Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 162 (2019) 539–546

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Information Technology and Quantitative Management (ITQM 2019)

Application of an opinion consensus aggregation model based on OWA operators to the recommendation of tourist sites

Itzcóatl Bueno[a], Ramón A. Carrasco[a], Raquel Ureña[b], Enrique Herrera-Viedma[c]

[a] *Faculty of Statistics at Complutense University, Avenida Puerta de Hierro, s/n, Madrid 28040, Spain*
[b] *Institute of Artificial Intelligence, De Montfort University, Gateway House, Leicester LE1 9BH, United Kingdom*
[c] *Department of Computer Science and Artificial Intelligence, University of Granada, Avenida de la Fuente Nueva S/N, Granada 18071, Spain*

Abstract

Given the growth in tourism online data as a result of a large number of users posting their personal opinions in social networks and other online platforms with the idea to help other visitants, many authors have proposed a huge variety of ways to classify the sentiments contained in these opinions in order to recommend services (hotels, restaurants, etc.) and destinations to the users with the intention of facilitating their trip planning. In this paper, the authors propose a model to rank tourist sites of a city, based on OWA operators, with the objective of being used as a recommender system.

Keywords: Sentiment Analysis, linguistic quantifiers, OWA operators, tourism.

1. Introduction

Tourism, as was defined in 1995[1] by *World Tourism Organization UNWTO*, include "the activities of persons traveling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes". Since that definition, tourism sector and how users demand their services have evolved rapidly. Consequently, tourism has been transformed from a classic offline industry to one of the most extended e-commerce businesses online ([2, 3]). This has led to an exponential increase in the number of searches for touristic information on the Internet. As a result, several number of reviews, opinions, descriptions, etc. can be found in different social media, blog or online travel agencies.

Paradoxically, this wide range of information, instead of helping users making their travel plans, sometimes could difficult the process. With this regards, a number of recommender systems have been proposed to help users in their trip planning by recommending destinations and other related touristic services. However not many focus

*Corresponding authors.
E-mail address: itzbueno@ucm.es.
E-mail address: ramoncar@ucm.es.
E-mail address: raquel.urena@dmu.ac.uk.
E-mail address: viedma@decsai.ugr.es.

10.1016/j.procs.2019.12.021

on recommending touristic sites in a given city.

In this contribution we propose, using the polarities obtained from the users reviews posted on social media, the application of Yagers' Ordering Weighting Aggregation Operator, OWA, the aim of ranking the most relevant touristic sites in a given city. We consider that this contribution would capture the fuzziness of overall opinions through the linguistic quantifiers such as "Most of the users think..." to give a unique rankable aggregated opinion.

This contribution is organized as follows. In Section 2, we present the related work. A brief summary of the aggregation operator and specifically of the OWA operator are shown in Section 3. In Section 4, we proposed our opinion consensus aggregation model. An application of the model with Paris city data is exposed in Section 5. Conclusions are presented in the last section.

2. Related Work

The tourism field has been explored by a number of authors, mainly through machine learning techniques such as classification or clusterization of destinations, tourist profiles identifications or extracting people sentiments about visited places using their online opinions. However, the literature about the application of OWA operators in tourism is very limited.

The application of machine learning has been widely used to develop recommender systems. For example in [4] it has been analyzed how developed are the content-based and knowledge-based systems in the tourism domain comparing them with other domains. Moreover, they propose the application of demographic recommender algorithms to predict the ratings of touristic attractions. In this line in [5] a method based on multi-criteria collaborative filtering recommender has been proposed. Moreover in [6] it has been developed a system that gives personalized touristic site recommendation using the analytic hierarchy process (AHP) and a bayesian network, taking into account the given users' travel behavior and other similar users' behaviour.

Using a large-scale text-based analysis, G. McKenzie and B. Adams in [7] measure the affinity and discrepancy between different kind of touristic reviews demonstrating their importance in defining how tourists view a city. The information contained in the travel blogs and social media has been also used to develop sentiment classifiers through supervised machine learning algorithms [8]. Sentiment analysis applied to tourism data is used in [9][10] where the authors study the concordance between users' sentiments and automatic sentiment-detection algorithms. An application of the OWA to the tourism industry has been done in [11], where the authors propose a decision making framework in which an expert system integrates with GIS-based ANP-OWA with the objective of finding the optimum location for new tourism facilities.

3. Preliminaries

There is a huge variety of aggregation operators classified in 4 groups according to their attitudes or position respect to the minimum and maximum operator (Dubois & Prade, 1985)[12]:

- The conjunctive attitude: It requires the satisfaction of all the used criteria and it is represented by

$$h(u_1, \ldots, u_q) \leq min(u_1, \ldots, u_q)$$

- The disjunctive attitude: Operators in which just some criteria must be satisfied and it is defined by the following axiom:

$$h(u_1, \ldots, u_q) \geq max(u_1, \ldots, u_q)$$

- The compromise attitude: This kind of operators describe a trade off between values returning a result within both extremes

$$min < h < max$$

- Hybrids: In this group are the operators with a mix attitude so it is not possible to include them in any of the other groups

In this paper we focus on the operators known as Ordered Weighted Averaging (OWA) which are included in the compromise attitude operators. They were introduced by Yager in 1988 as a new compensation operator. As they introduce weights they can be seen as the weighted means. However, in the OWA the weights do not affect to an specific criterion but to the position it has after ordering them: each weight ω_i is paired with the i_{th} bigger element.

DEFINITION 1. *(OWA Operator [13]) It is a function of dimension* κ, $F : \mathbb{R}^\kappa \longrightarrow \mathbb{R}$, *that has associated a set of weights or weighting vector* $\Omega = (\omega_1, \ldots, \omega_\kappa)$ *to it, so that* $\omega_i \in [0, 1]$ *and* $\sum_{i=1}^{\kappa} \omega_i = 1$, *and is defined to aggregate a list of real values* $\{u_1, \ldots, u_\kappa\}$ *according to the following expression:*

$$F(u_1, \ldots, u_\kappa) = \sum_{i=1}^{\kappa} \omega_i \cdot u_{\sigma(i)},$$

being $\sigma : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ *a permutation such that* $u_{\sigma(i)} \geq u_{\sigma(i+1)}, \forall i = 1, \ldots, n-1$, *i.e.,* $u_{\sigma(i)}$ *is the* i_{th} *highest value in the set* $\{u_1, \ldots, u_\kappa\}$.

Three notable OWA operators were defined in (Yager, 1988):

- $F(u_1, \ldots, u_\kappa) = max(u_1, \ldots, u_\kappa)$ when $\omega_1 = 1$ and $\omega_i = 0 \; \forall i \neq 1$
- $F(u_1, \ldots, u_\kappa) = min(u_1, \ldots, u_\kappa)$ when $\omega_\kappa = 1$ and $\omega_i = 0 \; \forall i \neq \kappa$
- $F(u_1, \ldots, u_\kappa) = avg(u_1, \ldots, u_\kappa)$ when $\Omega = \left[\frac{1}{\kappa}, \frac{1}{\kappa}, \ldots, \frac{1}{\kappa}\right]$

The weighting function establishes the difference between OWA operators. One of them is the fuzzy quantifier which is defined as a function $Q : [0, 1] \Longrightarrow [0, 1]$ where $Q(0) = 0, Q(1) = 1$ and $Q(x) \geq Q(y)$ for $x > y$. We can construct linguistic quantifiers such as *"at least half"*, *"most of"* and *"as many as possible"* Figure 1 using the following function (Zadeh, 1983):

$$Q(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } b < x \end{cases} \tag{1}$$

Based on the Q function, the elements of the weighting vector used in the OWA are defined from:

$$\omega_i = Q\left(\frac{i}{\kappa}\right) - Q\left(\frac{i-1}{\kappa}\right) \tag{2}$$

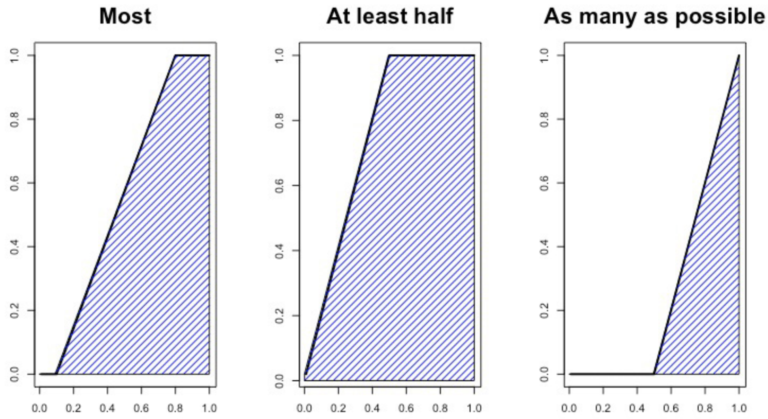

Fig. 1: Some linguistic quantifiers

In Figure 1 we show the form of the linguistic quantifiers we are going to use in our case of use to obtain the weights of the OWA operators. We have established the values of a and b at (0.1 , 0.8), (0 , 0.5) and (0.5 , 1) for *"the most of"*, *"at least half"* and *"as many as possible"* respectively.

4. An Opinion Consensus Aggregation Model

In this section we present our proposed consensus aggregation model, whose architecture has been outlined in 2.

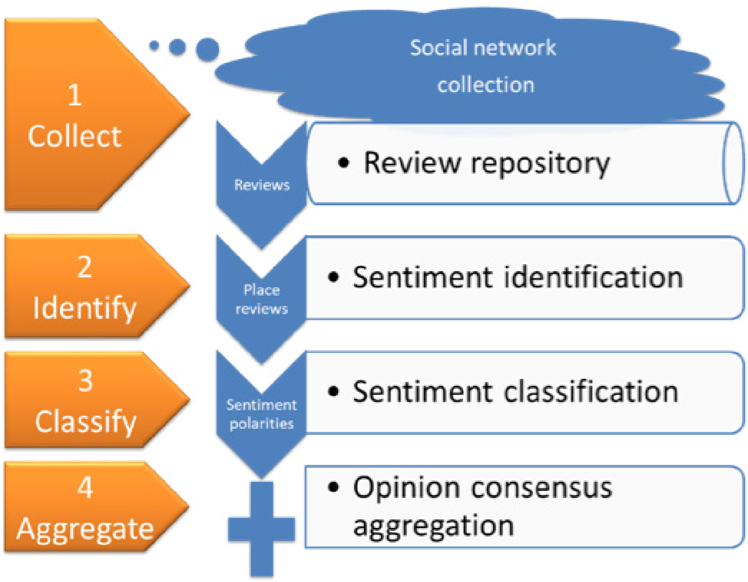

Fig. 2: Model architecture

For opinion aggregation have been proposed multiple architectures. In this contribution we use the ones proposed in [14, 15], composed of the following four main steps, depicted in 2:

- Collect: Information is extracted from social networks and stored into a repository.
- Identify: In this phase, pinion sentences on tourist sites are identified.
- Classify: A sentiment polarity is generated for each opinion sentence
- Aggregate: Finally, the sentiment polarities are aggregated by each tourist site.

4.1. Collect

In the first step city attractions information is obtained from the Web API of http://tour-pedia.org, which is part of the OpeNER project. Linked to this information are the reviews of each attraction ($A_i, i = 1, \ldots, \kappa$), obtained from four social network: Booking, Facebook, Foursquare and Google Places. Finally, the repository to use is composed by the name of the attraction, a text and the polarity of each review ($r_i, i = 1, \ldots, \kappa$). Although we have choose this data source for our analysis it is probably that public organisms have similar data in order to reproduce this example in their place attractions.

4.2. Identify and Classify

The steps to identifying the opinion sentences on tourist site and classifying them into sentiment polarities is solved given that the Web API includes the polarity, which is between 0 and 10, of each review. This scale has to be understood as the higher, the better, that is, 0 is the worst an attraction could be valued and 10 is the maximum. According to the polarity tagger section in the OpeNER project documentation[1], the tool tags words with polarity information, that is, positive or negative notions in a certain domain, and also with sentiment modifiers to inflate or attenuate the polarity, so it may be using a lexicon-based approach to tag the global polarity to a review. This tool support reviews in the following languages: Dutch, German, English, French, Italian and Spanish.

4.3. Aggregate

In the last step, we propose the aggregation of every review associated to an attraction through an OWA operator using different linguistic quantifiers to define the used weights. To summarize, for each touristic site a global priority is obtained, in order to rank them, following the scheme showed in Figure 3.

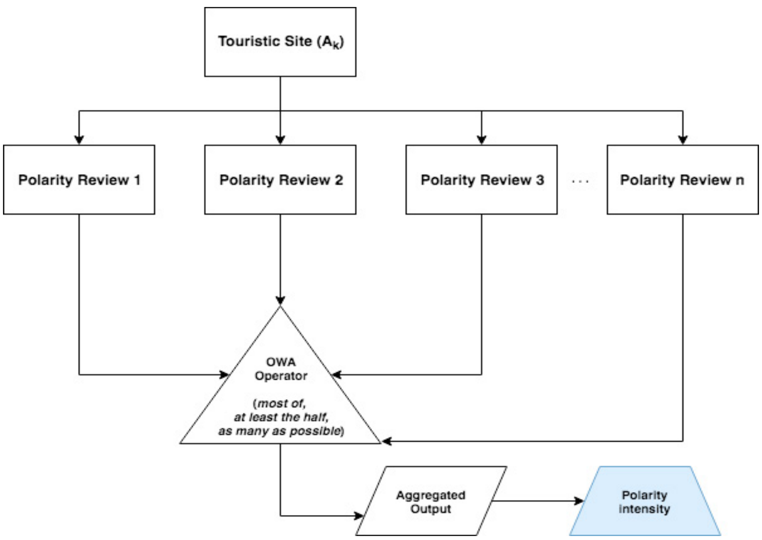

Fig. 3: Scheme of the opinion consensus aggregation model

[1] https://www.opener-project.eu/documentation/polarity-tagger.html

5. Using the proposed approach for evaluating the Paris' main touristic attractions

In this section we carry out an application of the exposed opinion consensus aggregation model to the city of Paris. With this purpose we follow the steps specified in Figure 2.

First of all we gather the data from the tour-pedia Web API, a dataset composed of 827 attractions (monuments, museums, gardens, etc.)[Figure 4]. We reduce it for this example to those places with at least 100 reviews, to guarantee the representativeness of opinions, and the quality of the tourist attractions, deleting pubs and other similar places.

name	address	lat	lng	numReviews
Tour Eiffel	5 Avenue Anatole France	48.858...	2.294...	863
Louvre Museum	Paris, France	48.861...	2.335...	763
Cathedrale Notre-Dame de Pa...	6 Parvis Notre-Dame	48.853...	2.349...	273
Orsay Museum	1 Rue de la L<ed><a9>gion d'Honneur, Paris, France	48.859...	2.326...	228
Luxembourg Gardens	Paris, France	48.846...	2.337...	188
Basilique du Sacre-Cour	35 Rue du Chevalier de la Barre	48.886...	2.343...	162
Moulin Rouge	82 Boulevard de Clichy	48.884...	2.332...	135
Tuileries Garden	Place de la Concorde, Paris, France	48.863...	2.327...	109

Fig. 4: Attraction repository

Once the dataset is generated the next stage consists on the identification and classification of the reviews associated with the tourist sites. From the 8 tourist sites selected (Tour Eiffel, Louvre Museum, Cathedrale Notre-Dame de Paris, Orsay Museum, Luxembourg Gardens, Basilique du Sacre-Cour, Moulin Rouge and Tuileries Garden) we obtain all the reviews for each of them[Figure 5].

place	language	polarity	source	text	date
Tour Eiffel	en	10	Foursquare	HighlightSee Paris in all its glory, on the observation ...	2010-07-26
Tour Eiffel	en	5	Foursquare	The Eiffel Tower is open every day all year long :- fro...	2010-07-28
Tour Eiffel	en	10	Foursquare	said people should never be alone to the Eiffel Tower,...	2010-07-29
Tour Eiffel	en	5	Foursquare	Wish I was there!	2010-07-31
Tour Eiffel	en	10	Foursquare	Can't forget the moment when you sip hot fresh brew...	2010-07-31
Tour Eiffel	fr	0	Foursquare	Sympa... Un peu venteux ;-)	2010-08-03
Tour Eiffel	fr	5	Foursquare	L'ours de la tour eiffel est magnifique	2010-08-12

Fig. 5: Reviews repository

After their classification the aggregation is carried out. In this last stage, the opinions of each site are aggregated into an overall polarity in order to rank the touristic attractions of Paris. We use the linguistic quantifiers exposed in Figure 1 to compute the weights used in the OWA operator to aggregate the reviews and construct three different rankings (see Table 1). The weights are obtained using Equation 2 where for each touristic site, κ is the number of reviews it has and $i = 1, \ldots, \kappa$.

Site	Most	At least the half	As many as possible
Basilica of the Sacred Heart of Paris	$5.97^{(4)}$	$8.48^{(4)}$	$1.88^{(3)}$
Notre-Dame Cathedral	$5.18^{(5)}$	$7.87^{(5)}$	$1.37^{(6)}$
Louvre Museum	$5.02^{(6)}$	$7.58^{(6)}$	$1.45^{(5)}$
Luxembourg Gardens	$6.07^{(3)}$	$8.67^{(3)}$	$1.83^{(4)}$
Moulin Rouge	$4.52^{(7)}$	$7.46^{(7)}$	$0.87^{(8)}$
Orsay Museum	$6.40^{(2)}$	$8.68^{(2)}$	$2.27^{(2)}$
Tour Eiffel	$4.51^{(8)}$	$7.40^{(8)}$	$0.91^{(7)}$
Tuileries Garden	$7.56^{(1)}$	$9.59^{(1)}$	$3.48^{(1)}$

Table 1: Overall polarity ranking with different linguistic quantifiers

As can be seen in Table 1, using the linguistic quantifiers *most* and *at least half* to obtain the weights, the OWA operator ordered the attractions equally although their polarity intensities change from one to another. Using *as many as possible* linguistic quantifier for the OWA we obtain a slightly change in the order of the touristic sites (for example, Tour Eiffle climbs a position) and a remarkable reduction in the polarity intensities. The table shows two measures for each ranking. Between parenthesis we can see the position in the ranking of the tourist sites to visit, while the other result is the valuation given by the users. The linguistic interpretation of the results can be understood through the following example: *The best valued tourist site in Paris is Toulleries Garden in such a way that the most of users give it a valuation of 7.56 out of 10.*

6. Conclusions and Future work

In this paper contribution proposed an opinion consensus aggregation model based on the OWA operators. The proposed approach applicability has been proved through an example using a real data set of the tourist sites in Paris. Some future extensions of the proposed model could be the aggregation of the polarities obtained using diverse machine learning algorithms, the integration of fuzzy logic in our model or the used of more complex OWA operators. Moreover, it has been observed that although some opinions for a given site are positive, their associated polarities are low given they contain, for example, a complaint about the weather during their visit. Thus another extension is to use multi-criteria decision making models such as AHP to take into account different aspects (weather, city geographical situation, amount of people, impression...) of the given opinion and applying the previous mentioned extensions in order to have a more complete and realistic ranking through the combination of the polarities of each aspect associated to the opinions.

Acknowledgments

The authors would like to acknowledge the financial support from the EU project H2020-MSCA-IF-2016-DeciTrustNET-746398.

This paper has been elaborated with the financing of FEDER funds in the Spanish National research project TIN2016-75850-R.

References

[1] UNWTO, Technical manual no. 2: Collection of tourism expenditure statistics (1995).
[2] J. Kracht, Y. C. Wang, Examining the tourism distribution channel: evolution and transformation, International Journal of Contemporary Hospitality Management 22 (4-5) (2010) 736–757. doi:10.1108/09596111011053837.
URL <GotoISI>://WOS:000281538200020
[3] U. Gretzel, M. Sigala, Z. Xiang, C. Koo, Smart tourism: foundations and developments, Electronic Markets 25 (3) (2015) 179–188. doi:10.1007/s12525-015-0196-8.
URL <GotoISI>://WOS:000359730200003
[4] Y. Y. Wang, S. C. F. Chan, G. Ngai, Applicability of demographic recommender system to tourist attractions: A case study on tripadvisor, 2012 Ieee/Wic/Acm International Conference on Web Intelligence and Intelligent Agent Technology Workshops (Wi-Iat Workshops 2012), Vol 3 (2012) 97–101doi:10.1109/wi-iat.2012.133.
URL <GotoISI>://WOS:000423017600021

[5] M. Nilashi, K. Bagherifard, M. Rahmani, V. Rafe, A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques, Computers & Industrial Engineering 109 (2017) 357–368. doi:10.1016/j.cie.2017.05.016.
URL <GotoISI>://WOS:000405052300030

[6] Y. X. Huang, L. Bian, A bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the internet, Expert Systems with Applications 36 (1) (2009) 933–943. doi:10.1016/j.eswa.2007.10.019.
URL <GotoISI>://WOS:000264182800092

[7] G. McKenzie, B. Adams, A data-driven approach to exploring similarities of tourist attractions through online reviews, Journal of Location Based Services 12 (2) (2018) 94–118. doi:10.1080/17489725.2018.1493548.
URL <GotoISI>://WOS:000446541600002

[8] Q. Ye, Z. Q. Zhang, R. Law, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, Expert Systems with Applications 36 (3) (2009) 6527–6535. doi:10.1016/j.eswa.2008.07.035.
URL <GotoISI>://WOS:000263817100095

[9] A. Valdivia, M. V. Luzon, F. Herrera, Sentiment analysis in tripadvisor, Ieee Intelligent Systems 32 (4) (2017) 72–77. doi:10.1109/mis.2017.3121555.
URL <GotoISI>://WOS:000408156300010

[10] A. Valdivia, M. V. Luzon, F. Herrera, Sentiment analysis on tripadvisor: Are there inconsistencies in user reviews?, Hybrid Artificial Intelligent Systems, Hais 2017 10334 (2017) 15–25. doi:10.1007/978-3-319-59650-1_2.
URL <GotoISI>://WOS:000432880600002

[11] K. A. Eldrandaly, M. A. AL-Amari, An expert gis-based anp-owa decision making framework for tourism development site selection, International Journal of Intelligent Systems and Applications 6 (7) (2014) 1.

[12] D. Dubois, H. Prade, A review of fuzzy set aggregation connectives, Information Sciences 36 (1-2) (1985) 85–121. doi:10.1016/0020-0255(85)90027-1.
URL <GotoISI>://WOS:A1985A019600007

[13] O. Appel, F. Chiclana, J. Carter, H. Fujita, A consensus approach to the sentiment analysis problem driven by support-based iowa majority, International Journal of Intelligent Systems 32 (9) (2017) 947–965. doi:10.1002/int.21878.
URL <GotoISI>://WOS:000406958100004

[14] R. A. Carrasco, P. Villar, M. J. Hornos, E. Herrera-Viedma, A linguistic multi-criteria decision making model applied to the integration of education questionnaires, International Journal of Computational Intelligence Systems 4 (5) (2011) 946–959. doi:10.1080/18756891.2011.9727844.
URL <GotoISI>://WOS:000297797300019

[15] R. A. Carrasco, P. Villar, M. J. Hornos, E. Herrera-Viedma, A linguistic multicriteria decision-making model applied to hotel service quality evaluation from web data sources, International Journal of Intelligent Systems 27 (7) (2012) 704–731. doi:10.1002/int.21546.
URL <GotoISI>://WOS:000304396100004