SDU
University of Southern Denmark

Identifying Innovative Idea Proposals with Topic Models

A Case Study from SPA Tourism

Sottocornola, Gabriele; Stella, Fabio; Symeonidis, Panagiotis; Zanker, Markus; Krajger, Ines; Faullant, Rita; Schwarz, Erich

Published in:
Big Data and Innovation in Tourism, Travel, and Hospitality

DOI:
10.1007/978-981-13-6339-9_8

Publication date:
2019

Document version:
Accepted manuscript

Document license:
Other

Citation for pulished version (APA):
Sottocornola, G., Stella, F., Symeonidis, P., Zanker, M., Krajger, I., Faullant, R., & Schwarz, E. (2019). Identifying Innovative Idea Proposals with Topic Models: A Case Study from SPA Tourism. In M. Sigala, R. Rahimi, & M. Thelwall (Eds.), *Big Data and Innovation in Tourism, Travel, and Hospitality: Managerial Approaches, Techniques, and Applications* (pp. 115-133). Springer. https://doi.org/10.1007/978-981-13-6339-9_8

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 22. May. 2022

This is a preprint, the published version can be accessed at
https://link.springer.com/chapter/10.1007/978-981-13-6339-9_8

Identifying Innovative Idea Proposals with Topic Models - A Case Study from SPA Tourism*

Gabriele Sottocornola, Fabio Stella, Panagiotis Symeonidis, Markus Zanker, Ines Krajger, Rita Faullant, and Erich Schwarz

Free University of Bozen-Bolzano, Italy
Gabriele.Sottocornola@stud-inf.unibz.it
{panagiotis.symeonidis,markus.zanker}@unibz.it
University of Milano-Bicocca
fabio.stella@unimib.it
Alpen-Adria-Universitaet Klagenfurt
{ines.krajger,erich.schwarz}@aau.at
Syddansk Universitet
ritaf@sam.sdu.dk

Abstract. This chapter builds on a dataset where online users of a spa platform participated in an online idea contest providing free-text descriptions of their proposals for spa services. A panel of domain experts annotated these idea descriptions with a score for their innovativeness that serves as ground truth for machine learning experiments. Thus, the contribution lies in the application of topic modeling techniques to free-text idea descriptions in order to automatically identify innovative proposals based on advanced text processing and machine learning. Results of this case study indicate that topic modeling can outperform the ZeroR baseline as well as traditional survey scales for lead user identification and therefore constitute a first step towards exploring this technique for innovation research.

Keywords: Spa Tourism, Topic Modeling, Online Idea Contest

1 Introduction

For services in general, and in particular for tourism offers, the customer has a crucial and integral role. Services are dependent upon the interaction quality between service provider and customer (Bolton & Drew, 1991; Grönroos, 1993; Parasuraman, Zeithaml, & Berry, 1988). Despite the acknowledgement of the customers' central role for tourism, it is surprising that customers are rarely integrated into new service development for tourism. In contrast, customer integration into new product development established on a systematic basis in the early 1980s for physical goods (Urban & Von Hippel, 1988; Von Hippel, 1978). In particular, the most demanding and advanced customers, so called

* This chapter extends the work published in (Faullant, Krajger, Zanker, et al., 2012)

lead users (Von Hippel, 1986), have been integrated into new product development. Several studies suggest that the integration of users into new product development is an appropriate means for companies to come up with faster and more customer-centered innovations, e.g. (Gruner & Homburg, 2000). Also for the service industry the potential for various stages and modes of user involvement has been highlighted (Alam, 2002). With our study we test whether the concept of lead-user integration is in principle also applicable for new service development in tourism. Users of an online platform focusing on spa vacations have been invited to participate in a virtual idea competition and to submit their ideas for new service offerings in spa tourism. In addition, participants' lead user characteristics were collected via a web-based questionnaire. Based on creativity theory an independent jury of spa experts evaluated these ideas, which were then correlated with the users' lead user characteristics.

This chapter extends the work presented in (Faullant et al., 2012) to automatically identify innovative ideas from users. Our proposed method consists of applying topic modeling, namely Latent Dirichlet Allocation (LDA) (Blei, 2012), to represent submitted ideas with their underlying topics. A topic can be seen as a sorted list of words that share a coherent semantic meaning and that are automatically extracted from corpora of natural language documents in an unsupervised manner. A panel of judges that scored each submitted proposal to an idea contest serve as ground truth. The hypothesis of this work is that the topical representation of each document can be effective to discern which formulated ideas (and thus also which participants) are innovative in the context of this study with respect to spa tourism.

In order to classify which of the participants' submissions better fit the innovation and customer's value expected by the judges we trained a tree-based classification model. The main advantages are that, once the model is trained, we can automatically have a degree of confidence on the innovation of a submitted idea together with an indication of the most important features, i.e. the topics, for the classification model. Since LDA provides an effective way to represent topics through their most representative words, we can try to explain which topics, and thus terms, can support the classification of an idea as innovative by the panel of judges.

The case study addresses the following research questions:

- Is LDA an effective technique to identify innovative traits or features within textual descriptions of ideas?
- Are tree-based classifiers together with LDA suitable to understand why a submitted idea is classified as innovative or not?

The rest of the paper is organized as follows: Section 2 presents the literature review of the field of lead users integration in new products development, together with the methodologies involved in this case study and some related works in the field of e-tourism; Section 3 describes the experiments conducted

on the specific dataset of spa tourism; Section 4 presents the results and finally conclusions are drawn.

2 Related Works

This chapter describes the application of text mining and machine learning techniques for the identification of opportunities for new product and service development. Thus, the related work covers all these aforementioned aspects. In addition, the subsections on LDA and classification provide a tutorial-like introduction.

2.1 Lead Users in new product development

Within the literature on innovation management, user integration into new product development has become an important research field. Instead of solely considering users as information providers, users can actively engage in the new product development process (Edvardsson, Kristensson, Magnusson, & Sundström, 2012; Von Hippel, 2005; Campos, Mendes, Valle, & Scott, 2018). Previous research confirms the ability of users to contribute to the NPD (new product development) process (Alam, 2006; Füller, Jawecki, & Mühlbacher, 2007; Lilien, Morrison, Searls, Sonnack, & Hippel, 2002; Oliveira & von Hippel, 2011; Skiba & Herstatt, 2009). However, only a small proportion of users, between 10 and 40 percent, has the know-how, creativity and expertise for truly innovative problem solutions that are not restricted to simple extensions or incremental innovations (Von Hippel, 2005). Since the value of customer contributions in the development of new products and services varies significantly, it is crucial to carefully select the right users to be integrated into new product or service development (Enkel, Perez-Freije, & Gassmann, 2005; Gruner & Homburg, 2000; Wellner, 2015). One group of users that has been shown to be able to deliver highly innovative suggestions for new product development are lead users. Lead Users are different from other users because they (a) have needs that will become commonplace in a market before the bulk of the other users encounters them and (b) they expect to benefit significantly from obtaining a solution to those needs (Von Hippel, 1986). These characteristics are also known as the *"Ahead of Trend" (AT)* dimension and the *"High Expected Benefit" (HEB)* dimension respectively (Franke, Von Hippel, & Schreier, 2006). Products developed in cooperation with lead users are appreciated as highly innovative by firms (Franke et al., 2006; Lilien et al., 2002; Lüthje, 2000). The ability to bear innovative solutions is fundamentally linked to a person's individual creativity (Faullant, Schwarz, Kraiger, & Breitenecker, 2009). In psychology, creativity is generally defined as "the production of novel, useful ideas or problem solutions" (Amabile, Barsade, Mueller, & Staw, 2005). The first aspect emphasizes the originality or unexpectedness of an idea (Sternberg & Lubart, 1999). The second aspect stresses that an idea must be of value, or "appropriate (i.e., useful, adaptive concerning task constraints)" (Sternberg & Lubart, 1999) which is especially important for new product development. Many

studies have confirmed that lead users are able to produce both novel and useful ideas.

Initial lead user studies concentrated predominantly on the industrial goods markets (Franke & Von Hippel, 2003; Herstatt & Von Hippel, 1992; Lüthje, 2003; Morrison, Roberts, & Von Hippel, 2000; Olson & Bakke, 2001; Urban & Von Hippel, 1988). The identification of lead users is also promising for user integration in consumer mass markets such as kite surfing, extreme sporting equipment, technical diving, and kitchen appliances (Franke & Shah, 2003; Füller, Bartl, Ernst, & Mühlbacher, 2006; Füller et al., 2007; Lüthje, 2004; Lüthje, Herstatt, & Von Hippel, 2005; Schwarz, Faullant, Krajger, & Breitenecker, 2009). Within new service development, systematic lead user identification and their integration for service innovation has been widely neglected (Skiba & Herstatt, 2009). Edvardson et al. (Edvardsson et al., 2012) proposed a conceptual framework of methods of customer integration into new service development. The lead user method was characterized as being able to generate highly novel service solutions, but at the same time requiring high methodological competences. Recent empirical evidence confirms the potential of user innovation for the service sector (Oliveira & von Hippel, 2011). In tourism so far, little is known about lead user identification and their involvement in new service development.

2.2 Virtual user integration for new product and service development

The use of the Web allows companies to reach potential users world-wide for new product development (Füller & Hienerth, 2004; Sawhney, Verona, & Prandelli, 2005). This is accompanied by the development of new tools and methods for virtual user integration, e.g. (Dahan & Hauser, 2002; Dahan & Srinivasan, 2000; Franke & Piller, 2004; Füller et al., 2007; Jeppesen, 2005; Verona, Prandelli, & Sawhney, 2006). Web- based methods such as idea competitions, toolkits for user innovation, virtual worlds, virtual stock markets and virtual communities have already diffused in practice supporting collaborative new product development (Bullinger, Neyer, Rass, & Moeslein, 2010; Ebner, Leimeister, & Krcmar, 2009). For the service sector in general Sigala (Sigala, 2010) provided insights from the Starbucks community that virtual user communities are able to generate, shape, and co-create ideas for new service development. The shared interpretation of an idea throughout the community can lead to different cultural interpretations of what a new service might constitute. Another study within the mobile service industry also demonstrates the potential of a firm hosted virtual lead user community for new service development (Mahr & Lievens, 2012). In tourism, the potential of user communities for the development of new tourism products was recognized in the early 2000s (Y. Wang, Yu, & Fesenmaier, 2002). Recent studies confirm the importance of customer co-creation in travel services and its impact on customer satisfaction and expenditure level (Grissemann & Stokburger-Sauer, 2012). These findings advocate against findings from earlier studies that demonstrate that user activities in user communities and blogs are still limited to

information exchange, such as sharing and documenting travel experiences and ratings of tourism products, more recently an active role of users in co-creation of tourism value creation has been acknowledged (Dippelreiter et al., 2008; Waldhör & Rind, 2008; Yoo & Gretzel, 2008; Rihova, Buhalis, Moital, & Gouthro, 2015). Especially through the use of technology users can actively engage in shaping their tourism experience (Neuhofer, Buhalis, & Ladkin, 2014). Meanwhile a range of successful examples of user involvement and user co-creation in the hospitality and leisure sector have been demonstrated (Egger, Gula, & Walcher, 2016). With our study we investigate whether users have the potential to substantially contribute to new service development and whether those users can be identified by web-based means.

2.3 Latent Dirichlet Allocation and Topic Models

Topic models are a suite of algorithms that aim to discover the main themes, denoted as topics, that pervade a large and otherwise unstructured collection of natural language documents. Topic models are able to annotate and summarize this corpus with the thematic information provided by topics.
The main contribution to probabilistic topic modeling was provided by (Blei, 2012) with the introduction of *Latent Dirichlet Allocation (LDA)*, where the underlying idea is that each document contains a mixture of multiple topics. LDA can be easily described through its *generative process*, a simple probabilistic procedure by which documents can be ideally generated.
Consider the visual representation of the LDA generative process for a corpus of tourism-related reviews, represented in Figure 1. A topic is formally defined as a probability distribution over a fixed vocabulary (the corpus dictionary) and a review document is associated with a probability distribution over topics. Assuming that the distribution of the topics over words is given (boxes on the left), the generative process of a generic document d in the corpus consists of the following steps:

1. Randomly choose a distribution over topics for document d (histograms on the right);
2. For each word in the document d:
   (a) Randomly choose a topic z from the distribution over topics sampled at step 1 (coins);
   (b) Randomly choose a word w from the corresponding distribution over the vocabulary, given the sampled topic z (word assignment in the text).

More formally the process specifies the probability of sampling a specific word token w_i as follows:

$$P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j)$$

where $P(z_i = j)$ is the probability that j-th topic was assigned to the i-th word token, $P(w_i|z_i = j)$ is the conditional probability to extract i-th word token

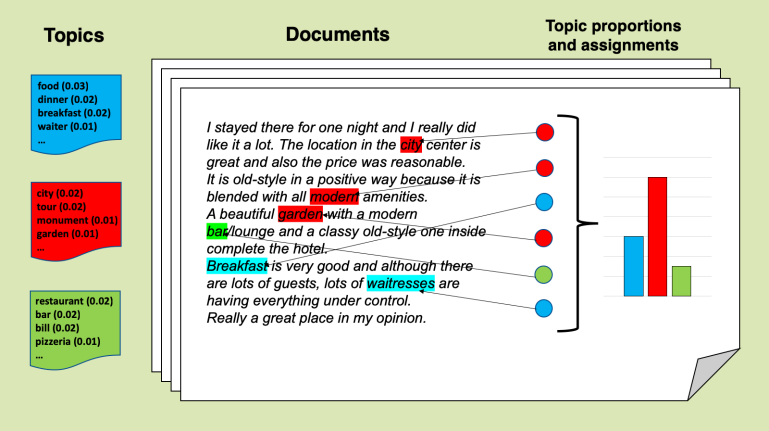

Fig. 1. A visual representation of the generative process underlying LDA.

given the assignment of topic j.

This statistical generative model reflects the intuition that documents contain multiple topics. Each document contains the topics in different proportion (step 1); each word in each document is drawn from one of the topics (step 2b), where the selected topic is chosen from the per-document distribution over topics (step 2a).
The aim of LDA model is to invert this generative process: the occurrences of words in the documents are the observed variables, while the topic structure (i.e. per-word topic distributions and per-document topic distributions) is hidden. By exploiting techniques of statistical inference and sampling (i.e. Gibbs sampling and variational bayesian inference), these probability distributions are inferred by observing the frequency of words within documents.

Some applications of topic models to e-tourism and recommendation systems in general have been described.

Rossetti et al. (Rossetti, Stella, & Zanker, 2016) provide a description of topic models with a particular focus on the tourism domain. They propose different application scenarios where the topic models effectively processes textual reviews in order to provide decision support and recommendations to online tourists as well as to build a basis for further analytics (i.e. provide additional semantics for explanation and understanding of the enormous amounts of user-generated data). Furthermore, the contribution consists of two new models based on LDA (namely, the *topic-criteria model* and the *topic-sentiment criteria model*) and results from experimenting with user-generated review data on restaurants and hotels.

Wang et al. (H. Wang, Lu, & Zhai, 2010) proposed a probabilistic generative model similar to LDA applied to textual reviews on hotels to estimate opinion

ratings on topical aspects (e.g. cleanliness, location or sleep quality), a problem defined as *Latent Aspect Rating Analysis (LARA)*. The underlying assumption is that each sentence in a review is related to a specific aspect. The proposed generative model assumes that for each sentence a user decides which aspect she wants to write about and chooses the words to write accordingly. To assign one or more aspects to each sentence a bootstrap procedure is defined in order to provide keyword sentences to different aspects. The method is able to distinguish between cases where the overall ratings are the same but aspect ratings are different. Furthermore, review analysis opens a range of possible applications, such as opinion summarization on topical aspects, ranking of entities based on these aspect ratings and the analysis of the rating behavior of reviewers.

Agarwal et al. (Agarwal & Chen, 2010) introduced a matrix factorization method for recommender systems where items have a natural bag-of-word representation named *fLDA*. The method works by regularizing both user and item factors simultaneously through user features and the bag of words associated with each item. In particular, each word in an item is associated with a discrete latent factor (i.e. the topic of the word); item topics are obtained by averaging topics across all words in an item. Then, a user rating on an item is modeled as user's affinity to the item's topics where users' affinity to topics (i.e. user factors) and topic assignments to words in items (i.e. item factors) are learned jointly in a supervised fashion. Topics extracted from item descriptions and user metadata are exploited as priors to regularize item and user latent factors. The posterior distribution of item and user factors depends on both the prior and user ratings on items, since the LDA model is exploited to regularize item latent factors, and the Gaussian linear regression regularizes user latent factors. The model has been proven to be accurate and capable to deal with warm-start and cold-start scenarios, as textual data related to new users and new items can be used to compute recommendations. Furthermore, it provides interpretable latent factors that can explain user-item interactions.

McAuley et al. (McAuley & Leskovec, 2013) aimed to combine latent rating dimensions (i.e. latent-factor recommender systems) with latent review topics (i.e. LDA topic models) in order to estimate the ratings from textual reviews on different datasets. The *Hidden Factors as Topics (HFT)* approach consists of two steps: first, latent factors for rating prediction are fitted, and second, topic assignments to item reviews are updated merging item-topic distributions to its latent factors. The proposed approach not only leads to more accurate predictions on recommendations, but can also solve side problems. First, it deals with the cold-start problem, exploiting content topics for items with only a few ratings. Second, it is able to discover and automatically categorize items in different categories based on the topics discussed in the reviews. Third, it is able to identify representative reviews, which can be shown to users as an explanation of item characteristics.

Another extension of LDA applied to user reviews is the *Joint Sentiment-Topic model (JST)* (Lin & He, 2009). In contrast to the majority of sentiment analysis models which are based on classification models, this model is able

to extract sentiment and topics simultaneously from text in an unsupervised way. The main difference with respect to the LDA model is that JST adds an additional sentiment layer between the document and the topic layer. In this way a four level hierarchy is defined where documents have distributions on sentiment labels, sentiment labels have distributions on topics and topics have distributions on words. The model has been evaluated on the movie review dataset to classify the review sentiment polarity to further improve the sentiment classification accuracy.

2.4 Decision Tree based Classification Models

Decision tree classification models (also referred to as decision tree learning) are a family of predictive modeling approaches used for supervised classification in the field of machine learning (Rokach & Maimon, 2014). A decision tree is used as a predictive model to go from observations of the features of an item (represented in the branches) to conclusions about the item's target class (represented by the leaves). Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

A decision tree can be learned by splitting the data into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the whole subset belongs to the same target class, or when splitting no longer adds value for the classification. Decision tree learning is the construction of a decision tree from class-labeled training tuples. A decision tree is a flow-chart-like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node. Many specific learning algorithms were proposed in the literature throughout the years, and the most relevant are: *ID3*, *C4.5*, and *CART*.

Algorithms for constructing decision trees usually work top-down, by choosing a variable at each step that best splits the set of items. Different algorithms use different metrics for measuring the best split. These generally measure the homogeneity of the target variable within the subsets. These metrics are applied to each candidate subset, and the resulting values are combined (e.g. averaged) to provide a measure of the quality of the split. The most used metrics to determine the quality of a split are *Gini impurity* and *information gain*.

A particular type of decision tree classifier that had lot of success in the machine learning community is the random forests classifier. *Random forests* were first introduced by Breiman et al. (Breiman, 2001), which describes a method of building a forest of uncorrelated trees, combined with randomized node optimization and bagging. Random forests are an ensemble learning method for classification, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. The main advantage of random forests over decision trees is the capability

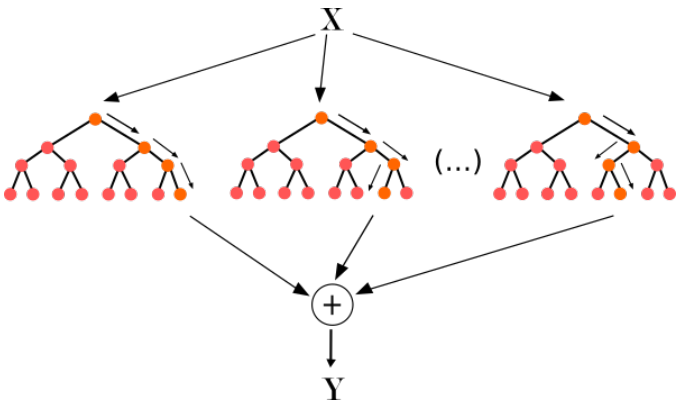

Fig. 2. A graphical representation of a Random forest.

of avoiding overfitting to the training set, thanks to the bagging (or bootstrap aggregating) procedure. Figure 2 provides a visual example of a random forest structure.

3 Methodology

In this section we describe the experimental framework to evaluate the effectiveness of the LDA methodology to identify innovative users with a case study from the spa domain. We first illustrate the process of data collection and the dataset characteristics. Then we describe in details the preprocessing process and the experiments performed on data. Finally, we present and discuss the results.

3.1 Dataset

For evaluating the proposed approach the dataset presented in (Faullant et al., 2012) was used. The data was collected through an empirical study executed in the field of a search engine to retrieve curated information on European spas. Visitors of the website were invited to submit innovative ideas for spa development and new service creation. After describing the details of their idea in a free text field, participants were asked to complete a standardized web-based questionnaire. To assess participants' for being a lead user, the two hallmark characteristics of being *ahead of trend* (AT) and *high expected benefit* (HEB) were used from existing measures in the literature (Lüthje, 2000). The scales were adapted to the spa context and were measured as a continuous variable on a 7-point Likert scale as denoted in Table 1. Furthermore, demographic data like gender, age, nationality and education as well as information about the actual spa usage of participants were collected.

To evaluate and rank the submitted ideas the *Consensual Assessment Technique* (CAT) (Amabile, 1982) was applied. According to this method "a product

Table 1. Codes of questionnaire items.

Code	Scale
AT1	I'm regarded as being well informed in the field of spa offers.
AT2	I usually determine new spa offers earlier than most other people.
AT3	I try to visit just recently opened spas.
HEB1	I have needs and preferences which are not satisfied by spa offers.
HEB2	During my past visits of spa resorts I noticed shortcomings several times.
HEB3	I'm dissatisfied with the existing spa resort offers.
SPA1	How often do you visit spa resorts each year?
SPA2	How many different spa resorts have you visited up to now?
SPA3	Which of the following eight recently opened spa resorts have you visited?

or response is creative to the extent that appropriate observers independently agree it is creative." (Amabile, 1982). The quality of the submitted ideas was independently evaluated by a jury of 4 experts in the spa domain, based on 3 dimensions: (i) originality of the idea (*orig*), (ii) customer value of the idea (*util*), (iii) overall impression (*over*). These dimensions were presented on a scale from 0 (no value) to 5 (very high value). The experts rated all ideas independently from each other.

In total 161 participants filled out the questionnaire, and submitted 122 ideas or suggestions for spa service development. The dataset contains only users which have submitted an idea (more than 1 term) and which have completed the questionnaire. For 6 users, a missing answer in the questionnaire was replaced with the average value for that question. Finally, after data cleaning the dataset resulted in 116 instances.

Table 2. Statistics on questionnaire items (Cronbach alpha = 0.74).

Item	Mean	StdDev
AT1	4.03	1.87
AT2	4.4	1.89
AT3	4.36	1.78
HEB1	2.53	1.68
HEB2	4.44	1.69
HEB3	2.72	1.66

Descriptive statistics of replies to the adapted questionnaire on AT and HEB items (scale 1-7, 1: I do not agree at all, 7: I fully agree) are given in Table 2. Before preprocessing the textual data, submitted ideas have an average length of 555 characters (standard deviation: 1411.5, min: 17, max: 14501) and Figure

3 depicts their length distribution with one extremely long outlier. The most dominant themes for innovation are the different needs of adults and children that should be addressed in separate locations and ideas for designing the relaxation and recreation areas. On average respondents have 7.18 spa visits per year (SPA1), know 6.25 different spa resorts (SPA2) and have already tried 1.4 out of the 8 newly opened spa resorts that were named in the questionnaire (SPA3).

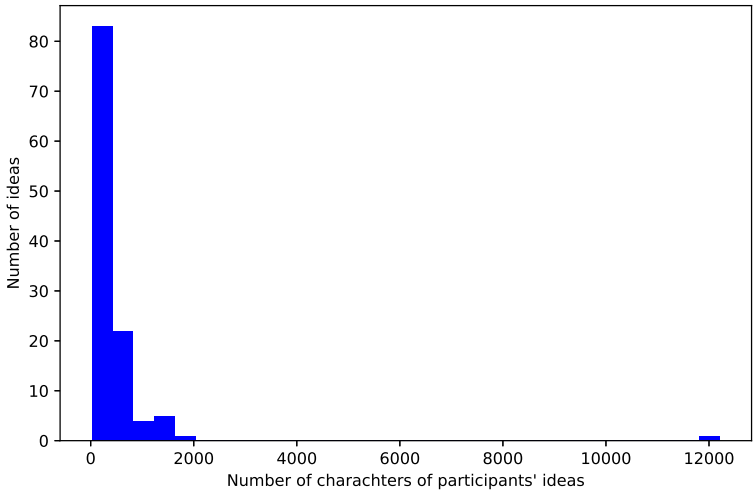

Fig. 3. Distribution of character length of the submitted ideas.

3.2 Preprocessing

We preprocessed the raw data before applying the topic model and the classification algorithm. First, we cast the problem to a binary classification problem, thus, for each one of the previously defined dimensions we compute the majority vote on the judges ratings. Formally, given the idea submitted by user i and a quality dimension d, the class $C_i^d = 1$ if at least 3 out of 4 judges have rated the idea greater or equal to 3 for the dimension d, otherwise $C_i^d = 0$.

We applied transformation to the corpus of users' submitted ideas. Textual content was tokenized, considering non-alphanumeric characters as separators, and all the extracted tokens were converted to lowercase. Furthermore, tokens with less than 3 characters were filtered out together with common German stopwords. Stemming wasn't applied to the tokens to avoid merging term roots with different meanings. The corpus is in German and the dictionary after this

process contains 2449 unique German terms.

The class label for our dataset is judges' overall impression (*over*), since it summarizes the other 2 dimensions. In Table 3 we report some statistics about the spa dataset with respect to positive and negative instances of this dimension.

Table 3. Summary of the statistics related to the positive and negative classes for the *over* dimension, i.e. number of users (*users*), total number of characters (*chars*), total number of terms (*terms*), and number of unique terms (*unique*) in the submitted ideas.

	users	*chars*	*terms*	*unique*
positive	33	24699	3422	1752
negative	83	14625	1984	1100
total	116	39324	5406	2449

Note the following observations from Table 3: (a) The number of instances in our dataset is relatively small, thus, we have to expect that a machine learning approach should perform poorly in terms of accuracy and confidence of the predictions. (b) The distribution of the two classes is unbalanced, with more than 70% of the observations being negative. Thus, it should be more difficult for a classifier to get good predictive performance on the minority class (i.e., the positive one). (c) Despite the fact that the majority of the observations are negative, we can see how the number of characters is almost doubled for positive ideas (average length 748 characters) with respect to the negative ones (average length 176 characters). The same observations can be drawn from the counts of the number of terms, i.e. the average number of terms for a positive idea is 104, while the average number of terms for a negative one is 24. (d) Finally, despite the big difference in the length of the ideas between the two classes, there's less difference in the number of unique terms.

3.3 Experimental Setup

We designed an evaluation protocol to cope with the very low number of instances to train and test our classification approach. First, we applied a 5-fold cross validation (CV). The data were sampled with a stratified sampling technique to maintain a similar distribution of the target class in every folder. At each CV loop, we learned different LDA topic models on the users' submitted ideas that belong to the training data. We trained LDA with number of topics $T = 3, 5, 7, 10, 13, 15, 20, 30$. To compute these models, we used the Python wrapper for the well-known NLP library MALLET[1]. The library implements an effective and efficient version of LDA, based on Gibbs sampling. We set the number of learning iterations to 1,000 with hyperparameter optimization every

[1] McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." `http://mallet.cs.umass.edu.2002.`

10 iterations (after 200 iterations of burn-in). The trained models are further exploited to infer the doc-topic distributions (i.e., the proportion of the topics hidden in each document) for the test corpus of ideas. We held-out test data from the training of the LDA model, in order to not overfit the model to that corpus, even with an unsupervised technique. After this process we came out with a new set of features (namely, the doc-topic distribution for each textual idea) for each trained topic model. These features were further exploited for the classification task.

For sake of clarity we wanted to narrow down the number of candidate topic models to learn. In literature several methods are provided, e.g. based on perplexity of hold-out documents (Wallach, Murray, Salakhutdinov, & Mimno, 2009) or coherence scores related to pointwise mutual information (Newman, Lau, Grieser, & Baldwin, 2010). Since our dataset is too small, we didn't have enough data to provide a significant generalization on this kind of unsupervised measures. Thus, we conducted a preliminary analysis on the whole set of topic models: we trained a separate 10-fold cross validation and we selected the optimal models, i.e. the ones which performed well in terms of accuracy on the prediction task. At the end of this process we manually set the topic models with $T = 5, 10, 15$.

We used each feature set generated by the LDA models to separately train a supervised classification model to predict the target class of interest. At every CV loop, the performances of all the models were evaluated on the same test folder, never seen by the training algorithm, in terms of accuracy, and precision-recall on the positive class. We decided to apply a tree-based classifier, for the advantage of automatically assessing the importance of each feature, based on the computed Gini index (Breiman, 2001). This characteristic provides a sort of explanation on which are the specific topics (thus, the most representative words for that topics) mainly contribute to the classification accuracy, to discern between innovative and not innovative ideas.

For our experiments we used the Random Forest classifier provided by Python library scikit-learn[2]. We trained the algorithm with 10 tree estimators with Gini index as a splitting criterion and all the features considered at each splitting point. We employed the same algorithm on the same training data points to build different baseline models to benchmark the classification performance of the topic models. Specifically, we trained two different kind of models: one on the features related to the self-reported characteristics of users and another on the features related to the length of the submitted idea (i.e. the number of words and the number of characters). Furthermore, we merged the two sets of features with the ones extracted by the topic models in order to train hybridized models that are able to capture the information provided by different sources.

[2] `http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.breiman:2001Classifier.html`

4 Results and Discussion

In this section we present and discuss the results achieved by our topic model classification approach on the *over* dimension. The results are presented together with three baselines: the classification model trained on the features related to the text length (i.e. *Length*), the classification model trained on the features extracted from the questionnaire (i.e., *Survey*), and the *ZeroR* classification model (i.e. the most trivial predictor that classifies every instance with the most frequent class).
Hence we present the results of the experiments related to the classification performance of the hybrid models, namely, the ones that are built upon the features of each candidate topic model merged together with the features of the two baseline models (i.e., length-related features and survey-related features).
Finally we provide to the reader an insight into the effectiveness of explaining the predictions provided by the proposed technique.

For each trained model we evaluate its performance with different classification metrics, directly derived from the confusion matrix. The confusion matrix is a table with two rows and two columns, that reports the number of *true positives* (TP, i.e. the number of instances that belong to the positive class predicted as positive), *false positive* (FP, i.e. the number of instances that belong to the positive class predicted as negative), *true negatives* (TN, i.e. the number of instances that belong to the negative class predicted as negative) and *false negatives* (FN, i.e. the number of instances that belong to the negative class predicted as positive), collected after the evaluation of a binary classifier. With these values we are able to compute the following metrics, that well describe the effectiveness of our classification algorithms:

- *Precision*, defined as $TP/(TP + FP)$.
- *Recall*, defined as $TP/(TP + FN)$.
- *Accuracy*, defined as $TP + TN/(TP + FP + TN + FN)$.

Table 4. Classification performances of the topic models (i.e. $T = 5$, $T = 10$, $T = 15$) and the baselines (*Length*, *Survey*, *ZeroR*) for the dimension $d = over$.

	T=5	**T=10**	**T=15**	**Length**	**Survey**	**ZeroR**
precision	0.533	0.533	0.4	0.564	0.04	0.0
recall	0.186	0.152	0.1	**0.448**	0.029	0.0
accuracy	0.734	0.742	0.743	0.748	0.614	0.716

In Table 4 we report the results achieved by the classifiers built upon the candidate topic models (i.e. $T = 5$, $T = 10$, $T = 15$) and the baseline methods (*Length*, *Survey*, *ZeroR*) for the dimension $d = over$.

First, increasing the number of topics does not automatically lead to better performances; the best performing topic models have 5 and 10 topics. While the accuracy is almost the same (with a maximum variation of 1%) for the three models, recall and precision on the positive class deviate significantly from the two best models to the one with 15 topics, i.e. there's a 13% improvement in the precision and more than 5% improvement in the recall. Again, we can see how each of the topic-related classification model strongly outperforms at least two of the baselines in every metric. In particular the *Survey* baseline, i.e. the self-reported characteristics w.r.t. the innovativeness, behave really poorly with respect to the other methods, confirming the findings in (Faullant et al., 2012). Furthermore, from the results it emerges that the classification model based only on the length of the submitted idea is able to closely outperform even the proposed topic model approach, especially for the recall metric (28% improvement). We take advantage of this finding and therefore explore the performance of our approach when we hybridize the features extracted by the LDA with the other two kinds of features, *Survey* and *length*.

Table 5. Classification performances of the hybrid models that combine the features of the candidate topic models (i.e. $T = 5$, $T = 10$, $T = 15$) with the baselines' features of Length (L) and Survey (S) for the dimension $d = over$.

	T=5+L	**T=5+S**	**T=10+L**	**T=10+S**	**T=15+L**	**T=15+S**
precision	0.603	0.267	**0.636**	0.55	0.587	0.6
recall	0.391	0.114	0.423	0.162	0.338	0.095
accuracy	0.759	0.706	**0.767**	0.752	0.76	0.742

In Table 5 we report the results achieved on the classification task by the hybrid models, constructed from the union of the features of the candidate topic models (i.e. $T = 5$, $T = 10$, $T = 15$) with the baselines' features of Length (L) and Survey (S) for the dimension $d = over$.
The results clearly highlight that the hybridization process improves the performance of each topic model configuration, in particular *Length* features in combinations with the topics are able to achieve optimal results. Each hybrid model built with length-based features outperform the results achieved by the topic models and the *Length* model in precision and accuracy. A significant case is represented by the 10-topics model, in which the combination with *Length* allows to achieve the best overall performance for all metrics (except that recall is still slightly outperformed by the simple length-based method). Even the *Survey* features in combination with the 10 topics are able to improve the model with respect to the separated ones. Finally, the noisy *Survey* features overwhelm the model with less topics, $T = 5$, with a significant decrease in accuracy and precision.

Finally, we provide an insight of the advantage of the proposed approach with respect to the *explainability* of the classification outcome. As stated above in this Section, a tree-based classifier is naturally able to rank features with respect to their importance in the classification process (i.e., the importance of a feature is computed as the Gini importance, which roughly represents the percentage of instances that a particular feature contributes to correct classification). Furthermore, each feature in our topic model approach can be represented as a ranked list of words, since LDA automatically extract a set of topics, which are a probability distribution over the vocabulary. The combination of the two models can provide an hint, that is easy to understand, on which are the clusters of words (i.e., the topics) that most effectively distinguish between innovative and not innovative ideas.
In Table 6 we report some illustrative examples of how this technique can positively impact our case study, providing a further explanation on the predictions. In particular, we give 4 examples, taken from the trained 5-topics models, of the most representative words for the most important features.

Table 6. Some examples of the explanation technique provided by our approach: the feature importance in the classification ($FeatureRel$) is related to a topic, which has a certain predominance in the corpus ($TopicRel$) and is represented by a ranked list of words ($TopWords$).

FeatureRel	*TopicRel*	*TopWords*
0.32	0.345	erding restaurants wasser ruheräume bad hotel moderne aufguss saunen hundebesitzer
0.24	0.362	außenbereich wasser erding restaurants ruheräume sauna warmes wellnessbereich achten erlebnisaufgüsse
0.41	0.016	thema vorstellen http www besucher show ideen watch youtube freizeitparks
0.42	0.043	man thermen idee vorstellen http besucher www thema show youtube

The examples show that there's not a clear correlation between the importance of a feature, as provided by the tree-based classifier ($FeatureRel$) and the relevance of a topic in a corpus ($TopicRel$, i.e. the percentage of documents that belong to it). The predominance of a topic in the corpus is not enough to indicate its predictive power. Furthermore, the relevant topics (automatically extracted from the corpus of ideas in a unsupervised manner) are the ones that are represented by significant words ($TopWords$), related to a particular type of room or to activity, e.g. ruheräume (relax room), außenbereich (outdoor activities), aufguss (infusions), hundebesitzer (dog owner), wellnessbereich (wellness area), restaurants, freizeitparks (leisure park), or related to particular video content

the users want to share with the judges (as in the last two examples presented).

5 Conclusions

This case study explored the application of LDA techniques for identifying innovative users based on free-text submission to an online idea contest. The contribution of this chapter lies in presenting the application of this technique in combination with learning classifiers to a wider audience of tourism and innovation researchers. Although sample size is a limitation for this case-study, the achieved results demonstrate that the prediction accuracy of traditional surveys for lead user identification as well as a ZeroR baseline can be surpassed. Furthermore, results can be even slightly improved if the feature space of identified topics is extended by considering also ideas' length or responses to the survey for lead user identification.

References

Agarwal, D., & Chen, B.-C. (2010). flda: matrix factorization through latent dirichlet allocation. In *Proceedings of the third acm international conference on web search and data mining* (pp. 91–100).

Alam, I. (2002). An exploratory investigation of user involvement in new service development. *Journal of the Academy of Marketing Science*, *30*(3), 250.

Alam, I. (2006). Removing the fuzziness from the fuzzy front-end of service innovations through customer interactions. *Industrial marketing management*, *35*(4), 468–480.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology*, *43*(5), 997.

Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative science quarterly*, *50*(3), 367–403.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Bolton, R. N., & Drew, J. H. (1991). A multistage model of customers' assessments of service quality and value. *Journal of consumer research*, *17*(4), 375–384.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Bullinger, A. C., Neyer, A.-K., Rass, M., & Moeslein, K. M. (2010). Community-based innovation contests: Where competition meets cooperation. *Creativity and innovation management*, *19*(3), 290–303.

Campos, A. C., Mendes, J., Valle, P. O. d., & Scott, N. (2018). Co-creation of tourist experiences: A literature review. *Current Issues in Tourism*, *21*(4), 369–400.

Dahan, E., & Hauser, J. R. (2002). The virtual customer. *Journal of Product Innovation Management: AN INTERNATIONAL PUBLICATION*

OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION, *19*(5), 332–353.

Dahan, E., & Srinivasan, V. (2000). The predictive power of internet-based product concept testing using visual depiction and animation. *Journal of product innovation management*, *17*(2), 99–109.

Dippelreiter, B., Grün, C., Pöttler, M., Seidel, I., Berger, H., Dittenbach, M., & Pesenhofer, A. (2008). Online tourism communities on the path to web 2.0: an evaluation. *Information technology & tourism*, *10*(4), 329–353.

Ebner, W., Leimeister, J. M., & Krcmar, H. (2009). Community engineering for innovations: the ideas competition as a method to nurture a virtual community for innovations. *R&d Management*, *39*(4), 342–356.

Edvardsson, B., Kristensson, P., Magnusson, P., & Sundström, E. (2012). Customer integration within service developmenta review of methods and an analysis of insitu and exsitu contributions. *Technovation*, *32*(7-8), 419–429.

Egger, R., Gula, I., & Walcher, D. (2016). Towards a holistic framework of open tourism. In *Open tourism* (pp. 3–16). Springer.

Enkel, E., Perez-Freije, J., & Gassmann, O. (2005). Minimizing market risks through customer integration in new product development: learning from bad practice. *Creativity and Innovation Management*, *14*(4), 425–437.

Faullant, R., Krajger, I., Zanker, M., et al. (2012). *Identification of innovative users for new service development in tourism.* Citeseer.

Faullant, R., Schwarz, E. J., Kraiger, I., & Breitenecker, R. J. (2009). Towards a comprehensive understanding of lead userness: the role of personality and creativity. In *16th international product development management conference.*

Franke, N., & Piller, F. (2004). Value creation by toolkits for user innovation and design: The case of the watch market. *Journal of product innovation management*, *21*(6), 401–415.

Franke, N., & Shah, S. (2003). How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research policy*, *32*(1), 157–178.

Franke, N., & Von Hippel, E. (2003). Satisfying heterogeneous user needs via innovation toolkits: the case of apache security software. *Research policy*, *32*(7), 1199–1215.

Franke, N., Von Hippel, E., & Schreier, M. (2006). Finding commercially attractive user innovations: A test of lead-user theory. *Journal of product innovation management*, *23*(4), 301–315.

Füller, J., Bartl, M., Ernst, H., & Mühlbacher, H. (2006). Community based innovation: How to integrate members of virtual communities into new product development. *Electronic Commerce Research*, *6*(1), 57–73.

Füller, J., & Hienerth, C. (2004). Engaging the creative consumer. In *European business forum* (Vol. 19, pp. 54–57).

Füller, J., Jawecki, G., & Mühlbacher, H. (2007). Innovation creation by online basketball communities. *Journal of business research*, *60*(1), 60–71.

Grissemann, U. S., & Stokburger-Sauer, N. E. (2012). Customer co-creation of travel services: The role of company support and customer satisfaction with the co-creation performance. *Tourism Management*, *33*(6), 1483–1492.

Grönroos, C. (1993). Toward a third phase in service quality research: challenges and future directions. *Advances in services Marketing and Management*, *2*(1), 49–64.

Gruner, K. E., & Homburg, C. (2000). Does customer interaction enhance new product success? *Journal of business research*, *49*(1), 1–14.

Herstatt, C., & Von Hippel, E. (1992). From experience: Developing new product concepts via the lead user method: A case study in a low-tech field. *Journal of Product Innovation Management: AN INTERNATIONAL PUBLICATION OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION*, *9*(3), 213–221.

Jeppesen, L. B. (2005). User toolkits for innovation: Consumers support each other. *Journal of product innovation management*, *22*(4), 347–362.

Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., & Hippel, E. v. (2002). Performance assessment of the lead user idea-generation process for new product development. *Management science*, *48*(8), 1042–1059.

Lin, C., & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th acm conference on information and knowledge management* (pp. 375–384).

Lüthje, C. (2000). Kundenorientierung im innovationsprozess. *Eine Untersuchung der Kunden-Hersteller-Interaktion in Konsumgütermärkten, Wiesbaden*.

Lüthje, C. (2003). Customers as co-inventors: An empirical analysis of the antecedents of customer-driven innovations in the field of medical equipment. In *Proceedings of the 32th emac conference, glasgow*.

Lüthje, C. (2004). Characteristics of innovating users in a consumer goods field: An empirical study of sport-related product consumers. *Technovation*, *24*(9), 683–695.

Lüthje, C., Herstatt, C., & Von Hippel, E. (2005). User-innovators and local information: The case of mountain biking. *Research policy*, *34*(6), 951–965.

Mahr, D., & Lievens, A. (2012). Virtual lead user communities: Drivers of knowledge creation for innovation. *Research policy*, *41*(1), 167–177.

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th acm conference on recommender systems* (pp. 165–172).

Morrison, P. D., Roberts, J. H., & Von Hippel, E. (2000). Determinants of user innovation and innovation sharing in a local market. *Management science*, *46*(12), 1513–1527.

Neuhofer, B., Buhalis, D., & Ladkin, A. (2014). A typology of technology-enhanced tourism experiences. *International Journal of Tourism Research*, *16*(4), 340–350.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evalua-

tion of topic coherence. In *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics* (pp. 100–108). Stroudsburg, PA, USA: Association for Computational Linguistics.

Oliveira, P., & von Hippel, E. (2011). Users as service innovators: The case of banking services. *Research policy*, *40*(6), 806–818.

Olson, E. L., & Bakke, G. (2001). Implementing the lead user method in a high technology firm: A longitudinal study of intentions versus actions. *Journal of Product Innovation Management: AN INTERNATIONAL PUBLICATION OF THE PRODUCT DEVELOPMENT & MANAGEMENT ASSOCIATION*, *18*(6), 388–395.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual-a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12–40.

Rihova, I., Buhalis, D., Moital, M., & Gouthro, M.-B. (2015). Conceptualising customer-to-customer value co-creation in tourism. *International Journal of Tourism Research*, *17*(4), 356–363.

Rokach, L., & Maimon, O. Z. (2014). *Data mining with decision trees: theory and applications (2nd ed.)*. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

Rossetti, M., Stella, F., & Zanker, M. (2016). Analyzing user reviews in tourism with topic models. *Information Technology & Tourism*, *16*(1), 5–21.

Sawhney, M., Verona, G., & Prandelli, E. (2005). Collaborating to create: The internet as a platform for customer engagement in product innovation. *Journal of Interactive Marketing*, *19*(4), 4–17.

Schwarz, E., Faullant, R., Krajger, I., & Breitenecker, R. (2009). Are leading edge users creative. In *Annual meeting of the 38th emac conference.*

Sigala, M. (2010). Managing customer involvement and roles in new service development (nsd): lessons learnt from www. mystarbucksidea. com. 28th eurochrie conference amsterdam. *The Netherlands*.

Skiba, F., & Herstatt, C. (2009). Users as sources for radical service innovations: opportunities from collaboration with service lead users. *International Journal of Services Technology and Management*, *12*(3), 317–337.

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. *Handbook of creativity*, *1*, 3–15.

Urban, G. L., & Von Hippel, E. (1988). Lead user analyses for the development of new industrial products. *Management science*, *34*(5), 569–582.

Verona, G., Prandelli, E., & Sawhney, M. (2006). Innovation and virtual environments: Towards virtual knowledge brokers. *Organization Studies*, *27*(6), 765–788.

Von Hippel, E. (1978). Successful industrial products from customer ideas. *The Journal of Marketing*, 39–49.

Von Hippel, E. (1986). Lead users: a source of novel product concepts. *Management science*, *32*(7), 791–805.

Von Hippel, E. (2005). Democratizing innovation: The evolving phenomenon of

user innovation. *Journal für Betriebswirtschaft*, *55*(1), 63–78.

Waldhör, K., & Rind, A. (2008). etbloganalysismining virtual communities using statistical and linguistic methods for quality control in tourism. *Information and communication technologies in tourism 2008*, 453–462.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1105–1112).

Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th acm sigkdd international conference on knowledge discovery and data mining* (pp. 783–792).

Wang, Y., Yu, Q., & Fesenmaier, D. R. (2002). Defining the virtual tourist community: implications for tourism marketing. *Tourism management*, *23*(4), 407–417.

Wellner, K. (2015). *User innovators in the silver market: An empirical study among camping tourists*. Springer.

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? *Information Technology & Tourism*, *10*(4), 283–295.