Climate tipping points: Detection and analysis of patterns using an ordinal regression approach

Type of activity: Standard Study

1. Background and Study Motivation

Many real-world learning applications in a wide range of domains, such as finance, aerospace, entertainment, or climate modelling involve time series data [7, 10]. A time series is defined as an ordered set of real-valued variables that are sampled or extracted on a continuous signal. Due to its sequential nature, variables that are close in a time series are often highly correlated. This is especially true in climate systems, as these variables are subject to strong internal feedback effects.

Generally, small changes in a certain parameter of a time series can cause equilibria to appear or disappear. These points are called tipping points [1]. Therefore, a climate tipping point is a point where a climate state changes from one stable state to another stable state. The detection of tipping points in climate systems is currently a vividly discussed topic in the scientific community [2].

According to the last report of the International Panel on Climate Change [8], the current global warming is very likely due the observed increase in anthropogenic greenhouse gas concentrations. Identifying the proximity to tipping points would be valuable for policy makers as it would provide early warnings of irreversible climate change.

Motivation

This study will attempt to use ordinal regression models to find and characterize patterns preceding tipping points. Via the application of the proposed models, the study aims at:

- The detection of tipping points in climate system data series.
- Gaining knowledge about the typical behavior of climate systems before the occurrence of a tipping point. Such insight would allow detecting warning signs for upcoming climate tipping points.

The application of time series analysis and ordinal regression approaches to predicting future climate tipping points is of interest because current global climate models cannot effectively predict climate tipping points. These models are large deterministic dynamical systems, which typically do not incorporate key small-scale processes and microscopic nonlinearities. However abrupt transitions can occur which would not necessarily have been predicted by the model because of the strong internal feedback effects occurring close to a tipping point. In addition the climate system is highly stochastic and therefore computer models do not perform well in short-term evaluations of the climate response. Studying tipping points using tools from time series analysis could help to identify such key small-scale processes and thus to improve our understanding of Earth's climate.

In case such a method works, it could be applicable to a large range of time series including economics.

Assumptions

The main assumption of the present study is that tipping points are preceded by typical patterns in time series in climate models. This assumption is based on research published by (add ref here of the Science papers). Therefore, the study proposes to use as input variables of the model the lagged values of the final dependent variable. This research project will analyse this assumption in depth considering machine learning models.

2. Study objectives

The main objective of this study is to assess whether the proposed ordinal regression model can be applied to detect characteristic patterns preceding tipping points in complex systems, and to apply such a method to climate data.

In order to achieve this objective, the study will need to:

- Propose new models for ordinal regression problems.
- Develop new methodologies for unbalanced classification problems in the ordinal regression context.
- Propose a new way to detect tipping points (using an ordinal regression approach).
- Propose new models to discretize a times series in order to transform it into an ordinal regression problem.
- analyse the found patterns and cross check them with physical insight

3. Proposed Methodology

The following study logic is proposed by the ACT, though universities and research centres are encouraged to propose a different approach to achieve the above described study objectives.

1. **Discretization stage**
The proposed approach is to discretise the dependent variable into different levels (classes) in order to treat it as a classification problem. This would allow to classify them into five regions:

- C5: The dependent variable increases sharply in the next period of time considered.
- C4: The dependent variable increases slightly in the next time step.
- C3: The time series is stable in the period of time t + 1.
- C2: The dependent variable decreases slightly in the next period of time.
- C1: In t + 1, the dependent variable decreases sharply.

2. **Ordinal regression stage**
Ordinal classification plays an important role in various machine learning tasks. In these tasks, the classes are ordered. However, little attention is paid to this type of learning tasks compared with traditional nominal classification learning (where no order is found between the classes). The characteristics of the climate modelling allows the problem to be defined as an ordinal classification problem (considering the discretisation of the original time series), in which the different classes (climate intervals), can be ordered from the smallest to the largest, in increasing order. The extreme classes, C1 and C5, represent tipping points.
Threshold models provide the regression framework from linear and non-linear models while treating the response variable rightfully as categorical. While threshold models are not the only type of ordinal regression models, they are by far the most popular class of ordinal regression models [11]. These methods project the patterns in a real line and determine a set of thresholds to divide the real line projection into consecutive intervals representing ordinal categories [11, 6, 3]. McCullagh's Proportional Odds Model (POM) [6] is a direct extension of the usual logistic regression model. Some authors refer to this model as the ordered logit model, because it is a generalization of the logit model to ordered response categories. This model may also be obtained from the latent variable formulation assuming that the error term has a standard logistic distribution.
To address the ordinal regression problem, an extension of the POM model is proposed. The model is composed by a potential function (f(x), to project the patterns into a real line) and a threshold vector (theta). The potential function of the model is a regression-type neural network. In this project, the selected basis function will be the product unit basis function [5]. The selection of the basis function is based on the nature of the problem: the detection of tipping points. In these problems, small changes in the input variables cause big modification in the dependent variable.
Besides, time series are usually highly correlated. Using the proposed model, the autocorrelation problem could be minimized (the product unit basis function models the interaction of the input variables).

3. **Imbalanced classification stage**
The most interesting classes are the extreme classes because they represent the tipping points. Therefore, the minority classes (C1 and C5) represent the interest to be accurately classified

(tipping points are usually a weird behaviour of the dynamical system). Data imbalance is a key source of degraded learning performance since state-of-the-art algorithms often assume a balanced class distribution. This imbalance often causes that the learning be biased towards the majority classes.

4. **Interpretation of the final model**

In order to detect the patterns characteristic to the time span preceding tipping points, an analysis of the impact of each input variable of the model (the lagged values of the dependent variable) on the final dependent variable is required. Two possibilities to interpret the final model are:

- Perform a sensitivity analysis.
- Convert the final model into a rule-based decision model.

The objective of the project in this last stage, is to validate the Scheffer hypothesis about the characteristics of the patterns preceding a tipping point [13]. To do that, we can apply hypothesis testing considering the model obtained in the project. To validate the project, the datasets published by Livina & Lenton in 2007 [14] will be utilized. This work could use data from the Greenland ice-core GISP2 paleo-temperatures (The datasets is available on the web page: http://mclean.ch/climate/Ice_cores.htm), though universities and research centres are encouraged to suggest different, potentially better-suited data sets.

4. ACT Contributions

This Ariadna project proposal is addressed at research groups with expertise in time series analysis, ordinal regression and imbalanced classification.

The project will be conducted in close scientific collaboration with ACT-researchers. In particular, ACT-researchers will provide expertise in time series analysis and climate modelling.

5. References

[1] A. D. Barnosky et al.. Approaching a state shift in Earth/'s biosphere. Nature, 486(7401):52-58, 2012.

[2] V. Dakoset al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PLoSONE, 7(7):e41010, 07 2012.

[3] R. Herbrich et al. Large margin rank boundaries for ordinal regression. In Advances in Large Margin Classifiers, pages 115-132. MIT Press, 2000.

[4] Miroslav Kubat and Stan Matwin. Addressing the curse of imbalanced training sets: one-sided selection. In Proc. 14th International Conference on Machine
Learning, pages 179-186. Morgan Kaufmann, 1997.

[5] A. C. Martinez-Estudillo, et al. Evolutionary product unit based neural networks for regression. Neural Networks, 19(4):477-486, 2006.

[6] P. McCullagh. Regression models for ordinal data. Journal of the Royal Statistical Society, Series B, 4, 1980, pages 109-142,.

[7] H. S. Oh, et al. Multi-resolution time series analysis applied to solar irradiance and climate reconstructions. Journal of Atmospheric and Solar-terrestrial Physics, 65(2):191-201, 2003.

[8] Intergovernmental Panel on Climate Change. Fourth Assessment Report: Climate Change 2007: Working Group I Report: The Physical Science Basis. Geneva: IPCC, 2007.

[9] M. Pazzani, et al. Reducing misclassification costs: Knowledge intensive approaches to learning from noisy data. In Proceedings of the 11th International Conference on Machine Learning (ICML-1994), pages 100-109, 1994.

[10] L. A. S. Romani, et al., Caetano Traina Jr., and Agma J. M. Traina. Mining relevant and extreme patterns on climate time series with clipsminer. JIDM, 1(2):245-260, 2010.

[11] Jan Verwaeren, et al. Learning partial ordinal class memberships with kernel-based proportional odds models. Computational Statistics and Data Analysis, 56(4):928-942, 2012.

[12] T. Lenton, Early warning of climate tipping points. Nature Climate Change, 1:201-209 (2011)

[13] M. Scheffer, Critical Transitions in Nature and Society. Princeton University Press (2012)

[14] Livina, V. N. and Lenton, T.M. A modified method for detecting incipient bifurcations in a dynamical system, Geophysical Research Letters 34, L03712. (2007)