Building confidence measures for tourist destination choice

Roberto Marmo
Department of Computer Engineering and Systems Science, Faculty of Engineering,
University of Pavia, Italy
Rodolfo Baggio
Master in Economics and Tourism,
Bocconi University, Milan, Italy

December 2016, forthcoming in International Journal of Tourism Sciences

Abstract
Reviews, social media, web sites, recommendations from trusted inner circle opinions are essential to tourists in the travel planning process, but the large amount of information they can collect increases the difficulty to choose a destination. Determining how improvements in data quality and the variety of touristic information sources affect the probability of a correct situational detection helps optimize the performance of the overall decision. This work presents a generalized method for estimating levels of individual source reliability and to assess the confidence level of a combined output. Initial reliability levels of touristic information sources are augmented by (i) combining the measures of multiple sources, (ii) incorporating the truth reinforcement of related elements, and (iii) incorporating the importance of the individual elements for determining the truth probability for the whole. The result is a measure of confidence in system output based on establishing links among the truth values of inputs.

Keywords
information fusion, confidence measures, destination choice, reputation

Introduction
The reputation of a tourist destination often depends on information and contents generated by travelers, suppliers, and residents. Recent research on the issues concerning information sources used for travel decisions (Jacobsen et al., 2012 ; Llodrà-Riera et al., 2015; Xiang et al., 2015), suggests that relevance of information sources, particularly those used in the pre-purchase hotel search process (Murphy et al., 2014), have not been fully explored.

Build confidence is a psychological state whereby actors make themselves vulnerable and ready to take risks based on the expectation of goodwill and positive decision. Accurate information for the selected tourism destination can be pulled from available online sources by using a percentage score representing the extent to which the reliability of an information source, in the wider scenario, matches the searched destination.

The purpose of this paper is to propose a mathematical method for providing a global measure of confidence based on a mixture of reliabilities from tourist and experts. A good fit with a low probability of accurately portraying the truth may not be as informative to the user as a lesser fit that is almost certainly a true representation of the world. Providing the confidence measure alongside the match percentage therefore improves the user's ability to make appropriate decisions.

System components
The method starts with known or estimated measures of the variety of sources' truth reliability and tracks their impact on final decision quality. Using varying levels of data independence it is possible to refine our confidence levels of individual elements. Incorporating the reinforcement effects of related

information provides a method of refining overall confidence levels. Not all elements are considered equal for determining the confidence level of a decision. Importance weights have also been accounted for in determining the combined reliability of the system output.

The confidence measure allows to compare final decision performance under varying conditions, to refine the mathematical model, and improve performance. Knowing which combinations of source types improves confidence the most can help optimize the effectiveness of data searches. Lacking a particular element, one can attempt to achieve a threshold confidence level by discovering enough related elements to compensate for the missing one. And by improving the confidence of the elements with the greatest importance, one can be assured that the improvement will have maximal impact.

The considered methods is explained in detail in Bramson (2004). It is based, mathematically, on a matrix product which is relatively easy to calculate and does not depend on specific data collection, extraction, fusion of information. At the end we have a unified measure of confidence incorporating: (i) truth overlap of multi-source elements, (ii) truth reinforcement from related elements, and (iii) the relative importance of each element in determining the whole.

Information collection and extraction

By browsing the Web (or by using offline sources) it is possible to determine the main sources of information for tourists' travel decisions. This is basically a sort of weight that the different sources employ in defining the overall construct. Here, as an example, we assume (although somehow arbitrarily) that information sources and their reliability are as shown in table 1. The values mainly come from a number of reports on the different tools that a traveler use for making a decision (see e.g. Jacobsen et al.,2012; Llodrà-Riera et al., 2015; Murphy et al., 2014; Xiang et al., 2015). The values reported here have the only purpose of showing the method. This case study is based on n=6 sources, it is possible to extend to wider number of sources.

Combining reliabilities from multiple sources

Looking up the truth overlaps for all possible pairwise combinations of sources, table 2 shows the reinforcement provided by multiple source for the same element. Each number c_{ij} (i is the row and j is the column) in table 2 is a user-defined individual reinforcement level (reinforcement direction is from i to j), the diagonal elements are arbitrarily set to one. The truth dependencies form a non-symmetric matrix because the temporal order in which the sources are combined is relevant. For example: a Facebook review followed by a review from a blog can create a different judgment compared to a blog review followed by a Facebook review.

The total reinforcement C as (elements i,j are those of table 2) is:

$$C = \frac{\sum_{i \neq j} c_{ij}}{n^2 - n}$$

The combined reliability X is determined by the following equation where r_i are provided by table 1:

$$X = \prod_{i=1}^{n} r_i + \left[\left(\frac{1}{n}\right) \sum_{i=1}^{n} r_i - \prod_{i=1}^{n} r_i \right] \cdot C$$

In order to incorporate the importance weights of the individual reliabilities of each element, this function needs individual values for the reliabilities, so it is necessary to split X down into its constituents:

$$x = \sqrt[n]{X}$$

$$r_k = x \cdot exp\left[\frac{1}{2}\left(1 + \left(1 - \frac{r_k}{\frac{1}{n}\sum_{k=1}^{n} r_k}\right) + \left(1 + \left(1 - \frac{n \cdot \sum_{i=1}^{n} c_{ik} \cdot r_i}{\sum_{j=1}^{n} \sum_{i=1}^{n} c_{ij} \cdot r_i}\right)\right)\right)\right] \text{ and } k=1,2,\ldots,n$$

The first addend redistributes the weight of the elements' original reliabilities, the second part redistributes the weight of how much an individual element's reliability was augmented by the reinforcement from surrounding elements.

Incorporating importance weights

The final decision is adjusted by the relative importance of information sources, because the importance of an element compared to other elements is an important aspect of combining reliability factors. Deciding how the levels of importance will affect the combined reliability is core to developing the appropriate combining function. There are several characteristics that can be identified as key to the problem, as described by Aaron (2004). Two type of weights are used:

- r_i reliability of each source i=1,2,…,n as opinion expressed by experts(could be seen as an "intrinsic" reliability); each weight assumes value in [0,1] as shown in table 1;
- a_i the importance weights of each source i=1,2,…,n expressed by user (tourist); each weight assumes value in [0,1], because the weights are relative importance weights, the values sum to 1.

These two constraints are sufficient to fully specify a function $R(r_i, a_i)$ for combining the reliabilities to have the final result as number in [0,1] where 0 denotes low reliability and 1 denotes high reliability. A simple function that combines the individual reliabilities incorporating their relative importance weights is:

$$R = \prod_{i=1}^{n} r_i^{\frac{\alpha_i}{m}} \text{ and } m = \max(\alpha_i)$$

Minimum value occur when the importance values are equal, and at these values the function equals the product of the reliabilities.

Some results

An example might serve to demonstrate the method. We use here the values r from table 1, matrix C of combined reliabilities from table 2, outcomes are shown in table 3.

Table 1: reliability for touristic information sources

ID	**Information source**	**Reliability**
A	Family & Friends	0.85
B	Online Social Networks/Reviews	0.75
C	Non-Official Websites/Blogs	0.70
D	Official Websites (DMOs/Travel agents)	0.55
E	Radio/TV/Movies	0.30
F	Printed media (newspapers/magazines/brochures etc.)	0.25
	Average reliability	***0.56***

Table 2: Combining reliabilities for multiple touristic information sources, rows are related to the first source and columns are related to the second source as order of evaluation time.

	A	**B**	**C**	**D**	**E**	**F**
A	1	.9	.8	.8	.2	.5
B	.9	1	.7	.7	.3	.4
C	.6	.8	1	.7	.5	.3
D	.6	.8	.6	1	.3	.4
E	.1	.1	.4	.2	1	.2
F	.1	.3	.4	.1	.4	1

Table 3: Combining reliabilities R for multiple touristic information sources, provided by tourists (a) and experts (r)

Row	**a**	**mean(a)**	**mean(r)**	**R(C,r_i,a_i)**
1	[1 0 0 0 0 0]	0.16	0.56	0.86

2	[0 1 0 0 0 0]	0.16	0.56	0.85
3	[0.6 0.3 0.1 0 0 0]	0.16	0.56	0.77
4	[0.5 0.5 0 0 0 0]	0.16	0.56	0.74
5	[0 0 0 0 0 1]	0.16	0.56	0.73
6	[0.4 0.3 0.2 0.1 0 0]	0.16	0.56	0.67
7	[0.5 0 0 0 0 0.5]	0.16	0.56	0.63
8	[0.1 0 0 0 0.3 0.6]	0.16	0.56	0.61
9	[0 0 0 0.2 0.3 0.5]	0.16	0.56	0.56
10	[0 0 0 0.33 0.33 0.33]	0.16	0.56	0.44
11	[0.2 0.2 0.2 0.2 0.1 0.1]	0.16	0.56	0.37
12	[0.16 0.16 0.16 0.16 0.16]	0.16	0.56	0.27

Table 3 shows the following columns:

1) **a** denotes the vector containing the judgment of the tourist related to information source;
2) **mean(a)** denotes the average or mean value of vector in (1);
3) **mean(r)** denotes the average or mean value of reliability of each source as opinion expressed by experts;
4) **R(C,r_i,a_i)** combines the individual reliabilities incorporating their relative importance weights.

Some relevant results in table 3 are:

1) row(1): maximum value of R is obtained when user makes only one choice corresponding to the highest score suggested by experts in array r; that is: the concordance between user and expert opinions increases the final score;
2) row (5): high value of R occurs even when a user assigns the highest vote to the worst judgment given by the experts; that is: the user's judgment is more valuable;
3) row (7): the user has issued two opposite scores, concerning only the best choice and the poor provided by the experts;
4) row (12): the minimum value of R is obtained when the user does not know what to choose, she gives the same importance to all opinions from experts, each value contained in array a corresponds to mean(a); similar result are on row (11) where values are almost equal; in this particular case of a fully undecided tourist, the support of external opinions from experts can increase the final result R, and combining opinions in this way acts as a reinforcement;
5) rows (1)-(8): when the user makes a few choices with high scores the final result R is greater than mean(r).

It is possible to choose a decision threshold *t* on final result, so when *R>t* the user can confidently make a decision.

Conclusion

The objective of this work was to determine a simple and rapid mathematical method to obtain a measure of confidence in touristic destination choice, based on weighing a number of possible information sources from experts and user opinions.

In the proposed analysis, different information sources carry different weights that affect the measure of global confidence. The weight of an element is also augmented by the presence of related elements, this reinforcement is realized by combining the reliabilities, using a simple function to determine the combined value.

As practical implications, the method discussed here can be easily implemented by using a small software script that facilitates the calculations.

References

Bramson, A. L. (2004). Methodology for building confidence measures. *Proceedings of SPIE Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, Bellingham, WA, 2004, edited by Belur V. Dasarathy*, *5434*, 270-281.

Jacobsen, J. K. S., & Munar, A. M. (2012). Tourist information search and destination choice in a digital age. *Tourism Management Perspectives, 1,* 39-47.

Llodrà-Riera, I., Martínez-Ruiz, M. P., Jiménez-Zarco, A. I., & Izquierdo-Yusta, A. (2015). A multidimensional analysis of the information sources construct and its relevance for destination image formation. *Tourism Management, 48*, 319-328.

Murphy, H. C., & Chen, M. M. (2014). Online Information Sources Used in Hotel Bookings Examining Relevance and Recall. *Journal of Travel Research*, available online at http://jtr.sagepub.com/content/55/4/523

Xiang, Z., Wang, D., O'Leary, J. T., & Fesenmaier, D. R. (2015). Adapting to the internet: trends in travelers' use of the web for trip planning. *Journal of Travel Research, 54(4)*, 511–527.