

Florian J. Zach Virginia Tech

Dejan Krizaj University of Primorska

Ajda Pretnar University of Ljubljana

Topic Modelling Tourism Literature on Innovation and Technology

Tourism research on technology and innovation has flourished for several decades. Especially the increased use of information and communication technologies in tourism triggered more interest in either topic since the beginning of the early 2000s. To learn more about the wealth of research at the intersection of tourism, technology and innovation this study applies topic modelling on abstracts of relevant studies to identify which research areas are covered to date. For today's body of research, we identified ten topics within five clusters. Hence, this study provides evidence that topic modelling is a meaningful approach to categorize past research and to identify opportunities for future research.

Key words: innovation, technology, topic modelling

Florian J. Zach Howard Feiertag Department of Hospitality and Tourism Management Pamplin College of Business Virginia Tech Wallace Hall 355 Blacksburg, VA 24061 USA Phone: +1 (540) 231-8425 Email: florian@vt.edu

Dejan Krizaj Faculty of Tourism Studies – Turistica University of Primorska Obala 11a 6320 Portorož Slovenia Phone: +386 5 617 70 21 Email: dejan.krizaj@fts.upr.si

Ajda Pretnar Bioinformatics Laboratory



Faculty of Computer and Information Science University of Ljubljana 1000 Ljubljana Slovenia Phone: +386 1 479 82 09 Email: ajda.pretnar@fri.uni-lj.si

Florian J. Zach is an Assistant Professor of innovation and entrepreneurship in tourism and hospitality at Virginia Tech with an interest in new technologies. His current research focuses on the implementation of technological advancements to create tourism experiences.

Dejan Krizaj is an Associate Professor at University of Primorska, Faculty of Tourism Studies – Turistica, focusing on R&D projects addressing specifics of tourism innovation and its stimulation and promotion measures.

Ajda Pretnar is a Researcher at the Faculty of Computer and Information Science, University of Ljubljana, where she teaches data mining and machine learning. Her current research deals with predicting tourism flows from traffic data.



Topic modelling tourism literature on innovation and technology

Tourism research on technology and innovation has flourished for several decades. Especially the increased use of information and communication technologies in tourism triggered more interest in either topic since the beginning of the early 2000s. To learn more about the wealth of research at the intersection of tourism, technology and innovation this study applies topic modelling on abstracts of relevant studies to identify which research areas are covered to date. For today's body of research, we identified ten topics within five clusters. Hence, this study provides evidence that topic modelling is a meaningful approach to categorize past research and to identify opportunities for future research.

Key words: innovation, technology, topic modelling

Introduction

Buhalis and Law (2008) and (Navío-Marco, Ruiz-Gómez, & Sevilla-Sevilla, 2018) argue to understand the role ICT in tourism scholars should focus on either consumers and their demand, technological innovation or the state of the industry. This research note focuses on innovation as it is a broad topic and incorporate several aspects of organizational decision making. Indeed, technological development and innovation are two commonly studied research areas in tourism and beyond. However, the variety of research ranges from firm performance (Salomo, Talke, & Strecker, 2008) to user acceptance in general (Venkatesh, Morris, Gordon, & Davis, 2003) to understanding determinants of innovation in tourism (Divisekera & Nguyen, 2018), social innovation in tourism (Kohler & Rutzler, 2018) and smart destinations (Lee, Lee, Chung, & Koo, 2017). To provide scholars with a better understanding of the tourism research area of innovation and technology investigated in the past, the purpose of this research note is to categorize existing research. To do so we employ topic modelling algorithms on existing research abstracts and identify ten meaningful research areas. Furthermore, we provide suggestions for future research avenues.



Literature Review

Innovation and technology

Technological innovation creates an economic disequilibrium that challenges current market setting (Iwai, 1984) and results in continuous changes and discontinuities in technological innovation (Dosi, 1982). Hence, the interplay of innovation and technology creates an opportunity for academics to investigate how market stakeholders plan for and develop new technologies (Hargadon & Sutton, 1997), adapt them (Chen & Tsou, 2007) and manage the related innovation activities and outcomes (Salomo et al., 2008). The importance of the connection between technology and innovation has also been studied in tourism research. Technology is often introduced from providers outside of tourism (Orfila-Sintes & Mattsson, 2009) requiring a confluence of absorptive capacity to integrate external knowledge and technology, internal knowledge and skills to apply internal and external knowledge to create something new (Souto, 2015). Gathering and absorbing knowledge for innovation activities is aided when firms collaborate with partners and stakeholders who are relevant drivers to leverage technologies for innovation (Aldebert, Dang, & Longhi, 2011).

Information and communication technologies are particularly relevant for innovation in tourism as it enables better communication as evidenced in the eTourism research stream (Navío-Marco et al., 2018) and aids in the development of smart and sustainable destinations (Romão & Neuts, 2017). Indeed, ICT is a key driver of process innovation (Hjalager, 2010) and its use has a positive effect on developing tourism marketing innovation (Divisekera & Nguyen, 2018), while a lack of knowledge in using ICT is a barrier for cooperation for innovation (Pikkemaat, Peters, & Chan, 2018).



Analysing Literature

The increase in both quantity and variety of research on the interplay of technology and innovation in tourism (e.g. Navío-Marco et al., 2018; Trunfio & Campana, 2019) resulted in a wide range of publications that can be assessed using a systematic review as conducted by Marasco, De Martino, Magnotti, and Morvillo (2018) for collaborative innovation in tourism. Other viable review strategies are critical/narrative reviews of the most significant items in the field in chronological order, meta-analyses, or reviews of text books and popular management book (Briner & Denyer, 2012). A limitation of these reviews is that not all publications can be reviewed as authors need to be selective to manage an ever-increasing number of publications. Recently, however, technological advancements, specifically in the area of machine learning, allow for the analysis of large swaths of publications (e.g. Amado, Cortez, Rita, & Moro, 2018).

Methodology

Topic Modelling

Topic modelling is a process of discovering topics in unstructured text and has been applied in tourism research to analyse – among many other topics – reviews by hotel guests (Calheiros, Moro, & Rita, 2017) and restaurant patrons (Rossetti, Stella, Cao, & Zanker, 2015). In its essence topic modelling presumes each document is composed of topics and each topic is composed of a set of words.

One of the popular techniques of topic modelling, Latent Dirichlet Allocation (LDA), finds topics by reverse-engineering the process of generating text. LDA starts with a key assumption, that each document is generated from a set of topics and further from a set of words defining these



topics. LDA then reverses the process by setting k topics (one of the parameters of the algorithm) and iteratively re-assigns words to topics based on their probability distributions in the document.

First, we applied extensive preprocessing to our documents. Besides lowercasing and tokenizing the words, we lemmatized them with WordNet lemmatizer (Bird, Loper & Klein, 2009), which transforms the words to their base forms (e.g. tourists to tourist). Then filtered out stopwords and corpus-specific words, which we manually defined. These words pertained mostly to the academic vocabulary, since we were dealing with articles, such as 'methodology', 'result', 'analysis' and so on. In the final bag-of-words representation we kept the words occurring in more than 10% and less than 90% of documents.

To determine the optimum number of topics, we ran three methods for the number of topics between two and fifty. Gensim's (Řehůřek & Sojka 2010) log perplexity computes how well the topics fit new data; we used 70% of the data for training and 30% for testing. Similarly, GridSearch algorithm from Python's sklearn suggested two topics. However, as reported in Chang et al. (2009), probability metrics do not correspond well to the human interpretation of topic coherence, hence we tried Gensim's Coherence Model (Röder, Both & Hinneburg, 2015) with Mallet LDA, which aims to capture human understanding sensible topics. The results reported local maxima around 10 topics. Finally, we ran LDA for 10 topics.

Data

We analysed abstracts of full-paper journal publications including the keywords "innovation", "tourism" and "technology" had to be mentioned in either title, abstract or author provided keyword list. Qualified publications had to be in written in English and published before 31st December 2018. Utilizing the ScienceDirect database, which includes publications from



Elsevier and all other major publishers, this effort resulted in 318 abstracts from 1982 to 2018

(Table 1).



(years with zero-count onnitied)						
Count	Year	Count	Year	Count		
2	1999	2	2010	15		
1	2001	4	2011	18		
1	2002	3	2012	16		
1	2003	4	2013	17		
2	2004	3	2014	23		
1	2005	7	2015	24		
1	2006	4	2016	41		
1	2007	9	2017	37		
1	2008	12	2018	57		
1	2009	10				
	Count 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	CountYear21999120011200212003220041200512006120071200812009	CountYearCount21999212001412002312003422004312005712006412007912008121200910	CountYearCountYear21999220101200142011120023201212003420132200432014120057201512006420161200792017120081220181200910-		

Table 1. Count of Qualifying Publications per Year(years with zero-count omitted)

Findings

We identified ten topics, of which some are closely related, while others are distinctively different. Projection of topic distribution is based on principal component analysis (PCA) transformation of word probabilities for each topic (Sievert & Shirley, 2014). PCA is a dimensionality reduction technique that projects the data into a lower dimensional space (in our case an easier to interpret a two-dimensional by maximizing the variance each component explains (Figure 1).

Upon manual analysis the retrieved topics with their corresponding words captured well with abstract texts and were highly interpretable. Saliency is the importance of each word for the topic (Chuang et al., 2012). If a word is frequent in a topic, but also in the entire corpus, it will get a lower saliency score than a word that is frequent in a topic alone (Table 2). Similarly, relevance is a weighted measure of word probability and lift, the ratio of a word's probability within a topic to its probability in the corpus.



Intertopic Distance Map (via multidimensional scaling)



Figure 1: Intertopic Distance Map (via multidimensional scaling)

Topic	Label	Top 10 salient terms
1	Climate change	emission, service, carbon, technology, process, carbon emission, product, two, context, allocation
2	Land use	technology, land, environmental, development, system, new, consolidation, land consolidation, use, business
3	Growth and economic development	geography, industry, technology, development, factor, economic, information, firm, hong kong, kong
4	Value creation	industry, sector, value, development, model, firm, district, product, chain, market
5	Smart destinations	technology, information, system, new, industry, development, smart, tour, city, application

 Table 2. Topics and Top 10 Salient Terms



6 Management of	technology, industry, development, knowledge,
technological	technological, process, management,
development	innovative, firm, model

Table 2 cont'd. Topics and Top 10 Salient Terms

Topic	Label	Top 10 salient terms
7	Destination websites	website, development, industry, information, technology, cluster, new, destination, service, travel
8	Collaboration for technology development	development, technology, industrial, website, child, park, company, industrial park, group, innovative
9	Travel experiences	new, technology, development, destination, industry, service, experience, sector, knowledge, building
10	ICT for destination management	technology, destination, information, ict, smart, use, management, development, new, virtual

Topic 1: Climate change

As seen in Figure 1 climate change is highly distinctive from all other topic. Literature of this topic investigates carbon emission rights, air travel innovation, sustainable travel mobility and issues related to the Paris Climate Change Agreement.

Topic 2: Land use

Optimizing the use of available land is a recurring issue in studies on smart destination, cocreation issues between tourism entities and hospitality businesses, site selection and studies specific to select geographies (e.g. Gran Canaria, Croatia, Scotland, Basque Country).

Topic 3: Growth and economic development



The use of technologies and innovation to gain an economic advantage is the focus of studies on specific geographic regions (e.g. Hong Kong, Great Plains) and includes issues such as open education, adoption of e-Tourism and Airbnb.



Topic 4: Value creation

Value for tourism stakeholders is created in many different forms. This topic includes studies on medical tourism, strategic and business model related studies, entrepreneurship, revitalization issues and issues involving sustainable tourism ecosystems and designated tourism districts.

Topic 5: Smart destinations

The development of smart destinations involves studies on information technologies and other technologies to enable "smartness" (e.g. near field communication), the opportunities that result from it such as increased experiences and visitor satisfaction.

Topic 6: Management of technological development

The development of new technologies requires not only management of that innovation process, but also management for implementation and adoption. This includes studies investigating the use of technologies for tourism education, drivers of technology development and adoption as well as innovations addressing environmental concerns. The latter is an overlap with topic 2 (land use).

Topic 7: Destination websites

Destination website studies assess managerial issues such as website control (overlap with topic 6), online vs. offline promotion in tourism, stimulating visitation, mobile website use, and website performance. Finally, opportunities to specialize and optimize destination usage with internet enabled wearables and website information relates to topic 2 (land use).





Topic 8: Collaboration for technology development

This topic covers several collaborative aspects supporting innovation and technology development ranging from crowdfunding to consumer co-creation as well as interregional and transatlantic partnerships. Aspects of economic development, such as alternatives for snow scarcity and research and development relate to topic 3 (growth and economic development).

Topic 9: Travel experiences

The myriad of drivers, determinants and outcomes of travel experiences is reflected in both the variety of issues covered by this topic and by its size – it is the largest topical area accounting for 17% of all terms. Issues covered include the use of ICT to reach potential travellers (overlap with topic 7 – destination websites), new experiences to stem survival in the face of technological change (overlap with topic 6 – management of technological development), creating new customer experiences by leveraging new technologies, as well as transportation innovations that affect aviation experiences and space tourism.

Topic 10: ICT for destination management

The use of ICT is of particular importance for smart destinations (overlap with topic 5 - smart destinations) and also covers studies on ICT adoption, usage of ICT to for heritage and cultural niche tourism to increase destination competitiveness.

Conclusion

We found that research on innovation and technologies in tourism is varied, but also that these distinctive groups overlap as they cover related topics. As can be seen from Figure 1, they



further imply five distinctive clusters. Topics 1 and 2 remain standalone; topic 1 focuses on technology and innovation', thus forming cluster **A** - **Environment Effects**. Topic 4, which addresses technology and innovation's value creation is cluster **B** - **Value Creation**. Topics 3 and 8 form cluster **C** - **Business Operations** aspects: growth, development and business collaboration. Topics 2, 6, 7 and 9 deal are cluster **D** - **Destination Aspects**: land, destination websites, travel experiences and management/development. Finally, topics 5 and 10 form cluster **E** - **Tourism ICT** ranging from essential to smart digital tools. An analysis of the available academic resources in the fields of technology and innovation in tourism thus reveals five focal factors that combine the two fields: technology (E) generates value (B) with business operations (C) at destinations (D) leaving a footprint (A).

Consequently, we found that topic modelling on abstracts, despite the limitation of utilizing only about up to 250 words, is a fruitful exercise to gain a comprehensive overview of an academic field. Such an exercise is a beneficial and relatively low-effort activity for new and established scholars to identify well covered research aspects and potential gaps. Hence, while topic modelling allows for an understanding of past research, it also highlights opportunities for future research.

However, given that the investigated field was narrow (innovation, tourism, technology) with only 322 abstracts, future research should develop in breadth and depth. First, breadth can be expanded by reducing keywords, thus broadening the field. Second, depth could be provided by further analysing a topic for different time windows to provide insights into the *evolution* of a field, its current *trends* and even its further development *forecasts*.

References

Aldebert, B., Dang, R. J., & Longhi, C. (2011). Innovation in the tourism industry: The case of Tourism@. *Tourism Management*, 32(5), 1204-1213. doi:10.1016/j.tourman.2010.08.010



- Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, 24(1), 1-7. doi:10.1016/j.iedeen.2017.06.002
- Briner, R. B., & Denyer, D. (2012). Systematic review and eidence synthesis as a pracitce and scholarship tool. In D. M. Rousseau (Ed.), *The Oxford Handbook Of Evidence-Based Management* (pp. 112-129). Oxford: Oxford University Press.
- Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet - The state of eTourism research. *Tourism Management*, 29(4), 609-623. doi:10.1016/j.tourman.2008.01.005
- Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675-693. doi:10.1080/19368623.2017.1310075
- Chen, J.-S., & Tsou, H. T. (2007). Information technology adoption for service innovation practices and competitive advantage: the case of financial firms. *Information Research*, *12*(3), paper 314. Retrieved from http://InformationR.net/ir/12-3/paper314.html
- Divisekera, S., & Nguyen, V. K. (2018). Determinants of innovation in tourism evidence from Australia. *Tourism Management*, 67, 157-167. doi:10.1016/j.tourman.2018.01.010
- Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3), 147-162. doi:10.1016/0048-7333(82)90016-6
- Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly*, 42(4), 716-749. doi:10.2307/2393655
- Hjalager, A.-M. (2010). A review of innovation research in tourism. *Tourism Management*, 31(1), 1-12. doi:10.1016/j.tourman.2009.08.012
- Iwai, K. (1984). Schumpeterian dynamics: An evolutionary model of innovation and imitation. Journal of Economic Behavior & Organization, 5(2), 159-190. doi:10.1016/0167-2681(84)90017-9
- Kohler, T., & Rutzler, L. (2018). Crowdsourcing Social Innovation in Tourism: Insights on Platform Design, Cham.
- Lee, J., Lee, H., Chung, N., & Koo, C. (2017). An Integrative Model of the Pursuit of Happiness and the Role of Smart Tourism Technology: A Case of International Tourists in Seoul, Cham.
- Marasco, A., De Martino, M., Magnotti, F., & Morvillo, A. (2018). Collaborative innovation in tourism and hospitality: a systematic review of the literature. *International Journal of Contemporary Hospitality Management*, 30(6), 2364-2395. doi:10.1108/IJCHM-01-2018-0043
- Navío-Marco, J., Ruiz-Gómez, L. M., & Sevilla-Sevilla, C. (2018). Progress in information technology and tourism management: 30 years on and 20 years after the internet -Revisiting Buhalis & Law's landmark study about eTourism. *Tourism Management*, 69, 460-470. doi:10.1016/j.tourman.2018.06.002
- Orfila-Sintes, F., & Mattsson, J. (2009). Innovation behavior in the hotel industry. *Omega*, 37(2), 380-394. doi:10.1016/j.omega.2007.04.002
- Pikkemaat, B., Peters, M., & Chan, C.-S. (2018). Needs, drivers and barriers of innovation: The case of an alpine community-model destination. *Tourism Management Perspectives*, 25, 53-63. doi:10.1016/j.tmp.2017.11.004



- Romão, J., & Neuts, B. (2017). Territorial capital, smart tourism specialization and sustainable regional development: Experiences from Europe. *Habitat International*, 68, 64-74. doi:10.1016/j.habitatint.2017.04.006
- Rossetti, M., Stella, F., Cao, L., & Zanker, M. (2015). *Analysing User Reviews in Tourism with Topic Models*, Cham.
- Salomo, S., Talke, K., & Strecker, N. (2008). Innovation field orientation and its effect on innovativeness and firm performance. *Journal of Product Innovation Management*, 25(6), 560-576. doi:10.1111/j.1540-5885.2008.00322.x
- Souto, J. E. (2015). Business model innovation and business concept innovation as the context of incremental innovation and radical innovation. *Tourism Management*, *51*, 142-155. doi:10.1016/j.tourman.2015.05.017
- Trunfio, M., & Campana, S. (2019). Drivers and emerging innovations in knowledge-based destinations: Towards a research agenda. *Journal of Destination Marketing & Management*, 14, 100370. doi:10.1016/j.jdmm.2019.100370
- Venkatesh, V., Morris, M. G., Gordon, B. D., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478. Retrieved from http://www.jstor.org/stable/30036540