symmetry

Editorial

Special Issue: "The Mixture Transition Distribution Model and Other Models for High-Order Dependencies"

André Berchtold †

Institute of Social Sciences & NCCR LIVES, University of Lausanne, CH-1015 Lausanne, Switzerland; andre.berchtold@unil.ch
† Current address: Institute of Social Sciences, University of Lausanne, Géopolis/SSP, CH-1015 Lausanne, Switzerland.

Citation: Berchtold, A. Special Issue: "The Mixture Transition Distribution Model and Other Models for High-Order Dependencies". *Symmetry* **2022**, *14*, 206. https://doi.org/10.3390/sym14020206

Received: 17 January 2022
Accepted: 19 January 2022
Published: 21 January 2022

1. Context

High-order Markov chains are very useful for the analysis of complex temporal relationships, but they generally require a very high number of parameters. A good approach is then to approximate them, and since its introduction in 1985 by Adrian Raftery, the mixture transition distribution (MTD) model has attracted much attention, thanks to its parsimony and versatility [1]. The MTD model uses a symmetry principle allowing a significant reduction in the complexity of a situation: instead of considering all possible interactions between past events to explain the present, it considers separately the effect of each lag in the model with an assumed similar form of effect across lags.

The MTD model has quickly found wide application for at least three reasons: First, many longitudinal data may require effects of order greater than one to be properly represented, but the use of fully parameterized models often proves to be impossible because of too many parameters to estimate given the number of data available. Second, and as symbolized by Figure 1, the MTD model can be easily extended to different types of data (continuous, multivariate, etc.); it can also incorporate covariates, and it can be combined with other models. Third, the results of MTD models are easy to interpret, in particular because they make it possible to distinguish the relative effect of the different explanatory variables of the model (the different lags for example), from the effect of the different modalities taken by these same explanatory variables.

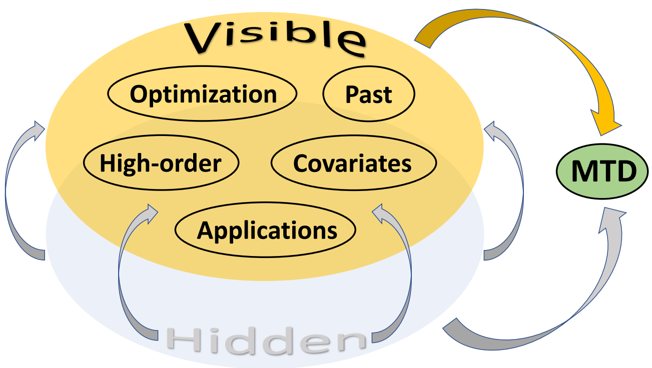

Figure 1. Graphical representation of the richness of the MTD principle.

In this context, the objective of this Special Issue of Symmetry was to propose both recent developments and original applications showing the richness and the diversity of the principles underlying the MTD model.

2. Contributions

The six articles in this Special Issue cover three different aspects of the MTD model: the development of the model itself, its interpretation, and its application to various fields.

To begin with, two articles deal specifically with the model itself. Danilo Bolano discusses the different methods of including covariates in the case of a model for categorical variables, which is notoriously more complicated than for continuous variables because of the large number of parameters involved [2]. The use of the same principle as the one used for the model's lags then allows easily comparing the influence of the covariates with that of the other components of the model. Such developments are essential, but for the end-user of statistical tools, it is sometimes difficult to quickly apply new approaches described in the literature because of a lack of implementation of these tools in standard software. This is why another article in this Special Issue presents a package for the open-source environment R called *march*, which allows estimating not only different variants of the MTD model, but also other types of Markovian models [3].

To develop its full potential, a model must be interpreted in a simple way, and we must have tools that allow us to judge its quality and interest in relation to a given situation. Confidence intervals are one of the tools that can be used for this purpose, and the article by André Berchtold proposes a new approach to obtain an approximation of these intervals for all the parameters of an MTD model [4]. By extension, this same approach can also be applied to other models such as hidden Markov models. Another article focuses on how to evaluate the amount of information contained in a transition matrix [5]. This is particularly useful here, since in the categorical case, the MTD model consists precisely of considering different transition matrices and using them sparingly. Criteria for differentiating the matrices are therefore useful, and this article links to the model estimation methods described in another article in this Special Issue [3].

The third area is applications. Here, the article proposed by Mohamed Yusuf Hassan is particularly interesting, because not only is the MTD model applied to a domain in which it has never been used before (respiratory infections), but moreover, the model is compared to other statistical tools; it is shown that it proves to be a very good competitor of a more classical model such as the SARIMA model [6]. Finally, the article by Zhivko Taushanov and Paolo Ghisletta proposes another innovative application using cognitive skills data [7]. Here again, a comparison is proposed with two types of GMM models. Moreover, this paper also establishes a strong link with the more theoretical developments of the MTD model by considering small sample sizes, including covariates and discussing the use of the model for clustering problems.

3. Further Developments

Even if it was proposed almost 40 y ago and if it finds part of its roots in an even older model, the Markov chain, the MTD model is still far from having revealed its full potential. The articles gathered in this Special Issue are an illustration of this, but we can think that much more is still to come. Indeed, we are currently witnessing an almost exponential increase in the complexity of the information available. The days of social scientists working only on patiently collected survey data seem to be over, and we have entered the era of "big data". If the sources of data and the very forms of these data become more diverse, it seems inevitable that the results derived from these data will only become more complicated. Therefore, anything that can efficiently explore this new jungle and present relevant results exploring symmetry or similarity effects is relevant. In this sense, the MTD model has its place in the toolbox of the modern researcher.

Obviously, for this to be the case, it is necessary to continue to develop and tailor the MTD model to new needs. We can then point to at least two areas in which work still needs to be performed. First, a better integration should be achieved between models for categorical variables and models for continuous variables, the ideal being to end up with a unified model allowing treating any variable in a similar way. The same should also be done for covariates. Then, it would be desirable to be able to analyze dyadic or even higher-order data more easily. Indeed, phenomena, especially social ones, are not easily represented by a single variable, but are most frequently presented as indicators of a latent concept or as networks. Thus, the challenge is to simultaneously identify and analyze everything that is significantly associated, while not getting lost in the arcana of complexity. In this game, the MTD model seems to have a role to play.

Funding: This publication benefited from the support of the Swiss National Centre of Competence in Research LIVES - Overcoming vulnerability: Life course perspectives, which is financed by the Swiss National Science Foundation (Grant Number 51NF40-160590). The author is grateful to the Swiss National Science Foundation for its financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; nor in the decision to publish the results.

References

1. Raftery, A.E. A model for high-order Markov chains. *J. R. Stat. Soc. B* **1985**, *47*, 528–539. [CrossRef]
2. Bolano, D. Handling Covariates in Markovian Models with a Mixture Transition Distribution Based Approach. *Symmetry* **2020**, *12*, 558. [CrossRef]
3. Berchtold, A.; Maitre, O.; Emery, K. Optimization of the Mixture Transition Distribution Model Using the March Package for R. *Symmetry* **2020**, *12*, 2031. [CrossRef]
4. Berchtold, A. Confidence Intervals for the Mixture Transition Distribution (MTD) Model and Other Markovian Models. *Symmetry* **2020**, *12*, 351. [CrossRef]
5. Berchtold, A. The Predictive Power of Transition Matrices. *Symmetry* **2021**, *13*, 2096. [CrossRef]
6. Hassan, M. The Deep Learning LSTM and MTD Models Best Predict Acute Respiratory Infection among Under-Five-Year Old Children in Somaliland. *Symmetry* **2021**, *13*, 1156. [CrossRef]
7. Taushanov, Z.; Ghisletta, P. The Use of a Hidden Mixture Transition Distribution Model in Clustering Few but Long Continuous Sequences: An Illustration with Cognitive Skills Data. *Symmetry* **2020**, *12*, 1618. [CrossRef]