Probabilistic Modelling of Influences on Travel Decision Making

Mario Pichler[a], Lisa Steiner[a], and Herta Neiß[b],

[a]Data Analysis Systems Group
Software Competence Center Hagenberg GmbH, Austria
[mario.pichler|lisa.steiner]@scch.at

[b]Tourism Management
Johannes Kepler University Linz, Austria
herta.neiss@jku.at

Abstract

For any tourism organisation or company, it is a necessity to know about the factors that are influencing tourists' travel decisions. The question, however, is how to model and represent heterogeneous influence factors in a way that a) human decision makers can easily understand, and b) allows for computer-based simulation and diagnoses to aid decision making. Currently, tourism suffers from meaningful and practically applicable representations of complex relations among influence factors stemming from different domains. This paper investigates Bayesian networks for modelling mutual influences of factors originating from heterogeneous data sources including tourism experts, and the integration of associated uncertainties in a single model. The authors are elaborating several development alternatives for the creation of a Bayesian network-based tourism knowledge model. Using this model, tourism professionals will be able to perform interactive decision analyses for determining, e.g., how to spend marketing budget most efficiently.

Keywords: Tourism influence factors; Web 2.0/social media; tourism knowledge model; Bayesian networks; probabilistic structural equation models; decision support.

1 Introduction

A number of factors that are changing in the course of time are influencing tourists' decision-making processes, e.g., destination or hotel choices. Market research aims at investigating these factors and capturing them in tourism statistics. Especially, recent advances in information and communication technologies (ICT) open up new ways for tourists to collect decision-relevant information and, as one consequence, to make travel decisions in an increasingly spontaneous manner. In particular, Web 2.0/social media developments allow tourists to make their decisions on alternative foundations that are currently finding their way into market research and tourism statistics.

Following these developments, the objectives of an ongoing tourism research project are: 1) to identify novel influence factors in Web and Web 2.0 channels, and 2) to find and model relationships among those new as well as already known influence factors (Pichler et al., 2014). The result of this approach will be a simulation/diagnosis model that allows for analysing interrelations of factors that are determining (purchase) decisions of tourists. Using this model, tourism professionals shall be provided with novel insights that are acting as foundations for investment decisions, e.g., how much effort and budget to invest in a specific communication channel to adequately reach a desired target group.

2 Related Work

This work addresses several recently mentioned issues of tourism (marketing) research. Dolnicar & Ring (2014) argue that research approaches in tourism marketing concerned with deepening the understanding of cause-effect relations as base for strategic principles are rare. Tourism knowledge representation represents a further concern. Hjalager (2010) points out the importance of adequate knowledge representation for tourism innovation research. Currently, tourism statistics (tables), textual explanations, and fragmented research work – mostly focussing on a limited number of specific aspects – are representing tourism domain knowledge.

Structural equations models (SEM) are representing a prominent and widespread adopted research method for investigating interrelations of aspects in tourism (Kline, 2011; Baggio & Klobas, 2011). To a certain degree, SEM allow for analyses of cause-effect relations among investigated aspects. Nunkoo & Ramkissoon (2012) present an overview of studies utilising SEM modelling until 2012. Among others, Marchiori et al. (2013) and Tussyadiah & Zach (2013) recently published SEM-based research analysing the impacts of novel ICT-based developments on tourism.

Beside the advantages of SEM, researchers and others are also controversially discussing them. The criticism include a) violations of proper method application or wrong conclusions (e.g. Nunkoo & Ramkissoon, 2012); b) results are primarily of academic nature and need to be further prepared for practical application (https://www.researchgate.net/post/How_would_you_rate_the_impact_practical_relevance_of_structural_equation_model_insights_on_the_tourism_industry [Oct. 16, 2014]); and c) published SEM mostly just cover a fraction of (practically) interesting relations of aspects.

3 Innovation and Approach

Based on aforementioned analyses of related work, this work aims at the development of comprehensible formal representations of complex interrelations of aspects in tourism. The approach foresees the development of a knowledge model that integrates a) existing tourism data, b) implicit tourism expert knowledge, and c) novel influence factors extracted from Web 2.0/ social media channels, to allow for interactive analyses of (complex) relations.

Explicitly addressing the mentioned criticism on SEM, this work investigates an alternative modelling approach. The authors are examining Bayesian network (BN) models (Pearl, 1985) – which represent a specific kind of probabilistic graphical models (Koller & Friedman, 2009) – for analysing mutual influences on tourists' travel decision making.

BNs are directed acyclic graphs. They allow for modelling of direct dependencies, independencies as well as conditional independencies between model nodes. For each model node, a probability distribution of its states needs to be defined. For model nodes that are dependent on others, the states' probabilities are conditioned on each combination of parent nodes' states. In this way, BNs allow for modelling uncertain relations among variables (model nodes).

Relevant features of BNs for the work at hand are: i) decision making under uncertainty through probabilistic relations among several influence factors; ii) possibility to integrate heterogeneous data sources and models, e.g., wide-spread structural equation models dealing with different aspects of tourism, local tourism data, and human expert knowledge; iii) updating the beliefs (prior probabilities) of the whole model as soon as new evidence on at least one of the model nodes is available, e.g., "real-time" data about website access; iv) possibility for interactive model investigation to perform what-if or diagnostic analyses.

In contrast to SEM, tourism research mostly neglected BNs so far. The works of Huang & Bian (2009) and Hsu et al. (2012) are among those few studies. Both are applying BN models for determining the probabilities of tourist attractions appealing to particular tourists.

4 Exploration of Bayesian Network Modelling Approaches

The following sections are introducing three potential approaches for the development of a Bayesian network-based tourism knowledge model. These are: 1) manual extraction and integration of model components from previous (SEM) studies, 2) linkage of existing structural equation models to Bayesian networks, and 3) semi-automated Bayesian network model learning from tourism data sources.

4.1 Manual Bayesian network model composition based on previous studies

For this first approach, previous studies investigating influence factors on tourism behaviour are acting as knowledge sources for model generation. Literature research covered major tourism research journals, computer science journals publishing BN-related research, and the ENTER conference series' proceedings. Because of space limitations, the references section of this work lists only an excerpt of identified research. A number of studies applied SEMs to identify causal relations among several of those factors. Therefore, single fragments (single influence factors or several related factors represented as SEMs) are available for integration in a subsequent step.

At that point, a number of question arises, e.g., how to parameterise source model nodes in the target BN model, or how to maintain causal relations of several source SEMs in a resulting BN combining those SEMs. The achievable model quality, however, will be high, because existing tourism domain knowledge builds the base.

The integration of novel influence factors extracted from novel Web and Web 2.0 channels requires close collaborations with tourism experts. Associated use case partners are offering this expertise in this work's frame. Overall, this approach tends to be very labour-intensive, as most of the work needs to be carried out manually by humans. For that reason, further alternatives are investigated as well.

4.2 Linkage of existing structural equation models to Bayesian networks

The second approach follows Gupta & Kim (2008), who propose and introduce how to link SEMs to BNs with the aim of combining the strength of both. These are the capabilities of SEMs in empirical validation, and the prediction and diagnosis capabilities of BN modelling. At first, an empirically validated SEM is used to deduce the structure of the BN model. Thereafter, the latent variable scores from the SEM are used for learning the conditional probabilities of the BN model.

As the authors already identified a number of SEM studies, as mentioned above, this second alternative seems to be quite suitable. However, the major drawback of this approach obviously lies in the requirements concerning data availability. The process as introduced by Gupta & Kim (2008) would require all the data sources used within the corresponding SEM studies. This requirement seems to be rather unrealistic.

4.3 Bayesian network model learning from tourism data sources

The third potential approach originates from a data-driven perspective. Data mining and machine learning techniques are creating an initial BN model structure, based on available tourism data sources. In a next step, (human) tourism experts are refining and validating this model structure.

Publicly available BN modelling software like GeNIe/Smile[1] is offering support for learning models from data. Conrady & Jouffe (2013) described an especially interesting approach that generates so-called probabilistic structural equation models (PSEMs) from data. Like the aforementioned second approach, also PSEMs aim at combining the advantages of SEMs and BNs in a single method. Moreover, a sophisticated commercial tool, BayesiaLab[2], implements this approach. Other than GeNIe/Smile, BayesiaLab supports the derivation of latent variables. In addition to the traditional SEM approach, the integration of BN features allows for what-if analyses and diagnoses of modelled relationships.

First experiments applying this approach on relevant tourism data sources, e.g., tourism statistics databases or website access data have shown that the achievable result quality and explanatory power depends on the characteristics of the data sources. The nature of the used tourism data sources considerably differed from the BayesiaLab example cases. Through further efforts in data preparation, the authors are expecting benefits of applying the PSEM approach for achieving the envisioned tourism knowledge model.

Derived from this exploration, the further course of action foresees a combination of manual extraction and integration of model components from previous tourism studies, and semi-automated Bayesian network model learning from tourism data sources, with a special focus on Web and novel Web 2.0 channels.

5 Conclusion and Outlook

This work investigated Bayesian networks for modelling and analysing factors that are influencing tourists' travel decisions with a specific focus on practical usage by tourism professionals. Tourism research has mostly neglected Bayesian networks so far, while the latter are successfully applied in a number of other disciplines. Most important, Bayesian networks a)represent an intuitive modelling approach that is easily understandable for experts of different disciplines, and b) allow for interactive

---

[1] https://dslpitt.org/genie/
[2] http://www.bayesia.com/en/products/bayesialab.php

analyses (simulation, diagnosis) of action options in the tourism sector. Novel developments, as depicted in the exploration section, are investigating combinations of structural equation models, which are widely adopted in tourism research, and Bayesian networks, to integrate simulation and diagnosis support. Hence, the authors will pursue this combined approach for developing a practically applicable tourism knowledge model that enables tourism professionals to carry out interactive decision analyses.

References

Baggio, R. & Klobas, J. (eds., 2011). *Quantitative Methods in Tourism: A Handbook*. Channel View, Bristol, UK.

Conrady, S. & Jouffe, L. (2013). Tutorial on Driver Analysis and Product Optimization with BayesiaLab. Available online: http://library.bayesia.com/display/whitepapers/ Driver+Analysis+and+Product+Optimization [last access: 2014/09/06].

Dolnicar, S. & Ring, A. (2014). Tourism marketing research: Past, present and future. *Annals of Tourism Research* 47(1): 31-47.

Gupta, S. & Kim, H. W. (2008). Linking structural equation modeling to Bayesian networks: Decision support for customer retention in virtual communities. *European Journal of Operational Research* 190(3): 818-833.

Hjalager, A.-M. (2010). A review of innovation research in tourism. *Tourism Management* 31(1): 1-12.

Hsu, F.-M., Lin, Y.-T. & Ho, T.-K (2012). Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps. *Expert Systems with Applications* 39(3): 3257-3264.

Huang, Y. & Bian, L. (2009). A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. *Expert Systems with Applications* 36(1): 933-943.

Kline, R.B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guildford Press.

Koller, D. & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press.

Marchiori, E., Milwood, P. & Zach, F. (2013). Drivers and Benefits of Analysing DMOs' eWOM Activities. In Cantoni, L. & Xiang, Z. (P. (Eds.), *Information and Communication Technologies in Tourism 2013*, Springer Berlin Heidelberg, 107-118.

Nunkoo, R. & Ramkissoon, H. (2012). Structural equation modelling and regression analysis in tourism research. *Current Issues in Tourism* 15(8): 777-802.

Pearl, J. (1985). Bayesian networks: a model of self-activated memory for evidential reasoning. *Proceedings Cognitive Science Society* 1985, August, UC Irvine, 329-334.

Pichler, M., Rutzinger, M. & Neiss, H. (2014). insightTourism – Towards Understanding Online Behavior of Tourists. *e-Review of Tourism Research (eRTR)* 11(3/4): 50-53.

Tussyadiah, I. & Zach, F. (2013). Social Media Strategy and Capacity for Consumer Co-Creation Among Destination Marketing Organizations. In Cantoni, L. & Xiang, Z. (P. (Eds.), *Information and Communication Technologies in Tourism 2013*, Springer Berlin Heidelberg, 242-253.

Acknowledgements

This work has been funded by the Austrian Research Promotion Agency (FFG) within the COIN program line (Prj.No. 841065) and the program 'Talente: Praktika für Schülerinnen und Schüler 2014'.