Expert Systems With Applications 169 (2021) 114306

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Criteria determination of analytic hierarchy process using a topic model

Jin Fang [a,*], Fariborz Y. Partovi [a]

[a] *Drexel University, USA*

ARTICLE INFO

Keywords:
Analytic hierarchy process
Topic model
Latent Dirichlet allocation
Hotel selection
Restaurant selection
Group decision

ABSTRACT

The purpose of this paper is to develop a technology-based model for identifying various criteria in a decision-making situation. We used topic modeling to discover critical criteria and their corresponding weights in the Analytic Hierarchy Process (AHP). Approximately 100,000 hotel reviews and 100,000 restaurant reviews were scraped from TripAdvisor.com for criteria determination. Next, an AHP model with criteria and 12 hotels/restaurants as alternatives were compared and ranked. The results compared favorably with more than 1000 reviews of these hotels/restaurants in TripAdvisor.com, thus validating the methodology.

1. Introduction

Analytic Hierarchy Process (AHP) (Saaty, 1988) is a useful method to assess the finite number of alternatives in multi-criteria decision-making (MCDM) problems. AHP provides judgments on both quantitative and qualitative criteria (Badri, 2001). Since its initial development, AHP has been applied to various decision areas in manufacturing and service (Partovi, Burton, & Banerjee, 1990; Mikhailov & Tsvetinov, 2004; Ishizaka & Labib, 2011). There are three main steps in using AHP: (1) construction of the hierarchy of the problem; (2) obtaining pairwise matrices of the criteria and the alternatives to provide relative importance; and (3) revealing priorities. In the first step, a hierarchy composed of decision elements is developed based on the problem. Albayrak and Erensal (2004) mentioned that a hierarchy has at least three levels: the overall objective of the decision problem at the top; multiple criteria and sub-criteria that evaluate the alternatives in the middle; the decision options or alternatives at the bottom. The second step is obtaining pairwise comparison matrices of the criteria and the alternatives. Decision-makers assign the relative importance based on their opinions for each pair of criteria and alternatives, separately using the 'Saaty scale' (Saaty, 1980). The alternatives' composite weights are then determined by aggregating the weights throughout the hierarchy in the final step. The outcome of this aggregation, so-called priorities, is an overall weight for each alternative.

It has been forty years since the AHP was introduced by Saaty (1980). There have been many modifications and improvements to the original framework, such as rank reversal, ANP, and fuzzy AHP. Belton and Gear (1983) introduced the rank reversal problem in the AHP. Barzilai and Golany (1994) showed that this issue could be avoided when the process's output is properly redefined as a weight-ratio matrix rather than a normalized-weight vector. The ANP technique, also developed by Saaty (1996), is another multi-criteria decision tool that allows for considering more complex interdependencies among levels of attributes and alternatives. ANP can include hierarchical relationships but does not require a strict hierarchical structure as does AHP. In traditional formulations of the AHP, human judgments are represented as exact numbers. However, decision-makers may find it more confident to give interval judgments than fixed value judgments. Therefore, Van and Pedrycz (1983) proposed the fuzzy AHP, which introduced fuzzy ratios using triangular membership functions. Researchers also combined AHP with other tools like linear programming, data envelopment analysis, artificial neural network, quality function deployment (Partovi & Corredoira, 2002), genetic algorithms, and SWOT-analysis (Ho, 2008).

All of the above improvements are related to the evaluation or mathematical aspects of AHP. One crucial aspect that has not been addressed scientifically is determining the criteria of AHP, which includes (a) selection of the criteria; (b) obtaining the weights of the criteria; and (c) evaluating the alternatives for these criteria. In part (a), the number of criteria is suggested to be at seven or less in the AHP (Saaty & Ozdemir, 2003) because when the number of criteria increases past seven, the inconsistency increases. Traditional methodologies of criteria selection in AHP mainly arise from the literature and the organization's expertise, a few cases were supported by external specialists (de FSM Russo & Camanho, 2015). However, it is both costly and challenging to find a suitable time slot to gather all experts or specialists together to have a live discussion about the criteria selection. Even in

* Corresponding author.
E-mail addresses: jf699@drexel.edu (J. Fang), partovi@drexel.edu (F.Y. Partovi).

https://doi.org/10.1016/j.eswa.2020.114306
Received 26 February 2020; Received in revised form 18 August 2020; Accepted 9 November 2020
Available online 24 December 2020

ideal cases, a consistent agreement is usually challenging to reach as the number of participants and related discussions increases. Some essential criteria may be omitted because one or two persons may dominate the discussions, and other participants hesitate to offer their opinion. In part (b), since it is difficult to compare different criteria in the group judgment, obtaining the criteria' weights is even more difficult. In the majority of AHP cases, decision-makers would build a consensus among themselves and use individual judgment aggregation (Escobar & Moreno-JimenTz, 2007) to calculate the criteria weight. However, the method in which the consensus was obtained, and inconsistency in the AHP application occurred were not mentioned. In part (c), evaluation of alternatives was grounded in the literature, and a few cases adopted the opinions of experts using the Delphi method (Yang, Chuang, & Huang, 2009; Vidal, Marle, & Bocquet, 2011). The Delphi method (Linstone et al., 1975) is based on multiple rounds of questionnaires sent to a panel of experts. There are two disadvantages. Firstly, the Delphi method's response time can be long, which slows the rate of discussion. Secondly, it is also possible that the information received from the experts may not provide innate value.

Although previous AHP applications provide the criteria determination, they are highly dependent on human judgments. Hence, it could be the case that they omit some essential criteria. It is both time-consuming and costly to carry out the group decision-making processes as mentioned above systematically. Also, determining the weights of criteria is demanding. To overcome these shortcomings, there is a need to propose an innovative and effective intelligent system to determine certain judgments' criteria. The rapid development of information technology and the internet has dramatically changed the searching process, Nielsen (2012) concluded that 70% of customers trust online reviews. Therefore review data are considered as one of the most influential components in the decision-making process of customers. Incorporating an intelligent system into analyzing the review data is beneficial as it allows decision-makers to learn criteria determination based on the large volume of documents written by various customers. This insightful knowledge from the intelligent system can then be used by decision-makers to rethink their decision-making process. To analyze the review data, this intelligent system is constructed using LDA for topic modeling. A topic model is a statistical model for discovering the hidden "topics" that occur in a collection of documents. This paper utilizes LDA to extract the latent topics from the TripAdvisor reviews. Also, LDA assigns a probabilistic mixture of these topics into the reviews. The LDA does not require any prior knowledge of the original text documents as it learns topics from analyzing text data. Also, LDA can assign topics to future reviews.

This paper aims to demonstrate the applicability of the topic modeling approach to analyze reviews in determining the criteria within an AHP setting. We use TripAdvisor reviews to identify the criteria of hotel selection and restaurant selection as examples. After obtaining the criteria and their weights from topic modeling, we will apply the results to twelve four-star hotels and twelve restaurants with more than 1000 reviews in Philadelphia, respectively. These twelve hotels or restaurants become alternatives in the AHP. There has been no published paper analyzing criteria determination in AHP using text mining methodologies to the best of the authors' knowledge.

2. Background

2.1. Latent Dirichlet allocation

LDA is a generative probabilistic topic model that uncovers latent thematic structures from a corpus D (Blei, Ng, & Jordan, 2003). The latent thematic structure, expressed as topics and topic proportions per document, is represented by hidden variables that LDA posits about the corpus. Documents are modeled as finite mixtures over an underlying set of these latent topics. LDA is based on the 'bag of words' assumption, in which the order of words is neglected (Blei et al., 2003). The generative process is defined as follows:

1. For every topic k = {1, …, K}, sample a topic-word distribution, $\beta_k \sim \mathrm{Dir}(\eta)$
2. For every document d
   (a) sample a document-topic distribution, $\theta_d \sim \mathrm{Dir}(\alpha)$
   (b) for each word w within document d
      i. sample a word-topic assignment $z_{d,n} \sim \mathrm{Mult}(\theta_d)$, where $z_{d,n} \in \{1, \ldots, K\}$
      ii. sample a word $w_{d,n} \sim \mathrm{Mult}(\beta_{z_{d,n}})$, where $w_{d,n} \in \{1, \ldots, V\}$

Each topic β_k is a multinomial distribution over the vocabulary V and comes from a Dirichlet distribution $\beta_k \sim \mathrm{Dir}(\eta)$. Additionally, every document is represented as a distribution over K topics and comes from a Dirichlet distribution $\theta_d \sim \mathrm{Dir}(\alpha)$. The Dirichlet parameter α denotes the smoothing of topics within documents, and η denotes the smoothing of words within topics.

However, we only observe the words within documents and need to infer the hidden structure, which are the topics and topic proportions per document. This inference aims to answer: Which hidden structure or topic model is most likely to have generated these documents? To answer this question, we obtain the posterior distribution that captures the hidden structure given the observed documents. The word topic assignment z_D, the document topic distribution θ_D, and the topics β_K are the latent variables and are not observed. We would have to condition the only observed variable, the words within the documents w_D, to infer the hidden structure with statistical inference. This can be viewed as a reversal of the generative process. The conditional probability, also known as the posterior, is expressed by the following equation:

$$p\left(\beta_K, \theta_D, z_D | w_D\right) = \frac{p(\beta_K, \theta_D, z_D, w_D)}{p(w_D)} \tag{1}$$

Blei et al. (2003) concluded that the posterior computation is intractable due to the denominator. The marginal probability $p(w_D)$ is the sum of the joint distribution over all instantiations of the hidden structure and is exponentially large (Blei, 2012). Although the posterior cannot be computed precisely, a close enough approximation to the true posterior can be achieved with posterior inference. There are mainly three types of inference techniques: Gibbs sampling (Griffiths & Steyvers, 2004), variational method (Blei et al., 2003), and expectation propagation (Minka & Lafferty, 2012). This paper has used Gibbs sampling (Griffiths & Steyvers, 2004), a Markov chain Monte Carlo algorithm to infer parameters.

2.2. Hotel and restaurant selection based on TripAdvisor

TripAdvisor.com is a popular tourism website that contains more than 435 million travel reviews submitted by travelers. These reviews represent the honest, unbiased opinions of travelers who have visited various hotels and restaurants. Recently, many researchers from various fields have contributed to the study of hotel selection using TripAdvisor reviews. For example, Casalo, Flavian, Guinaliu, and Ekinci (2015) discussed the influences of online comments regarding tourists' intentions in selecting a hotel. They indicated that comments are reliable and helpful when posted on popular tourism websites, such as TripAdvisor. Zaman, Botti, and Thanh (2016) calculated the weight of criteria using AHP by considering tourists' points of view. This study used six criteria proposed on TripAdvisor and interrogated 250 tourists in Paris to determine the weights. Rossetti, Stella, and Zanker (2016) provided decision support and recommendations to online tourists with a topic model method using TripAdvisor dataset. This paper focuses on rating prediction and recommendation. It does not determine hotel criteria selection or the corresponding weight of each criterion. Yu, Wang, and Wang (2018) proposed an acronym in Portuguese of interactive and multi-criteria decision-making (TODIM) method to prioritize hotels on TripAdvisor.com. They evaluated hotels using intuitionistic linguistic numbers. A non-linear programming model calculates the criteria weights.

There have been previous research works discussing restaurant selection using TripAdvisor reviews. Lei and Law (2015) did a content analysis of Tripadvisor reviews on Restaurants in Macau and showed that overall customer satisfaction on Macau's dining experience was positive. Zhang, Ji, Wang, and Chen (2017) introduced fuzzy sets to denote restaurant reviews from TripAdvisor and utilized the Bonferroni mean to consider interdependence among criteria. Laksono, Sungkono, Sarno, and Wahyuni (2019) classified Surabaya restaurant customer satisfaction using Naive Bayes from the sentiment analysis of restaurant reviews on TripAdvisor. Most of the literature mentioned above merely implemented auxiliary explorations and needs opinions of high-level experts. These are causing difficulties in building an intelligent system for hotel and restaurant selection. The increasing volume of reviews from numerous customers is the primary concern for hotel and restaurant selections. The existing studies do not solve these problems effectively.

3. Dataset

3.1. Preparing hotel and restaurant review datasets

The hotel and restaurant reviews were scraped from TripAdvisor.com using ParseHub, a user-friendly visual web scraper. We denote hotel reviews dataset as DS_1 and restaurant reviews dataset as DS_2. DS_1 contains 101,744 customer reviews of various hotels in Philadelphia and includes the hotel name, reviewer name, submission date, and review description. DS_2 contains 128,855 customer reviews of various restaurants in Philadelphia and includes the restaurant name, reviewer name, submission date, and review description. It was observed that some records repeat others, hence we first removed the repeated records. Also, a review may not exist for every record because guests may not choose to share their experiences. Reviews are the main focus of this research, therefore we removed the records without reviews. This gave 101,706 and 127,344 records for DS_1 and DS_2, respectively. Figs. 1 and 2 present the average number of reviews by month in DS_1 and DS_2, concluding that July is the most popular tourism month in Philadelphia.

Fig. 3 presents the average number of reviews from 1999 to 2020 in DS_1. Fig. 4 presents the average number of reviews from 2010 to 2020 in DS_2. Table 1 shows the summary statistics of reviews for DS_1 and DS_2. While in some cases, there may be only two words in the DS_1, however, the mean document length in DS_1 is 89.29, and the median is 82. Compared with DS_1, DS_2 has less mean and median document length, 40.14 and 45, respectively.

3.2. Data preprocessing

After obtaining the datasets from Section 3.1, we started to preprocess the data with LDA using python. The data preprocessing includes six steps: (1) Tokenize and remove special characters, (2) create an n-gram model, (3) remove stop words, (4) make n-grams, (5) lemmatization, and (6) create a term-document matrix. In step 1, each sentence is tokenized into a list of words with special characters and punctuations removed. This step is achieved by using 'simple_preprocess' function in the python library Gensim. Step 2 builds an n-gram model, which is the union of unigrams and bigrams using Phrases function in Gensim. Each word is a unigram, and every two adjacent words create a bigram. For example, the unigrams of the sentence from the reviews 'Super kind staff went beyond to accommodate our situation' are 'Super', 'kind', 'staff', 'went', 'beyond', 'to', 'accommodate', 'our', 'situation'. While the bigrams for this sentence are 'Super kind', 'kind staff', 'staff went', 'went beyond', 'beyond to', 'to accommodate', 'accommodate our', 'our situation'. We set the 'min_count' argument in Phrases function equal to 5, ignoring all unigrams and bigrams with total collected count lower than 5. Another essential argument in Phrases function is 'threshold', representing a score threshold for forming the phrases (higher means fewer phrases). A phrase of words A followed by B is accepted if the phrase's score is greater than the threshold. The score of phrase of words A followed by B is calculated by bigram scoring function in Mikolov, Sutskever, Chen, Corrado, and Dean (2013), which is

$$\frac{\text{Number of co}-\text{occurrences for phrase ''A\_B''}-\text{Minimum collocation count threshold}}{\text{Number of occurrences for A}\times\text{Number of occurrences for B}}$$

We set the 'threshold' argument equal to 10, which is the minimum score for a bigram to be taken into account in our n-gram model. Step 3 is a crucial stage in the data preprocessing. A stop word is a commonly used word (such as "the", "a", "an", "in") that a search engine has been programmed to ignore. This is implemented using Python natural language toolkit (NLTK) library (Joakim, 2012). Step 4 is to make unigrams and bigrams for the implementation of Step 5. The goal of Step 5 is to consider the use of vocabulary and morphological analysis of words. It usually aims to remove inflectional endings and return the root of a word, known as the lemma. For example, the lemma of the word

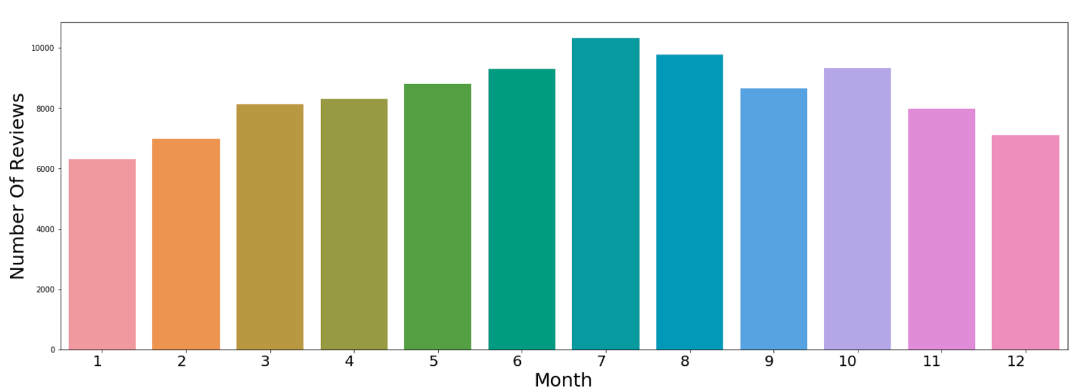

Fig. 1. The average number of reviews by month in DS_1.

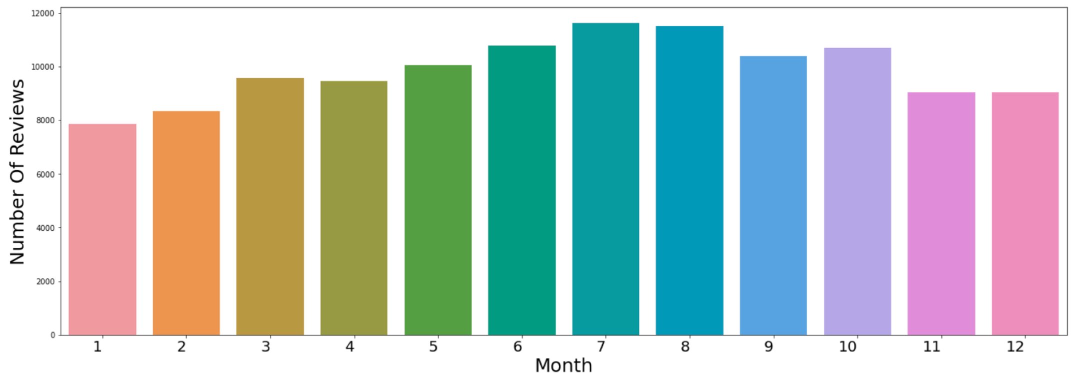

Fig. 2. The average number of reviews by month in DS_2.

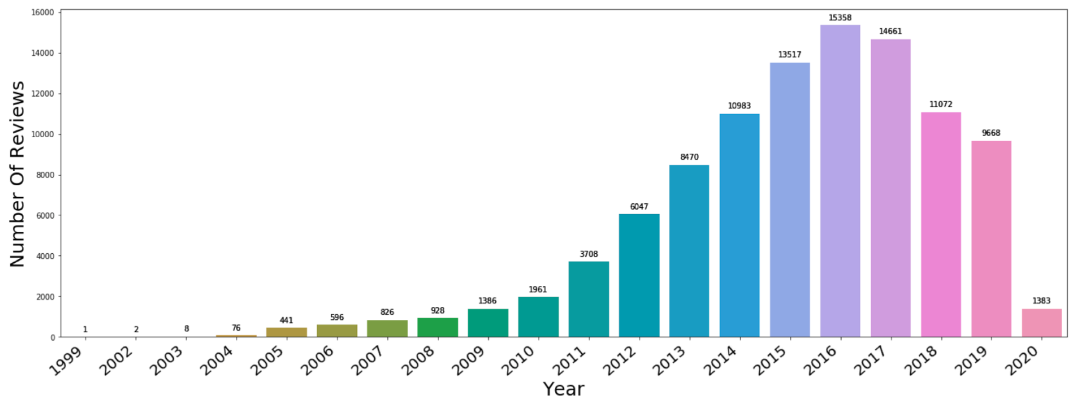

Fig. 3. The average number of reviews by year in DS_1.

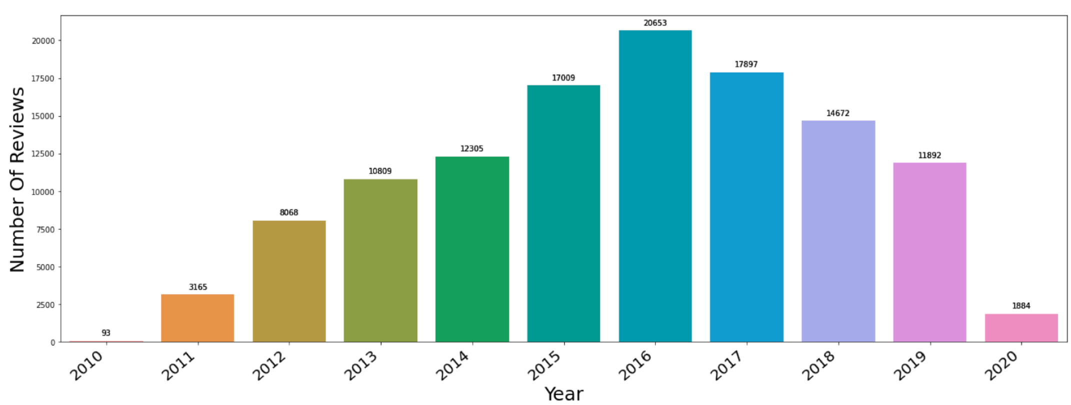

Fig. 4. The average number of reviews by year in DS_2.

Table 1
Overview of the DS_1 and DS_2.

	DS_1 (hotel reviews dataset)	DS_2 (restaurant reviews dataset)
Time range	1999–2020	2010–2020
Number of documents	101,706	127,344
Mean document length	89.29	40.14
Median document length	82.00	45.00
Minimum document length	2.00	1.00
Maximum document length	180.00	48.00

'machines' is 'machine'. Likewise, 'walking' is converted to 'walk', 'mice' is converted to 'mouse'. This lemmatization step is very significant in text mining. We use unigrams and bigrams from Step 4 as the input and only keep nouns, adjectives, verbs, and adverbs, producing 42790 unique features for DS_1 and 39,228 unique features for DS_2. Step 6 is creating a term-document matrix using Gensim. This matrix is the frequency of terms (rows) in documents (columns), which is the LDA algorithm's corpus. Gensim creates a unique ID for each word in the document. The produced corpus is a mapping of "word_identity" to "word_frequency". Data preprocessing is an essential step in our analysis because it removes noise from the original data, which could affect LDA's performance.

4. The model

The proposed model can be divided into five sub-sections: (1) creating LDA models, (2) topic coherence, (3) topic representation and assignments, (4) determination of criteria, weights, and alternatives in AHP, and (5) hotel and restaurant rankings via AHP.

4.1. Creating LDA models

The LDA analysis was conducted by Python wrapper from McCallum (2002), the Java topic modeling toolkit. The Mallet topic model package includes a fast and highly scalable implementation of Gibbs sampling, efficient methods for document-topic hyperparameter optimization, and tools for inferring topics for new documents given trained models. The LDA algorithm requires two hyper-parameters α, η, and the number of topics K. α controls the division of documents into topics and η controls division of topics into words. Larger values of η yield coarser topics and larger values of α yields coarser distribution of document into topics. Based on the suggestion of the creators of Mimno (2013), we initially set α to 5.0 and η to 0.01 and then allow Mallet to optimize these hyper-parameters every ten iterations. We set the number of training iterations equal to 1000. For both datasets, we created 59 different LDA models by varying the K parameter (the number of topics) from 2 to 60 and repeating this process three times using the random seed equal to 4, 5, and 6 to ensure consistent results. Therefore, we created $3 \times 59 \times 2 = 354$ LDA models in total.

4.2. Topic coherence

The number of topics K is also required by the LDA model. Previous researchers determine the number of topics by trial and error procedures (Blei et al., 2003; Steyvers & Griffiths, 2007; Blei, Carin, & Dunson, 2010). They select the number K, which produces the most interpretable and meaningful outcome. In recent years, an innovative nonparametric Bayesian approach hierarchical Dirichlet process (HDP) (Teh, Jordan, Beal, & Blei, 2005) has been proposed. This approach can be used to determine the number of topics automatically. However, compared with LDA, the performance of HDP was not good concerning the formation of semantically meaningful topics. Therefore, our analysis built many LDA models with different values of the number of topics (K) and picked the one that gave the optimal coherence score. We calculated the corehence score C_v (Röder, Both, & Hinneburg, 2015) using the Gensim library. C_v is based on four parts (Syed & Spruit, 2017): (i) segmentation of the data into word pairs, (ii) calculation of word or word pair probabilities, (iii) calculation of a confirmation measure that quantifies how strongly a word set supports another word set, and (iv) aggregation of individual confirmation measures into an overall coherence score.

In part (i), data segmentation pairs each of the topic's top-N words with every other top-N word. Let W be the set of a topic's top-N most probable words $W = \{w_1, \dots, w_N\}$. Each word $W' \in W$ paired with all words $W^* \in$ W, generating a segmented pair S_i. S is defined as $S = \{(W', W^*)|W' = \{w_i\}; w_i \in W; W^* = W\}$. Part (ii) uses Boolean sliding window calculation, where a new virtual document is created for every window size of s when sliding over the document at a rate of one word token per step. For example, document d_1 with words w results in virtual documents $d_1' = \{w_1, \dots, w_s\}$ and $d_2' = \{w_2, \dots, w_{s+1}\}$ and so on. Then the number of virtual documents where (w_i) and (w_i, w_j) occurs is divided by the total number of documents to estimate the probabilities of single words p(w_i) or the joint probability of two words p(w_i, w_j). In part (iii), we calculate a confirmation measure ϕ to indicate how strongly W^* supports W' for every segmented pair S_i. We first calculate the similarity of W' and W^* about all the words in W. W' and W^* are represented as context vectors $\overrightarrow{v}$(W') and $\overrightarrow{v}(W^*)$ by pairing them to all words in W (Aletras & Stevenson, 2013) in the following equation:

$$\overrightarrow{v}\left(W'\right) = \left\{\sum_{w_i \in W'} NPMI\left(w_i, w_j\right)^{\gamma}\right\}_{j=1,\dots,|W|} \quad (2)$$

The NPMI (normalized pointwise mutual information) between individual words w_i and w_j is calculated as the follows:

$$NPMI\left(w_i, w_j\right)^{\gamma} = \left(\frac{log\frac{P\left(w_i, w_j\right)+\epsilon}{P\left(w_i\right)P\left(w_j\right)}}{-log\left(P\left(w_i, w_j\right)+\epsilon\right)}\right)^{\gamma} \quad (3)$$

Additionally, ϵ is used to account for the logarithm of zero and γ to place more weight on higher NPMI values. The confirmation measure ϕ of a pair S_i is obtained by calculating the cosine vector similarity of all context vectors $\phi_{S_i}(\overrightarrow{u}, \overrightarrow{w})$ within S_i. Here $\overrightarrow{v}(W') \in \overrightarrow{u}$ and $\overrightarrow{v}(W^*) \in \overrightarrow{w}$ are expressed in the following equation.

$$\phi_{S_i}\left(\overrightarrow{u}, \overrightarrow{w}\right) = \frac{\sum_{i=1}^{|W|} u_i w_i}{||\overrightarrow{u}||_2 ||\overrightarrow{w}||_2} \quad (4)$$

Part (iv) calculates the final coherence score, which is the arithmetic mean of all confirmation measures ϕ.

Figs. 5 and 6 show the obtained C_v coherence scores for all 354 LDA models created. Figs. 5 and 6 display the results for the hotel review and restaurant review datasets, respectively. The lines represent the mean coherence scores from 3 runs, where the number of topics was varied from 2 to 60. Tables 2 and 3 display the actual coherence score values for uncovered topics for hotel review and restaurant review datasets, respectively. They show the coherence scores X_1, X_2, and X_3 from 3 different runs where the random seed is 4, 5, and 6. They also show the mean C_v coherence score ($\overline{X}$) and the standard deviation(s) calculated from all three runs for $K = \{2, \dots, 60\}$. The standard deviation(s) from both Tables 2 and 3 show that the variation among the calculated coherence scores from three different runs is small. DS_1 achieved the optimal coherence score with an elbow method (the point with maximum absolute second derivative) at k = 16, and DS_2 achieved this at k = 14 (Syed & Spruit, 2017).

4.3. Topic representation and assignment

The results of the optimal LDA model on DS_1 and DS_2 are shown in Tables 4 and 5, respectively. Since LDA analysis gives the outcomes of topics β_k and topic proportions θ_d, the most frequently used ten words in the descending order of frequency are provided for each topic in the second column of Tables 4 and 5. The topic proportions for each topic are in the fourth column of Tables 4 and 5. To better summarize the

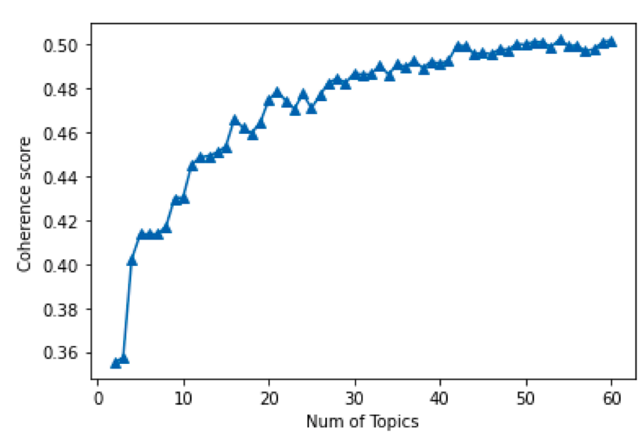

Fig. 5. Choosing the optimal LDA model with a coherence score for DS_1.

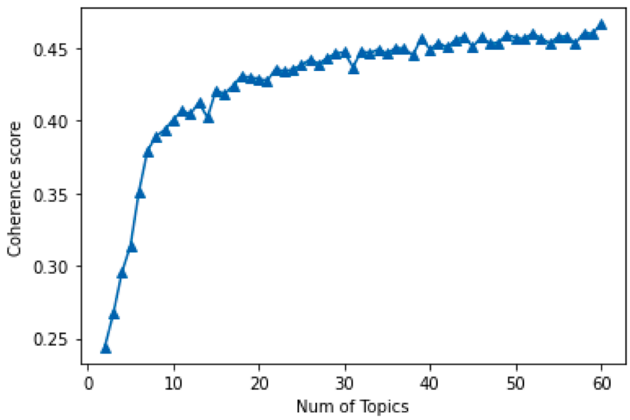

Fig. 6. Choosing the optimal LDA model with a coherence score for DS_2.

topics, we had to condense and label them instead of showing all ten words. According to Chang, Gerrish, Wang, Boyd-Graber, and Blei (2009), identifying topics is an unsupervised learning problem, and therefore automatic labeling of topics is impossible. Hence, to deliver the contents of topics more precisely, human judgments and intervention are required to assign a topic label/title, and is done based on the semantic similarity of significant words in the topic matrix (Chang et al., 2009). The TripAdvisor.com has also designed several aspects for both hotels and restaurants. For hotels, it provides the aspects 'Location', 'Cleanliness', 'Service', 'Value', which allow guests to evaluate their experience. It also displays the list of 'Property Amenities' and 'Room Features' for each hotel. For restaurants, it provides the aspects of 'Food', 'Service', 'Value', 'Atmosphere', which allow guests to evaluate their experience. These factors are predetermined labels proposed by website experts, which identify the main issues of concern by guests, and therefore we will use these factors as the criteria to evaluate the alternatives in AHP. Additionally, Parasuraman, Zeithaml, and Berry (1988) described the development of a 22-item instrument (SERVQUAL) for assessing customer perceptions of service quality in service and retailing organizations. This paper suggested the following labels for the five dimensions of service quality, and we also considered these five dimensions to determine the sub-criteria.

- Tangibles: Physical facilities, equipment, and appearance of personnel.
- Reliability: Ability to perform the promised service dependably and accurately.
- Responsiveness: Willingness to help customers and provide prompt service.
- Assurance: Knowledge and courtesy of employees and their ability to inspire trust and confidence.
- Empathy: Caring, individualized attention the firm provides its customers.

Considering the criteria/sub-criteria mentioned above and the semantic similarity of significant words in topics, we labeled topics and put them in the third column of Tables 4 and 5.

Table 2
Calculated coherence score for DS_1. K = number of topics, X_1 = coherence score where the random seed is 4, X_2 = coherence score where the random seed is 5, X_3 = coherence score where the random seed is 6, $\overline{X}$ = mean coherence score, s = standard deviation of coherence score.

K	X_1	X_2	X_3	$\overline{X}$	s
2	0.353	0.360	0.353	0.355	0.003
3	0.358	0.357	0.357	0.357	0.000
4	0.404	0.404	0.399	0.402	0.002
5	0.416	0.413	0.412	0.414	0.002
6	0.420	0.415	0.406	0.414	0.006
7	0.415	0.414	0.412	0.414	0.001
8	0.418	0.420	0.413	0.417	0.003
9	0.433	0.427	0.430	0.430	0.002
10	0.428	0.435	0.428	0.430	0.003
11	0.449	0.445	0.442	0.445	0.003
12	0.447	0.450	0.449	0.449	0.001
13	0.454	0.445	0.449	0.449	0.004
14	0.448	0.454	0.451	0.451	0.002
15	0.457	0.457	0.447	0.454	0.005
16	0.463	0.468	0.468	0.466	0.002
17	0.455	0.460	0.471	0.462	0.007
18	0.459	0.462	0.458	0.460	0.002
19	0.468	0.463	0.462	0.464	0.003
20	0.469	0.477	0.478	0.475	0.004
21	0.488	0.474	0.475	0.479	0.006
22	0.474	0.474	0.475	0.474	0.000
23	0.469	0.471	0.472	0.471	0.001
24	0.479	0.476	0.478	0.478	0.001
25	0.466	0.479	0.468	0.471	0.006
26	0.480	0.481	0.470	0.477	0.005
27	0.483	0.485	0.480	0.483	0.002
28	0.483	0.495	0.476	0.485	0.008
29	0.482	0.485	0.481	0.483	0.002
30	0.488	0.489	0.484	0.487	0.002
31	0.493	0.481	0.485	0.487	0.005
32	0.484	0.493	0.483	0.487	0.004
33	0.488	0.491	0.493	0.491	0.002
34	0.486	0.489	0.484	0.486	0.002
35	0.494	0.491	0.488	0.491	0.002
36	0.487	0.495	0.488	0.490	0.004
37	0.491	0.487	0.500	0.493	0.005
38	0.484	0.494	0.490	0.489	0.004
39	0.492	0.491	0.494	0.492	0.001
40	0.493	0.491	0.490	0.491	0.001
41	0.496	0.489	0.493	0.493	0.003
42	0.501	0.496	0.502	0.500	0.003
43	0.496	0.502	0.500	0.500	0.002
44	0.494	0.494	0.499	0.496	0.002
45	0.497	0.494	0.498	0.496	0.002
46	0.496	0.494	0.497	0.496	0.001
47	0.496	0.498	0.500	0.498	0.002
48	0.496	0.500	0.497	0.498	0.002
49	0.495	0.502	0.504	0.500	0.004
50	0.502	0.498	0.500	0.500	0.002
51	0.503	0.503	0.498	0.501	0.002
52	0.504	0.503	0.496	0.501	0.004
53	0.500	0.502	0.495	0.499	0.003
54	0.505	0.503	0.499	0.502	0.002
55	0.500	0.499	0.499	0.499	0.000
56	0.501	0.500	0.498	0.500	0.001
57	0.496	0.498	0.497	0.497	0.000
58	0.495	0.501	0.498	0.498	0.002
59	0.501	0.503	0.499	0.501	0.002
60	0.500	0.506	0.500	0.502	0.003

4.4. Determination of criteria, weights, and alternatives in AHP

This step is crucial for any hotel or restaurant selection problem. From the topic modeling results of 100,000 reviews for DS_1 and DS_2, the criteria, sub-criteria, and weights are given in columns 3–4 of Tables 4 and 5. The six criteria considered in hotel selection are 'Property Amenities', 'Value', 'Service', 'Cleanliness', 'Room Features', and 'Location'. Also, some criteria have sub-criteria in the next level. For example, 'Property Amenities' has the sub-criteria 'Restaurant & Bar', 'Breakfast & Pool', 'Airport Shuttle', 'Valet Parking', and 'Front Desk'. The four criteria considered in restaurant selection are 'Food', 'Service', 'Value', and 'Atmosphere'. Based on the determined criteria and sub-criteria, two decision hierarchies are developed as demonstrated for hotel and restaurant selection in Figs. 7 and 8.

To calculate each criterion's weight, we first check the decision hierarchy to determine whether this criterion has any sub-criteria. Take the hotel selection as one example, the criterion 'Property Amenities' has several sub-criteria. Therefore the weight of 'Property Amenities' is the sum of weights for each sub-criteria. Then by checking the last

Table 3
Calculated coherence score for DS_2. K = number of topics, X_1 = coherence score where the random seed is 4, X_2 = coherence score where the random seed is 5, X_3 = coherence score where the random seed is 6, $\overline{X}$ = mean coherence score, s = standard deviation of coherence score.

K	X_1	X_2	X_3	$\overline{X}$	s
2	0.244	0.244	0.244	0.244	0.000
3	0.282	0.252	0.267	0.267	0.012
4	0.307	0.281	0.298	0.295	0.011
5	0.319	0.319	0.303	0.314	0.008
6	0.376	0.334	0.343	0.351	0.018
7	0.379	0.398	0.361	0.379	0.015
8	0.379	0.385	0.403	0.389	0.010
9	0.401	0.394	0.386	0.394	0.006
10	0.405	0.397	0.399	0.400	0.003
11	0.405	0.404	0.411	0.407	0.003
12	0.402	0.405	0.406	0.404	0.002
13	0.416	0.409	0.412	0.412	0.003
14	0.407	0.405	0.394	0.402	0.006
15	0.414	0.422	0.426	0.421	0.005
16	0.431	0.414	0.410	0.418	0.009
17	0.425	0.420	0.427	0.424	0.003
18	0.429	0.436	0.427	0.431	0.004
19	0.427	0.434	0.428	0.430	0.003
20	0.433	0.429	0.424	0.429	0.004
21	0.427	0.413	0.442	0.427	0.012
22	0.435	0.431	0.441	0.436	0.004
23	0.438	0.431	0.432	0.434	0.003
24	0.435	0.441	0.431	0.435	0.004
25	0.430	0.436	0.448	0.438	0.007
26	0.446	0.440	0.440	0.442	0.003
27	0.443	0.436	0.438	0.439	0.003
28	0.447	0.437	0.445	0.443	0.004
29	0.445	0.446	0.447	0.446	0.001
30	0.442	0.449	0.450	0.447	0.004
31	0.430	0.440	0.440	0.437	0.005
32	0.451	0.443	0.449	0.448	0.003
33	0.440	0.442	0.457	0.446	0.008
34	0.455	0.445	0.447	0.449	0.004
35	0.449	0.440	0.449	0.446	0.004
36	0.449	0.448	0.453	0.450	0.002
37	0.451	0.447	0.450	0.449	0.002
38	0.447	0.446	0.442	0.445	0.002
39	0.456	0.452	0.461	0.456	0.004
40	0.449	0.451	0.447	0.449	0.002
41	0.450	0.457	0.453	0.453	0.003
42	0.440	0.448	0.463	0.450	0.010
43	0.461	0.449	0.456	0.455	0.005
44	0.453	0.460	0.460	0.458	0.003
45	0.446	0.452	0.454	0.451	0.003
46	0.458	0.464	0.450	0.457	0.006
47	0.450	0.460	0.448	0.453	0.005
48	0.453	0.453	0.455	0.454	0.001
49	0.450	0.462	0.465	0.459	0.006
50	0.459	0.458	0.454	0.457	0.002
51	0.451	0.463	0.454	0.456	0.005
52	0.458	0.464	0.458	0.460	0.003
53	0.457	0.459	0.453	0.456	0.002
54	0.451	0.451	0.458	0.453	0.003
55	0.459	0.456	0.457	0.457	0.001
56	0.461	0.461	0.452	0.458	0.004
57	0.455	0.458	0.446	0.453	0.005
58	0.463	0.465	0.453	0.460	0.005
59	0.458	0.459	0.462	0.460	0.002
60	0.473	0.462	0.464	0.466	0.005

column of Table 4, the weight of the criteria 'Property Amenities' is 0.0282 (the weight of 'Restaurant & Bar') + 0.0373 (the weight of 'Breakfast & Pool') + 0.0368 (the weight of 'Airport Shuttle') + 0.0408 (the weight of 'Valet Parking') + 0.0809 (the weight of 'Front Desk') = 0.2240. For the criteria that do not have any sub-criteria such as 'Cleanliness', we directly use the weight of 'Cleanliness' from the last column of Table 4, which is 0.0671.

In this research, Philadelphia has been chosen as the pilot region. Only four-star hotels or restaurants with more than 1000 reviews in

Table 4
Extracted topics for DS_1 using LDA.

ID	Topics (β_k)	Label	Proportions (θ_d)
1	Stay hotel room property good service philadelphia price marriott location	Price	0.0546
2	Staff front_desk make great service helpful friendly time check experience	Responsiveness	0.0816
3	Bed bathroom large comfortable small nice shower suite room area	Room Features	0.0547
4	Bathroom dirty bed clean floor carpet shower smell bad stain	Cleanliness	0.0671
5	Staff great clean good location nice room friendly comfortable restaurant	Empathy	0.1800
6	Breakfast food good order coffee service restaurant eat day bar	Restaurant & Bar	0.0282
7	Nice free good lobby coffee breakfast pool clean water small	Breakfast & Pool	0.0373
8	Stay great staff clean location philly time friendly night family	Reliability	0.0765
9	Airport shuttle night flight clean check stay good free minute	Airport Shuttle	0.0368
10	Street market location reading_terminal convention_center walk great close restaurant block	Close to Attractions	0.0455
11	Parking car park valet lot night check pay street nice	Valet Parking	0.0408
12	Breakfast stay home house philadelphia inn comfortable penn lovely wonderful	Assurance	0.0336
13	Great location restaurant walk city distance philadelphia view staff within_walk	Close to Business	0.0768
14	Great service love staff beautiful stay lobby view bar restaurant	Tangibles	0.0608
15	Check call front_desk arrive day book time night reservation wait	Front desk	0.0809
16	Night elevator noise sleep door room hear loud floor people	Quiet	0.0448

Table 5
Extracted topics for DS_2 using LDA.

ID	Topics (β_k)	Label	Proportions (θ_d)
1	City_tavern history philadelphia tavern great visit experience lunch time historic	History	0.0804
2	Review enjoy glad hope experience hear time visit feedback service	Assurance	0.0959
3	Hotel good stay walk philadelphia street lunch philly find decide	Location	0.0498
4	Pizza good order chicken great delicious dish sushi fresh noodle	Food Variety	0.1093
5	Good service price order bad average give time quality bit	Price	0.0245
6	Table reservation wait seat dinner arrive order time bar sit	Responsiveness	0.0627
7	Service great friendly staff good excellent atmosphere dinner delicious attentive	Empathy	0.0428
8	Good delicious salad order appetizer excellent dinner great chicken dish	Food Quality	0.0603
9	Sandwich good philly cheesesteak order meat cheese_steak great steak cheese	Cheesesteak Sandwich	0.1507
10	Visit philly philadelphia year good time dinner friend wife eat	Visit	0.0831
11	Menu good dish great service excellent delicious choice option item	Menu Options	0.0726
12	Breakfast good brunch great coffee delicious service lunch egg order	Brunch	0.0402
13	Great good beer bar drink nice service selection atmosphere friendly	Bar	0.0677
14	table great bar atmosphere decor nice area room view beautiful	Decoration	0.0600

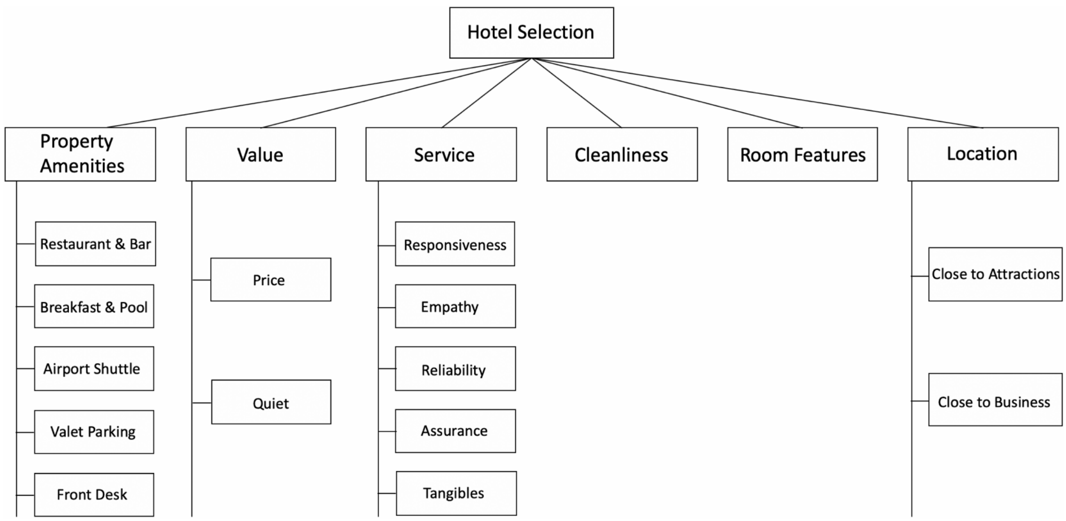

Fig. 7. Decision hierarchy for hotel selection.

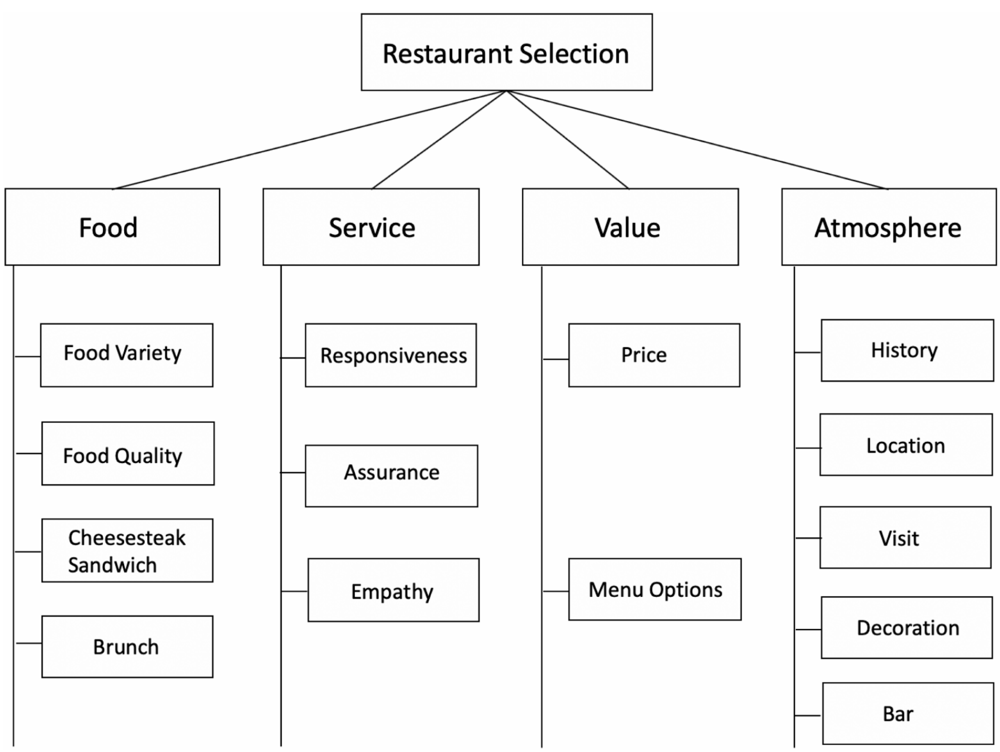

Fig. 8. Decision hierarchy for restaurant selection.

Philadelphia are considered in the hotel or restaurant selection. They are determined as alternatives in the AHP, which are shown in Tables 6 and 15. We also provided the overall bubble rating from TripAdvisor.com for all hotels in the third column of Tables 6 and 15. These bubble ratings are based on the average of ratings from the reviewers that each hotel has received. Reviewers are asked “Your overall rating of this property from 1 bubble meaning terrible to 5 bubbles meaning excellent.” Bubble rating is an overall score that considers the quality, quantity, and age of individual travelers.

4.5. Hotel and restaurant rankings via AHP

After the hotel/restaurant selection’s hierarchy structure has been constructed, the prioritization procedure determines hotels/restaurants’

Table 6
Appointed alternatives for the case of four-star hotels in Philadelphia.

Label	Hotel name	Overall bubble rating
A1	The Windsor Suites	4.5
A2	Kimpton Hotel Monaco Philadelphia	4.5
A3	Kimpton Hotel Palomar Philadelphia	4.5
A4	Sofitel Philadelphia at Rittenhouse Square	4.5
A5	Loews Philadelphia Hotel	4.5
A6	The Inn at Penn, A Hilton Hotel	4.5
A7	The Logan Philadelphia, Curio Collection by Hilton	4.5
A8	Sonesta Philadelphia Downtown	4.5
A9	Le Meridien Philadelphia	4.0
A10	The Bellevue Hotel	4.0
A11	Warwick Hotel Rittenhouse Square	4.0
A12	Hilton Philadelphia City Avenue	3.5

Table 7
Rating for the case of four-star hotels in Philadelphia.

Label	Value	Service	Cleanliness	Location	Property amenities	Room features
A1	4	4.5	4.5	4.5	28	10
A2	5	4.5	5	5	32	7
A3	4.5	4.5	5	5	32	7
A4	4	4.5	4.5	5	32	10
A5	4	4.5	4.5	4.5	42	9
A6	4	4.5	4.5	4.5	28	8
A7	4	4.5	4.5	5	37	7
A8	4	4.5	4.5	4.5	33	6
A9	4	4.5	4.5	4.5	26	5
A10	4	4.5	4.5	4.5	27	6
A11	4	4	4.5	4.5	33	5
A12	3.5	3.5	4	4	30	4

relative importance. All of the alternatives are evaluated by reviewers from TripAdvisor, and therefore this is a group decision-making analysis. For each alternative in hotel selection, TripAdvisor provides the bubble rating according to value, service, cleanliness, and location, shown in columns 2–5 in Table 7. TripAdvisor also provides the list of property amenities and room features for each hotel, and we count the number of them in columns 6 and 7 of Table 7. For each alternative in restaurant selection, TripAdvisor provides the bubble rating according to food, service, value, and atmosphere, shown in columns 4–7 in Table 15.

To rank the alternatives in hotel selection listed in Table 6, six pairwise comparison matrices are used to estimate the relative quality of hotels in Philadelphia from Tables 8–13. Let the pairwise comparison n-by-n matrix be $A = (a_{ij})$, i, j $= 1, 2, \ldots$, n. The entry a_{ij} is defined by $\frac{w_i}{w_j}$. Here w_i and w_j are weights of alternative i and alternative j regarding certain criteria. To make the comparison, 12 hotels are listed on the left and at the top, and the judgement is made to how strongly hotels on the left dominate hotels on the top. In Table 8, when hotel A1 on the left is compared with hotel A2 at the top concerning the criteria 'Value', we first checked the Table 7, and the ratings of 'Value' for A1 and A2 are 4 and 5, respectively. Therefore, a 4/5 is entered in the first row and second column position of Table 8. A 5/4 is automatically entered in the second row and first column position of Table 8. In this case, the quality of hotel A1 concerning the criteria 'Value' on the left does not dominate that of the hotel A2 on the top. To obtain each table's priorities, we first compute each matrix's principal right eigenvector and then normalize its entries by dividing by their sum (Saaty, 1988). All preference matrices from Tables 8–13 are consistent since we use quantitative values. Then in Table 14,15, we obtain the priorities by calculating the weighted sum score. For example, the priorities of A1 $= 0.0994$ (the weight of 'Value') $\times$ 0.08163265 $+ 0.4325$ (the weight of 'Service') $\times$ 0.08571429 $+ 0.0671$ (the weight of 'Cleanliness') $\times$ 0.08256881 $+ 0.1223$ (the weight of 'Location') $\times$ 0.08108108 $+ 0.2240$ (the weight of 'Property Amenities') $\times$ 0.07368421 $+ 0.0547$ (the weight of 'Room Features') $\times$ 0.11904762 $= 0.08365947$. Then we normalize the priorities by dividing their sum to obtain the totals for each alternative.

In Table 14, the rankings of alternatives from our proposed model in hotel selection are given. We find they are in the exact order as the overall bubble rating obtained from TripAdvisor (Table 7). This means our analysis outcome is highly accurate compared with reviewers' opinions about hotels. Based on the reviews' overall rating about each hotel in column 3 of Table 6, A1–A8 (Group 1) are rated as 4.5, A9–A11 (Group 2) are rated as 4.0 and A12 (Group 3) is rated as 3.5. The proposed model gives the result in column 2 of Table 14. It ranks Group 1 above Group 2, and Group 3 is ranked as the lowest. The only difference is that our model is more detailed within each group, while the bubble ratings from Table 6 show all hotels from the same group as equal. For example, we ranked the Loews Philadelphia Hotel (A5) as the best hotel. By checking Table 7, we find that the number of property amenities in this hotel is 42, which is the largest number compared with other hotels.

Table 8
Pairwise comparison matrix for the alternatives with respect to value.

Value	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Priorities
A1	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A2	5/4	1	5/4.5	5/4	5/4	5/4	5/4	5/4	5/4	5/4	5/4	5/3.5	0.10204082
A3	4.5/4	4.5/5	1	4.5/4	4.5/4	4.5/4	4.5/4	4.5/4	4.5/4	4.5/4	4.5/4	4.5/3.5	0.09183673
A4	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A5	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A6	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A7	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A8	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A9	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A10	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A11	1	4/5	4/4.5	1	1	1	1	1	1	1	1	4/3.5	0.08163265
A12	3.5/4	3.5/5	3.5/4.5	3.5/4	3.5/4	3.5/4	3.5/4	3.5/4	3.5/4	3.5/4	3.5/4	1	0.07142857

Note: Consistency ratio $= 0.000$.

Table 9
Pairwise comparison matrix for the alternatives with respect to service.

Service	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Priorities
A1	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A2	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A3	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A4	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A5	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A6	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A7	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A8	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A9	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A10	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/3.5	0.08571429
A11	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	1	4/3.5	0.07619048
A12	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4.5	3.5/4	1	0.06666667

Note: Consistency ratio $= 0.000$.

Table 10
Pairwise comparison matrix for the alternatives with respect to cleanliness.

Cleanliness	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Priorities
A1	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A2	5/4.5	1	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.09174312
A3	5/4.5	1	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.09174312
A4	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A5	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A6	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A7	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A8	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A9	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A10	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A11	1	4.5/5	4.5/5	1	1	1	1	1	1	1	1	4.5/4	0.08256881
A12	4/4.5	4/5	4/5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	1	0.0733945

Note: Consistency ratio $= 0.000$.

Table 11
Pairwise comparison matrix for the alternatives with respect to location.

Location	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Priorities
A1	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A2	5/4.5	1	1	1	5/4.5	5/4.5	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.09009009
A3	5/4.5	1	1	1	5/4.5	5/4.5	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.09009009
A4	5/4.5	1	1	1	5/4.5	5/4.5	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.09009009
A5	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A6	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A7	5/4.5	1	1	1	5/4.5	5/4.5	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.09009009
A8	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A9	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A10	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A11	1	4.5/5	4.5/5	4.5/5	1	1	4.5/5	1	1	1	1	4.5/4	0.08108108
A12	4/4.5	4/5	4/5	4/5	4/4.5	4/4.5	4/5	4/4.5	4/4.5	4/4.5	4/4.5	1	0.07207207

Note: Consistency ratio $= 0.000$.

Table 12
Pairwise comparison matrix for the alternatives with respect to property amenities.

Property amenities	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Priorities
A1	1	28/32	28/32	28/32	28/42	1	28/37	28/33	28/26	28/27	28/33	28/30	0.07368421
A2	32/28	1	1	1	32/42	32/28	32/37	32/33	32/26	32/27	32/33	32/30	0.08421053
A3	32/28	1	1	1	32/42	32/28	32/37	32/33	32/26	32/27	32/33	32/30	0.08421053
A4	32/28	1	1	1	32/42	32/28	32/37	32/33	32/26	32/27	32/33	32/30	0.08421053
A5	42/28	42/32	42/32	42/32	1	42/28	42/37	42/33	42/26	42/27	42/33	42/30	0.11052632
A6	1	28/32	28/32	28/32	28/42	1	28/37	28/33	28/26	28/27	28/33	28/30	0.07368421
A7	37/28	37/32	37/32	37/32	37/42	37/28	1	37/33	37/26	37/27	37/33	37/30	0.09736842
A8	33/28	33/32	33/32	33/32	33/42	33/28	33/37	1	33/26	33/27	1	33/30	0.08684211
A9	26/28	26/32	26/32	26/32	26/42	26/28	26/37	26/33	1	26/27	26/33	26/30	0.06842105
A10	27/28	27/32	27/32	27/32	27/42	27/28	27/37	27/33	27/26	1	27/33	27/30	0.07105263
A11	33/28	33/32	33/32	33/32	33/42	33/28	33/37	1	33/26	33/27	1	33/30	0.08684211
A12	30/28	30/32	30/32	30/32	30/42	30/28	30/37	30/33	30/26	30/27	30/33	1	0.07894737

Note: Consistency ratio $= 0.000$.

Table 13
Pairwise comparison matrix for the alternatives with respect to room features.

Room features	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Priorities
A1	1	10/7	10/7	1	10/9	10/8	1	10/6	10/5	10/6	10/5	10/4	0.11904762
A2	7/10	1	1	7/10	7/9	7/8	1	7/6	7/5	7/6	7/5	7/4	0.08333333
A3	7/10	1	1	7/10	7/9	7/8	1	7/6	7/5	7/6	7/5	7/4	0.08333333
A4	1	10/7	10/7	1	10/9	10/8	10/7	10/6	10/5	10/6	10/5	10/4	0.11904762
A5	9/10	9/7	9/7	9/10	1	9/8	9/7	9/6	9/5	9/6	9/5	9/4	0.10714286
A6	8/10	8/7	8/7	8/10	8/9	1	8/7	8/6	8/5	8/6	8/5	8/4	0.0952381
A7	7/10	1	1	7/10	7/9	7/8	1	7/6	7/5	7/6	7/5	7/4	0.08333333
A8	6/10	6/7	6/7	6/10	6/9	6/8	6/7	1	6/5	1	6/5	6/4	0.07142857
A9	5/10	5/7	5/7	5/10	5/9	5/8	5/7	5/6	1	5/6	1	5/4	0.05952381
A10	6/10	6/7	6/7	6/10	6/9	6/8	6/7	1	6/5	1	6/5	6/4	0.07142857
A11	5/10	5/7	5/7	5/10	5/9	5/8	5/7	5/6	1	5/6	1	5/4	0.05952381
A12	4/10	4/7	4/7	4/10	4/9	4/8	4/7	4/6	4/5	4/6	4/5	1	0.04761905

Note: Consistency ratio $= 0.000$.

Table 14
Numerical values for ratings of the alternatives in hotel selection.

	Ranking	Totals	Criterion priorities	0.0994	0.4325	0.0671	0.1223	0.2240	0.0547
			Priorities	Value	Service	Cleanliness	Location	Property amenities	Room features
A1	7	0.08365946	0.08365947	0.08163265	0.08571429	0.08256881	0.08108108	0.07368421	0.11904762
A2	3	0.08780976	0.08780976	0.10204082	0.08571429	0.09174312	0.09009009	0.08421053	0.08333333
A3	5	0.08679547	0.08679547	0.09183673	0.08571429	0.09174312	0.09009009	0.08421053	0.08333333
A4	4	0.08711916	0.08711916	0.08163265	0.08571429	0.08256881	0.09009009	0.08421053	0.11904762
A5	1	0.09126091	0.09126091	0.08163265	0.08571429	0.08256881	0.08108108	0.11052632	0.10714286
A6	8	0.08235708	0.08235709	0.08163265	0.08571429	0.08256881	0.08108108	0.07368421	0.0952381
A7	2	0.08811296	0.08811296	0.08163265	0.08571429	0.08256881	0.09009009	0.09736842	0.08333333
A8	6	0.08400207	0.08400207	0.08163265	0.08571429	0.08256881	0.08108108	0.08684211	0.07142857
A9	11	0.07922456	0.07922457	0.08163265	0.08571429	0.08256881	0.08108108	0.06842105	0.05952381
A10	9	0.08046523	0.08046523	0.08163265	0.08571429	0.08256881	0.08108108	0.07105263	0.07142857
A11	10	0.07923183	0.07923184	0.08163265	0.07619048	0.08256881	0.08108108	0.08684211	0.05952381
A12	12	0.06996149	0.06996149	0.07142857	0.06666667	0.0733945	0.07207207	0.07894737	0.04761905

Table 15
Appointed alternatives for the case of restaurants in Philadelphia.

Label	Restaurant name	Overall bubble rating	Food	Service	Value	Atmosphere
B1	Ristorante Pesto	5.0	5.0	5.0	5.0	4.5
B2	Zahav	4.5	4.5	4.5	4.0	4.5
B3	Talula's Garden	4.5	4.5	4.5	4.0	4.5
B4	The Dandelion	4.5	4.5	4.5	4.0	4.5
B5	Parc Brasserie	4.5	4.5	4.5	4.0	4.5
B6	City Tavern	4.5	4.5	4.5	4.0	4.5
B7	Buddakan	4.5	4.5	4.5	4.0	4.5
B8	Amada	4.5	4.5	4.5	4.0	4.5
B9	El Vez	4.5	4.5	4.5	4.0	4.5
B10	Fogo de Chão Brazilian Steakhouse	4.5	4.5	4.5	4.0	4.5
B11	Maggiano's Little Italy	4.0	4.5	4.5	4.0	4.0
B12	Red Owl Tavern	4.0	4.0	4.5	4.0	4.0

Also, 'Property Amenities' weights approximately 22% of the evaluation process based on Table 14. Simultaneously, the rating of service for A5 is 4.5, which is also the highest among all alternatives. From Table 14, 'Service' weights approximately 43% of the evaluation process. Therefore, we ranked Loews Philadelphia Hotel in the first place. In contrast, Warwick Hotel Rittenhouse Square (A11), Le Meridien Philadelphia (A9), and Hilton Philadelphia City Avenue (A12) are placed in the bottom.

Similar calculations are performed to rank the alternatives in restaurant selection, which are listed in Table 15. Four pairwise comparison matrices are constructed from Tables 16–19 to evaluate the restaurants with the criteria 'Food', 'Service', 'Value', and 'Atmosphere'. Table 20 displays the rankings of alternatives in restaurant selection, which are in the exact order as the overall bubble ratings from TripAdvisor in Table 15. Based on the reviews' overall rating about each restaurant in column 3 of Table 15, B1 (Group 1) is rated as 5.0, B2–B10 (Group 2) are rated as 4.5 and B11–B12 (Group 3) are rated as 4.0. The proposed model gives the rankings in column 2 of Table 20. It ranks Group 1 above Group 2, and Group 3 is ranked as the lowest. The significant difference between the ranking outcomes of restaurant selection and hotel selection is that we cannot differentiate between B2 and B10 (Group 2) in the restaurant selection, while we obtain a complete ranking list for hotel selection. The reason is that the ratings from reviewers on TripAdvisor are the same for all four criteria 'Food', 'Service', 'Value', and 'Atmosphere' of the restaurants B2–B10 (Group 2). Therefore, by implementing the AHP, we cannot distinguish between the restaurants B2–B10 (Group 2). However, we can still differentiate between the restaurants B11–B12 (Group 3), and we rank B11 above B12. From Table 15, although the overall rating of B11 and B12 is the same, the rating of B11 regarding 'Food' and 'Service' are higher compared with B12. Therefore, the ranking outcome of our proposed model makes sense.

5. Conclusion

An intelligent system based on LDA was proposed in this paper to determine criteria/sub-criteria and their corresponding weights in AHP. LDA generates semantically meaningful topics to summarize reviews from TripAdvisor.com into a mixture of topics that would not be possible by human judgments. We picked twelve four-star hotels with more than 1000 reviews in Philadelphia as alternatives for the hotel selection. We obtained the ratings of value, service, cleanliness, location, the number of property amenities, and room features for all alternatives. By applying AHP, we found the ranking list of twelve four-star hotels in Philadelphia.

Table 16
Pairwise comparison matrix for the alternatives with respect to food.

Food	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Priorities
B1	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4	0.08547009
B2	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B3	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B4	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B5	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B6	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B7	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B8	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B9	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B10	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B11	4.5/5	1	1	1	1	1	1	1	1	1	1	4.5/4	0.07692308
B12	4/5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	1	0.06837607

Note: Consistency ratio = 0.000.

Table 17
Pairwise comparison matrix for the alternatives with respect to service.

Service	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Priorities
B1	1	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	5/4.5	0.08547009
B2	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B3	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B4	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B5	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B6	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B7	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B8	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B9	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B10	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B11	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308
B12	4.5/5	1	1	1	1	1	1	1	1	1	1	1	0.07692308

Note: Consistency ratio $= 0.000$.

Table 18
Pairwise comparison matrix for the alternatives with respect to value.

Value	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Priorities
B1	1	5/4	5/4	5/4	5/4	5/4	5/4	5/4	5/4	5/4	5/4	5/4	0.09433962
B2	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B3	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B4	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B5	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B6	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B7	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B8	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B9	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B10	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B11	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717
B12	4/5	1	1	1	1	1	1	1	1	1	1	1	0.0754717

Note: Consistency ratio $= 0.000$.

Table 19
Pairwise comparison matrix for the alternatives with respect to atmosphere.

Atmosphere	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Priorities
B1	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B2	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B3	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B4	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B5	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B6	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B7	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B8	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B9	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B10	1	1	1	1	1	1	1	1	1	1	4.5/4	4.5/4	0.07964602
B11	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	1	1	0.07079646
B12	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	4/4.5	1	1	0.07079646

Note: Consistency ratio $= 0.000$.

Table 20
Numerical values for ratings of the alternatives in restaurant selection.

	Ranking	Totals	Criterion priorities	0.3605	0.2014	0.0971	0.3410
			Priorities	Food	Service	Value	Atmosphere
B1	1	0.08434531	0.08434531	0.08547009	0.08547009	0.09433962	0.07964602
B2	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B3	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B4	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B5	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B6	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B7	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B8	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B9	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B10	2	0.07771067	0.07771067	0.07692308	0.07692308	0.0754717	0.07964602
B11	11	0.07469297	0.07469297	0.07692308	0.07692308	0.0754717	0.07079646
B12	12	0.07161177	0.07161178	0.06837607	0.07692308	0.0754717	0.07079646

Compared with ratings found online, our proposed model's ranking outcome is highly accurate and can also distinguish between hotels. Based on the result, Loews Philadelphia Hotel was ranked first due to its relatively high number of property amenities and the highest rating of service. For the restaurant selection, we picked twelve restaurants with more than 1000 reviews in Philadelphia as alternatives. We obtained the rating of food, service, value, and atmosphere for all alternatives. By applying AHP, we obtained the ranking list of the twelve restaurants mentioned above in Philadelphia. Compared with ratings found online, our proposed model's ranking outcome is also highly accurate and can also distinguish between restaurants in Group 3.

The proposed approach can be useful in analyzing human judgments in multi-criteria group decision-making problems. This approach is not intended to replace experts, but it is designed to be a supplemental tool to identify additional criteria and corresponding weights. Also, it narrows and prioritizes seemingly equal alternative ratings. The proposed intelligent system is not specifically limited to the usage of TripAdvisor. Similar techniques could be utilized to analyze reviews in various directions, such as product reviews, movie reviews, and many other multi-criteria decision-making scenarios.

CRediT authorship contribution statement

Jin Fang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Fariborz Y. Partovi:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Albayrak, E., & Erensal, Y. C. (2004). Using analytic hierarchy process (ahp) to improve human performance: An application of multiple criteria decision making problem. *Journal of Intelligent Manufacturing, 15*(4), 491–503.

Aletras, N., & Stevenson, M. (2013). Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th international conference on computational semantics (iwcs 2013)–long papers* (pp. 13–22).

Badri, M. A. (2001). A combined ahp–gp model for quality control systems. *International Journal of Production Economics, 72*(1), 27–40.

Barzilai, J., & Golany, B. (1994). Ahp rank reversal, normalization and aggregation rules. *INFOR: Information Systems and Operational Research, 32*(2), 57–64.

Belton, V., & Gear, T. (1983). On a short-coming of saaty's method of analytic hierarchies. *Omega, 11*(3), 228–230.

Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic topic models: A focus on graphical model design and applications to document and image analysis. *IEEE Signal Processing Magazine, 27*(6), 55.

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

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Casalo, L. V., Flavian, C., Guinaliu, M., & Ekinci, Y. (2015). Do online hotel rating schemes influence booking behaviors? *International Journal of Hospitality Management, 49*, 28–36.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., & Blei, D.M. (2009). Reading tea leaves: How humans interpret topic models. In Advances in neural information processing systems (pp. 288–296).

de FSM Russo, R. & Camanho, R. (2015). Criteria in ahp: a systematic review of literature. Procedia Computer Science, 55, 1123–1132.

Escobar, M. T., & Moreno-JimenTz, J. M. (2007). Aggregation of individual preference structures in ahp-group decision making. *Group Decision and Negotiation, 16*(4), 287–301.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5228–5235.

Ho, W. (2008). Integrated analytic hierarchy process and its applications–a literature review. *European Journal of Operational Research, 186*(1), 211–228.

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications, 38*(11), 14336–14345.

Joakim, C. (2012). *Explore python, machine learning, and the nltk library*. IBM Developer Works.

Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment analysis of restaurant customer reviews on tripadvisor using nanve bayes. In *2019 12th international conference on information & communication technology and system (icts)* (pp. 49–54).

Lei, S., & Law, R. (2015). Content analysis of tripadvisor reviews on restaurants: A case study of macau. *Journal of Tourism, 16*(1).

Linstone, H. A., Turoff, M., et al. (1975). *The delphi method*. MA: Addison-Wesley Reading.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. (http://mallet.cs.umass.edu).

Mikhailov, L., & Tsvetinov, P. (2004). Evaluation of services using a fuzzy analytic hierarchy process. *Applied Soft Computing, 5*(1), 23–33.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111–3119).

Mimno, D. (2013). Machine learning with mallet.

Minka, T. P. & Lafferty, J. (2012). Expectation-propogation for the generative aspect model. arXiv preprint arXiv:1301.0588.

Nielsen, A. (2012). Global trust in advertising and brand messages. Report available at Nielsen.com.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perc. *Journal of Retailing, 64*(1), 12.

Partovi, F. Y., Burton, J., & Banerjee, A. (1990). Application of analytical hierarchy process in operations management. *International Journal of Operations & Production Management, 10*(3), 5–19.

Partovi, F. Y., & Corredoira, R. A. (2002). Quality function deployment for the good of soccer. *European Journal of Operational Research, 137*(3), 642–656.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth acm international conference on web search and data mining* (pp. 399–408).

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

Saaty, T.L. (1980). The analytic hierarchy process mcgraw-hill. New York, 324.

Saaty, T. L. (1988). What is the analytic hierarchy process?. In *Mathematical models for decision support* (pp. 109–121). Springer.

Saaty, T. L. (1996). *Decision making with dependence and feedback: The analytic network process* (Vol. 4922). RWS Publ.

Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. *Mathematical and Computer Modelling, 38*(3–4), 233–244.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis, 427*(7), 424–440.

Syed, S., & Spruit, M. (2017). Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In *2017 IEEEE international conference on data science and advanced analytics (DSAA)* (pp. 165–174).

Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. (2005). Sharing clusters among related groups: Hierarchical dirichlet processes. In Advances in neural information processing systems (pp. 1385–1392).

Van Laarhoven, P. J., & Pedrycz, W. (1983). A fuzzy extension of saaty's priority theory. *Fuzzy Sets and Systems, 11*(1–3), 229–241.

Vidal, L.-A., Marle, F., & Bocquet, J.-C. (2011). Using a delphi process and the analytic hierarchy process (ahp) to evaluate the complexity of projects. *Expert Systems with Applications, 38*(5), 5388–5405.

Yang, C.-L., Chuang, S.-P., & Huang, R.-H. (2009). Manufacturing evaluation system based on ahp/anp approach for wafer fabricating industry. *Expert Systems with Applications, 36*(8), 11369–11377.

Yu, S.-M., Wang, J., & Wang, J.-Q. (2018). An extended todim approach with intuitionistic linguistic numbers. *International Transactions in Operational Research, 25*(3), 781–805.

Zaman, M., Botti, L., & Thanh, T. V. (2016). Weight of criteria in hotel selection: An empirical illustration based on tripadvisor criteria. *European Journal of Tourism Research, 13*(1), 132–138.

Zhang, H.-Y., Ji, P., Wang, J.-Q., & Chen, X.-H. (2017). A novel decision support model for satisfactory restaurants utilizing social information: A case study of tripadvisor.com. *Tourism Management, 59*, 281–297.