained to encode Y_t into a lower-dimensional space.

6.2. Reverse Diffusion (Predictive Model)

The reverse diffusion process aims to reconstruct or predict the original data from the noisy state. This can be modeled as a reverse SDE:

$$dx = \left[\mu(x, t) - \sigma^2(x, t)\nabla_x \log p_t(x)\right] dt + \sigma(x, t)\, dW_t \tag{5}$$

Here, $\nabla_x \log p_t(x)$ is the gradient of the log probability of x at time t, which is crucial for 'denoising'.

6.3. Training Objective (Loss Function)

The training of a diffusion model involves minimizing a loss function that typically measures the difference between the original data and its reconstruction. A common loss function is the Mean Squared Error (MSE):

$$L = \frac{1}{N} \sum_{i=1}^{N} \|Y_i - \hat{Y}_i\|^2 \tag{6}$$

Where:

- Y_i is the original time series data.
- $\hat{Y}_i$ is the predicted data from the model.
- N is the number of data points.

6.4. Time Series Prediction

For predicting future values $\hat{Y}_{t+1}$ of a time series Y_t, the model uses the learned reverse diffusion process:

$$\hat{Y}_{t+1} = g^{-1}(\text{ReverseDiffusion}(Z_t)) \tag{7}$$

Where:

- g^{-1} is the decoding function, inverse of f, transforming latent representations back to the original data space.

7. The diffusion process

$p_\theta\left(x_{t-1}|x_t\right) = \mathcal{N}\left(\mathrm{x_{t-1}}; \mu_\theta\left(\mathrm{x_t}, \mathrm{t}\right), \sigma_{\mathrm{t}}^2\mathbf{I}\right)$

7.1. Forward Pass: Adding Noise

The forward pass of a diffusion process can be described by a series of transformations that gradually add noise to the data. If x_0 represents the original data, the noisy data at step t can be represented as x_t. The process is often modeled as a Markov chain, where each step is only dependent on the previous step.

The equation for the forward process at step t can be given by:

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon_t$$

where:

- x_t is the data at step t.
- α_t is a coefficient that determines the amount of noise added at step t (typically $0 < \alpha_t \leq 1$).
- ϵ_t is a noise term, often assumed to be Gaussian, $\epsilon_t \sim \mathcal{N}(0, I)$.

7.2. Backward Pass: Denoising

The backward pass in a diffusion process aims to reverse the noise addition of the forward pass to reconstruct or generate new data. This involves learning a denoising function that can predict the original data from the noisy data.

The denoising step can be modeled as:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}}\epsilon_\theta(x_t, t)\right)$$

where:

- $\epsilon_\theta(x_t, t)$ is the predicted noise at step t by the model with parameters θ.
- Other terms are as defined in the forward pass.

7.3. Algorithms

Algorithm 1 shows an example.

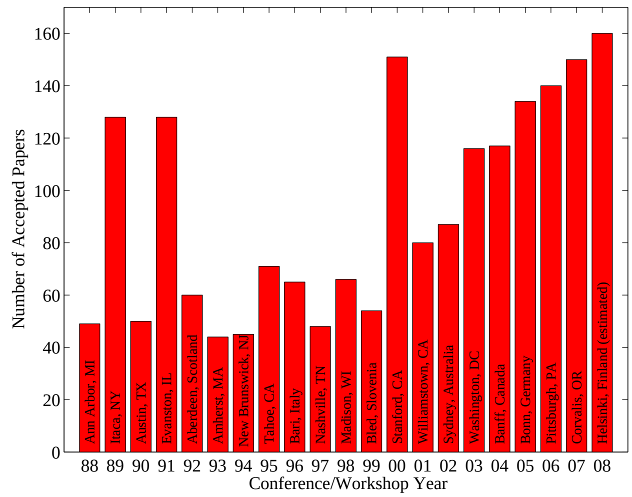

Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

Algorithm 1 Bubble Sort

Input: data x_i, size m
repeat
  Initialize $noChange = true.$
  for $i = 1$ **to** $m - 1$ **do**
    if $x_i > x_{i+1}$ **then**
      Swap x_i and x_{i+1}
      $noChange = false$
    end if
  end for
until $noChange$ is $true$

7.4. Tables

References

Author, N. N. Suppressed for anonymity, 2021.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.).

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	$\surd$
CLEVELAND	83.3± 0.6	80.0± 0.6	$\times$
GLASS2	61.9± 1.4	83.8± 0.7	$\surd$
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	$\times$
META	67.1± 0.6	76.5± 0.5	$\surd$
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	$\surd$

Machine Learning: An Artificial Intelligence Approach, Vol. I. Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.

A. Appendix

A.1. Asimov psychohistoiry

Isaac Asimov's concept of "psychohistory" in his "Foundation" series is a remarkable parallel to the idea described above. Psychohistory, in Asimov's universe, is a fictional scientific discipline that combines history, sociology, and mathematical statistics to make general predictions about the future behavior of very large groups of people, such as the Galactic Empire.

A.1.1. Key Elements of Asimov's psychohistory:

- Predictive Power on a Large Scale:

Psychohistory is effective in predicting the broad flow of future events in large populations. It doesn't predict individual actions but can forecast the aggregate behavior of billions.

- Need for Secrecy:

In the "Foundation" series, it's crucial that the existence and predictions of psychohistory remain secret. If the public were aware of these predictions, they could change their behavior in response, thus invalidating the predictions — very similar to the feedback loop described in the model.

- Mathematical Underpinnings: While Asimov didn't provide detailed mathematical formulas for psychohistory, he suggested that it's based on complex mathematical models. This aligns with the idea of defining public uncertainty (ΔI_p) and oracle error (ΔI_o) in mathematical terms.

A.1.2. Reflections on psychohistory and the concept of history consistency:

- Complexity and Chaos:

Both psychohistory and this concept recognize that even if one could predict the future to some extent, the inherent complexity and chaotic nature of human behavior might limit the accuracy of these predictions, especially when the subjects of prediction become aware of it.

- Narrative as a Tool for Exploration:

Asimov used psychohistory as a narrative tool to explore themes of destiny, free will, and the role of knowledge in shaping the future. Similarly, the concept of historical consistency , though hypothetical, serves as a thought-provoking tool to explore the dynamics of prediction, knowledge, and their impacts on human behavior.

A.1.3. Ethical and Philosophical Implications:

The ethical and philosophical implications of psychohistory are important. The knowledge of future events, especially on a large scale, brings up questions about determinism, free will, and the ethical use of such knowledge.

- Determinism vs. Free Will:

The concept of psychohistory challenges the notion of free will. If the future can be predicted accurately, it implies a level of determinism in human behavior. This raises profound questions about individual agency and responsibility. Are we merely following predetermined paths, or do our choices matter?

- The Ethics of Foreknowledge:

The moral implications of knowing the future are significant. If one can predict disasters or societal upheavals, what is the ethical responsibility to intervene or warn others? Conversely, could such foreknowledge lead to fatalism or apathy, believing that certain outcomes are inevitable?

- Impact on Society:

Knowledge of future events could dramatically alter societal behavior. It might lead to attempts to change the course of predicted events, potentially causing new, unforeseen outcomes. This interplay between knowledge and action is a central theme in discussions about the ethics of prediction.

A.1.4. Real-World Predictive Models

While psychohistory is fictional, it does bear resemblance to certain real-world disciplines. For example, in economics and social sciences, models are used to predict broad trends, but these predictions can be thrown off by unforeseen factors, including the subjects' knowledge of the predictions.

- Complexity and Unpredictability:

Economic and social models often struggle with the complexity and unpredictability of human behavior. The Butterfly Effect in chaos theory – where small changes can lead to vastly different outcomes – is a key challenge in these models. This unpredictability is compounded when individuals react to predictions, potentially altering the outcome.

- Limitations and Assumptions:

Predictive models are based on assumptions that may not hold in all circumstances. The limitations of these models become apparent during unprecedented events (like a global pandemic), where the models may fail to accurately capture new dynamics and human responses.

- Continuous Evolution:

Predictive modeling is an evolving field. As we gather more data and develop better algorithms, our models become more sophisticated. However, the inherent uncertainty in human behavior and the influence of knowledge on future events remain significant challenges.

- Philosophical conclusion

In summary, the exploration of predictive models, whether in science fiction like Asimov's "Foundation" or in real-world applications, reveals much about our understanding of the universe and ourselves. It raises critical questions about the nature of knowledge, the ethics of prediction, and the balance between determinism and free will.onclusion

A.2. The GDELT Project

A.2.1. Definition and Scope

- Data Repository:

The GDELT Project is a large-scale repository of global events, media, and socio-economic data. Its comprehensiveness makes it an ideal dataset for attempting large-scale predictions about global trends and events.

- Real-Time Analysis:

GDELT's real-time data collection could allow for continuous updating of predictive models, making them more responsive to emerging trends and changes.

Global Database of Events, Language, and Tone (GDELT): A large-scale, open-source database that monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100 languages. Data Collection: It processes billions of real-time and historical news reports worldwide.

- 2. Core Components

Events Database: Contains geocoded entries of events, capturing what happened, where, and who was involved. Global Knowledge Graph (GKG): Maps people, places, organizations, themes, sources, emotions, counts, quotes, images, and events from the news into a massive interconnected network. Visual Knowledge Graph (VGKG): Extracts visual imagery from news media to create a searchable index of the visual world.

- 3. Data Features

Temporal Coverage: Ranging from near-real-time data to historical archives. Geographic Detail: High-resolution geographic information, allowing for localized analysis. Thematic Breadth: Covers a wide range of themes, from political and economic to social and environmental.

A.2.2. Applications already develloped in Geostrategic Prediction

- Conflict and Crisis Forecasting

Identifying Hotspots: Analyzing trends and patterns in conflict-related data to anticipate geopolitical hotspots. Crisis Management: Providing real-time data for humanitarian and disaster response.

- Political Analysis

Election Monitoring: Tracking political discourse and sentiment around elections. Policy Impact Assessment: Evaluating the global reaction to policy changes or major political events.

- Economic and Business Insights

Market Intelligence: Assessing global market trends and their geopolitical impacts. Risk Assessment: Identifying economic risks and opportunities in different regions.

- Social and Cultural Trends

Public Opinion and Sentiment Analysis: Gauging public sentiment on various issues across different regions. Cultural Impact Studies: Understanding the global influence of cultural events and movements.

- Environmental and Health Monitoring

Climate Change Analysis: Tracking news related to climate change and environmental policies. Public Health Surveillance: Monitoring global health trends, including disease outbreaks and public health responses.

A.3. Challenges and Considerations

- Complexity and Computation:

The sheer complexity of global socio-economic systems and the computational power required to process and analyze this data are significant challenges. It requires not only advanced algorithms but also immense computational resources.

- Data Quality and Bias:

The accuracy of predictions is heavily dependent on the quality of the data. Biases in data collection and reporting can skew predictions, leading to misleading results.

- Ethical and Privacy Concerns:

Using such data for predictive modeling raises ethical questions, especially regarding privacy and the potential misuse of predictive information.

- Interdisciplinary Approach:

This endeavor would require collaboration across multiple disciplines – computer science, physics, economics, sociology, and more – to effectively model and interpret the data.