TEST (2014) 23:473–477
DOI 10.1007/s11749-014-0387-1

DISCUSSION

Comments on: Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates

Leonard J. Paas

Published online: 21 August 2014

1 Introduction

First of all I find that the authors have written an excellent, comprehensive framework on latent Markov modeling. This model and related models, such as mixture models, latent class models and hidden Markov models, have been covered extensively in the literature (e.g., De Angelis and Paas 2013; Hokimoto and Shimizu 2014; Paas et al. 2007; Wedel and Kamakura 2000). Bartolucci, Farcomeni and Pennoni (BFP) addressed many important issues from this extensive body of literature, which resulted in an exhaustive theoretical framework.

In this comment, I will discuss the practical application of the latent Markov model, which deserves additional attention. I propose that the latent Markov model has potential for predictive purposes, which is highly relevant for practitioners, as will be discussed in Sect. 2 of this comment. Section 3 discusses an important caveat for the practical application of latent Markov modeling, involving the selection of an appropriate number of latent classes. The limited space of a comment does not facilitate addressing a large range of major issues for the application of the latent Markov model, but can provide a starting point.

This comment refers to the invited paper available at doi:10.1007/s11749-014-0381-7.

I am grateful to Meike Morren and Jeroen Vermunt for the useful suggestions on a previous version of this comment.

L. J. Paas (✉)
Department of Marketing, Faculty of Economics and Business, VU University,
Room 6A31, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
e-mail: l.j.paas@vu.nl

2 The applications of predicting future observed responses in latent Markov models

Applied academics and practitioners often use statistical models for prediction purposes rather than for explaining phenomena (Hair et al. 2009). Section 3.4 of the BFP-paper, based on Paas et al. (2007), addresses the application of the latent Markov model for predicting binary ownership indications for financial products. Moreover, promising findings on the application of hidden Markov modeling for prediction have been reported by De Angelis and Paas (2013) and Hokimoto and Shimizu (2014). The hidden Markov model is closely related to the latent Markov model, but is usually applied for time series predictions for a single unit, as also pointed out in the BFP-paper. Latent Markov modeling is for assessing responses of multiple units over time. BFP used ideas stemming from the hidden Markov literature for discussing the latent Markov model. I suggest that findings on using the hidden Markov model for prediction purposes can also could also be relevant when applying latent Markov models for this purpose. This will be addressed further below.

The BFP-paper distinguishes between the latent and measurement component in the latent Markov model. The former component describes class sizes at the first measurement occasion and transition probabilities between latent classes across consecutive time-points. The measurement component provides conditional probabilities for having specific values on observed responses, given a subject's latent class membership probabilities. Based on the BFP-notation, predicted values on the r observed response variables (product ownership indications in this comment) at a future time-point $T+1$ for each subject i are collected in the vector $\boldsymbol{Y}_{\boldsymbol{i}}^{(\boldsymbol{T+1})}$. Paas et al. (2007) introduced equations for predicting values in $\boldsymbol{Y}_{\boldsymbol{i}}^{(\boldsymbol{T+1})}$ for all n subjects in the analyzed dataset, as pointed out in section 3.4 of the BFP-paper. This results in a matrix $\boldsymbol{Y}^{(\boldsymbol{T+1})}$, with n rows and r columns. $\boldsymbol{Y}^{(T+1)}$ is derived from the values that the n subjects have on the r observed responses and the x covariates at time-point T and all previous times, and the latent Markov model that was derived from these data. For subject i prior latent class membership probabilities for all k latent classes at $T+1$ are predicted and stored in the vector $\boldsymbol{U}_{\boldsymbol{i}}^{(\boldsymbol{T+1})}$. This is based on the Baum–Welch algorithm, which is also discussed in section 5 of the BFP-paper. These predictions are combined with the conditional distributions across all observed response variables over the latent classes, stored in the matrix $\boldsymbol{Y}^{(\boldsymbol{t})}|\boldsymbol{U}^{(\boldsymbol{t})}$, i.e., the measurement component. This yields values for the vector $\boldsymbol{Y}_{\boldsymbol{i}}^{(\boldsymbol{T+1})}$ for each n, resulting in the matrix $\boldsymbol{Y}^{(\boldsymbol{T+1})}$. To explain the rationale underlying this prediction method, consider that a subject is more likely to own a credit card at $T+1$ when she/he has a larger probability to belong to a latent class at $T+1$ in which a high-conditional ownership probability occurs for this specific product (Paas et al. 2007). The equations introduced in Paas et al. (2007) provide the exact probabilities that the n subjects own this product at $T+1$ and the other $r-1$ products. These predicted values can be calculated in computer programs (Vermunt and Magidson 2013).

Such predictions are obviously useful for banks. Those consumers who do not own, for example, a credit card at T, but are likely to switch to a latent class with a high-conditional probability for owning this product at $T+1$, could be offered this product.

Such an offer may prevent acquisition of this product at a competing bank. Alternative models for such prospect selection purposes have been discussed in the literature (e.g., Knott et al. 2002; Paas and Molenaar 2005). I propose that the predictive potential of latent Markov modeling should be compared with the alternatives. The literature on the predictive potential of the hidden Markov model can provide avenues for such further research.

In a recent paper, De Angelis and Paas (2013) apply hidden Markov modeling for predicting the value of an important US stock index at $T + 1$ and later time-points. Such predictions are useful for supporting investment decisions. If the stock index is more likely to increase in value, it is more prudent to invest and vice-versa. The model can also be used to pinpoint the end of a financial crisis. If the hidden Markov model predicts a stable or increasing value of the stock index for a period of at least 13 weeks, it is likely that stock market has entered a phase with little volatility (De Angelis and Paas 2013). De Angelis and Paas (2013) do point out that further research is required, but potential implications for investors and policy-makers are obvious. Interestingly, the hidden Markov model that is reported in De Angelis and Paas (2013) outperforms alternatives for predicting stock index developments, e.g., the commonly used GARCH model.

In another practical application, Hokimoto and Shimizu (2014) used two alternative hidden Markov models for predicting sea surface elevation, refer to their paper for descriptions. The predictions in Hokimoto and Shimizu (2014) are also practically relevant, addressing the evaluation of human safety risks in sea activities, such as navigation and fishing. One of the two applied hidden Markov models in Hokimoto and Shimizu (2014) outperforms the other hidden Markov model and also alternatives such as the ARIMA model. The promising hidden Markov model applications and results (De Angelis and Paas 2013; Hokimoto and Shimizu 2014), suggest that future research could aim to assess the conditions under which different formulations of the latent Markov model provide more accurate or useful predictions than alternative statistical models in a large range of applied domains.

3 Assessing the optimal number of latent classes

This section addresses an important caveat of using the latent Markov model for applied (prediction) purposes, i.e., selecting an appropriate number of latent classes in large datasets. This is salient due to increased availability of large datasets, also referred to as big data (Manyika et al. 2011). The BFP-paper mentions various so-called information criteria, and suggests BIC as effective for selecting an optimal number of latent classes. However, research comparing the effectiveness of information criteria, and other approaches for selecting an optimal number of latent classes, is usually based on relatively small datasets. Tuma and Decker (2013) presented an extensive overview on approaches for selecting an appropriate number of latent classes for the more general family of segmentation models. Latent Markov modeling can be considered as a technique for conducting longitudinal segmentation analyses. In the overview, the largest simulation datasets contain 2,500 cases, most are smaller. In a recent paper, which was also cited in the BFP-paper, Bacci et al. (2013) evaluated criteria for determining the number of latent classes in latent Markov models, using simulated datasets with up to 500 cases.

The latent Markov model is often applied to larger datasets, e.g., $n = 7{,}676$ in Paas et al. (2007), see section 3.4 of the BFP-paper. Such datasets often lead to a large number of latent classes. The values on information criteria such as AIC and BIC keep declining when the number of classes increases. This also applies for the CAIC criterion, which penalizes additional parameters more strongly. Some of these latent classes may contain very small (fractions of) percentages of the subjects. An interesting approach that may mitigate such issues was discussed in Magidson and Vermunt (2004), see also Vermunt and Magidson (2013). This bivariate residual (BVR) approach quantifies the extent to which a model reproduces associations between the observed responses. The approach is based on the Pearson chi-square statistic and the number of degrees of freedom. Lower values imply a better reproduction of the observed associations (Magidson and Vermunt 2004). If all BVR's have an approximate maximum value of 1, the model is considered adequate. Magidson and Vermunt (2004) suggest that larger BVR's may sometimes occur for one or a few specific observed bivariate association(s). They mention an example with four observed responses, A, B, C and D. In this specific example, a poor fit of a two-class model is found, but this only applies for the association between the observed responses to C and D. Magidson and Vermunt (2004) suggest that researchers can accommodate for this by adding a direct effect between C and D, implying partially relaxing the local independence assumption, referred to as contemporary dependence in the BFP-paper. This prevents increasing the number of latent classes, on the account of a single bivariate association.

The increasing interest in big data implies that future simulation studies on selecting an optimal number of classes in latent Markov models, and related models, should include larger simulated datasets. Such studies may assess the effectiveness of the more traditional information criteria, e.g., AIC and BIC, and potential alternatives, e.g., the BVR approach (Magidson and Vermunt 2004), the likelihood ratio test or the quality of classification approach mentioned in the BFP-paper. Moreover, statisticians can aim to develop new criteria that ensure latent classes fulfill a specific minimum size in large datasets, as sizability is an important criterion for many applications of models that are used for classifying subjects (Wedel and Kamakura 2000).

In sum, the latent Markov model has already been applied in an impressive range of domains, as is apparent in the BFP-paper. Research on selecting the optimal number of latent classes in large datasets may further enhance relevance of the latent Markov model for applied academics and practitioners, e.g., database analyst in firms. This also applies for studies comparing the predictive potential of latent Markov models with alternative models in different domains, as discussed in Sect. 2 of this comment. The theoretical framework presented in the BFP-paper is highly useful for these directions for further research and also for other future research avenues on the practical applications of the latent Markov model.

References

Bacci S, Pandolfi S, Pennoni F (2013) A comparison of some criteria for states selection in the latent Markov model for longitudinal data. Adv Data Anal Classif 1–21. doi:10.1007/s11634-013-0154-2

De Angelis L, Paas LJ (2013) A dynamic analysis of stock markets using a hidden Markov model. J Appl Stat 40:1682–1700

Hair JF, Black WC, Babin BJ, Anderson RE (2009) Multivariate data analysis, 7th edn. Prentice Hall, Upper Saddle River

Hokimoto T, Shimizu K (2014) A non-homogeneous hidden Markov model for predicting the distribution of sea surface elevation. J Appl Stat 41:294–319

Knott A, Hayes A, Neslin SA (2002) Next-product-to-buy models for cross-selling applications. J Interact Market 16:59–75

Magidson J, Vermunt, J (2004) Latent class models. In: Kaplan D (ed) The SAGE handbook of quantitative methodology for the social sciences. Sage, Thousand Oaks

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity. The McKinsey Globa Institute, Seoul, San Fransico, London, Washington DC

Paas LJ, Molenaar IW (2005) Analysis of acquisition patterns: a theoretical and empirical evaluation of alternative methods. Int J Res Market 22:87–100

Paas LJ, Vermunt JK, Bijmolt THA (2007) Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisition patterns of financial products. J R Stat Soc Ser A 170:955–971

Tuma M, Decker R (2013) Finite mixture models in market segmentation: a review and suggestions for best practices. Electron J Bus Res Methods 11:2–15

Vermunt JK, Magidson J (2013) Latent GOLD 5.0 upgrade manual. Statistical Innovations Inc, Belmont

Wedel M, Kamakura WA (2000) Market segmentation: conceptual and methodological foundations, 2nd edn. Kluwer, Amsterdam