AAAI Technical Report SS-12-04
Intelligent Web Services Meet Social Computing

# Priorities-Based Review Computation

**Gianpiero Costantino** and **Fabio Martinelli** and **Marinella Petrocchi**
Istituto di Informatica e Telematica - Consiglio Nazionale delle Ricerche
via G.Moruzzi 1, 56124, Pisa, Italy

**Abstract**

Recently, online vendors and providers manage review systems as a mechanism to advertise their services and goods over the Web. In making their choice, clients can rely on feedback expressing the degree of satisfaction of past users with respect to such services and goods. This set of feedback, or reviews, may be filtered by categories of users, they may be affected by multiple factors, and they are elaborated in order to obtain an overall score, representing a global indicator aimed at summarising the level of quality of that service. In this paper, we concentrate on multi-factor review, *i.e.,* a review whose value is computed assembling the scores given to a set of parameters that may affect the quality level of a service. Our interest is evaluating the relevance, or dominance, of some parameter with respect to the others. We advocate the use of the Analytic Hierarchy Process, a well-known technique born in the area of multi-criteria decision making, to derive the priorities to assign to the scores of the single parameters. We illustrate the proposal on the example of hotel reviews.

## 1 Introduction

Making a choice on the Web is commonly supported by reviews and feedback that users post on the Internet to report their personal experience towards products and/or services. A review set is usually elaborated to obtain an overall rating that identifies the degree of satisfaction that the users have met towards the product or service. The subsequent users can rely on that rating for comparing two or more items and making a choice. The overall rating could be the result of the evaluation of multiple parameters. For example, popular websites giving traveller advices, such as Booking.com, allow customers to review hotels based on, *e.g.,* cleanness, staff, and comfort. The overall rating results from the elaboration of the scores attached to the single parameters. The simplest way to elaborate such scores is to make their plain average.

However, being posted by humans, a review may be influenced by subjective attitudes and preferences. Thus, different typologies of users might attach a different importance to a given parameter. For instance, when reviewing hotels, cleanness could be more important than staff friendliness for families with young children. Analogously, on a scale of five stars, a businessman used to travel with high-standards airlines, with as few delays as possible, might rate the punctuality of an airline with one flight delayed over ten as two stars. However, a backpack traveller used to travel with budget airlines might give four stars for the punctuality of the same airline. Also, individuals belonging to the same typology, *e.g.,* two businesswomen, may have different opinions regarding the same objective, not necessarily concerned with business transactions: one of them would never live in downtown, while the other could prefer to have various kind of services available 24 hours a day.

The above considerations may lead to re-consider how to elaborate the scores of the single parameters. Instead of adopting a plain average of the single scores (as it is the case, *e.g.,* of Booking.com), these could be weighed with *priorities*, representing the dominance, or relevance, of each parameter with respect to the others. In this paper, we propose a way to obtain such priorities.

We adopt a well known technique, the *Analytic Hierarchy Process* (AHP) (Saaty 1980), successfully adopted in various settings to make decisions, see, *e.g.,* (Saaty 2008) and Section 2. Here, we move away from the traditional AHP field of application, *i.e.,* decision making, and we apply this process for prioritising the rating of a service.

The original intent of AHP is helping the decision makers to rank the alternatives of a decision. AHP relies on the fact that the importance of each factor influencing the decision is evaluated with respect to the importance of all the other factors, through pairwise comparisons. As a simple example, suppose that the objective of a decision is: *booking a restaurant for dinner*. Criteria that may reasonably help in fulfilling this objective could be *cost*, *food quality*, *staff kindness*, and *location*. The methodology imposes to first compare cost against food quality, staff kindness, and location. Then, food quality must be compared with staff kindness and location. Finally, staff kindness is compared with location.

Usually, the scale of importance is typically left to the subjectivity of the decision maker, or to the judgment of experts in the field. Here, we ask about one hundred persons which factors they consider more important when choosing a service. By elaborating the set of answers, we derive a general scale of importance among the factors. This serves as the starting point to apply AHP and find the priorities to weigh the scores of the single factors. Other approaches are possible for deriving the input data to AHP, *e.g.,* collecting reviews posted by real users on websites specialised in services advices, and analysing the dataset in order to discover, at least statistically, some dominance among the various factors affecting the overall scores. This concept will be further elaborated throughout the paper. Such a combined use of an experimental approach and the application of the AHP methodology may bring added value in computing more accurate ratings for products and services advertised on the Web.

The paper is structured as follows. Section 2 recalls the main steps of the Analytic Hierarchy Process. In Section 3, we present the survey carried out to empirically derive a scale of importance between pairs of factors. Section 4 shows an application of AHP to hotel reviews. In Section 5, we further elaborate on improvements and further suggestions to find a set of input data to the methodology. Section 6 discusses related work in the area. Finally, Section 7 concludes the paper.

## 2 The Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) (Saaty 1980) has been largely adopted in decision making processes, *i.e.,* the cognitive processes resulting in a selection among several alternatives (Figueira, Greco, and Ehrgott 2005). The selection of an alternative with respect to the others may be supported by the evaluation of a set of criteria. AHP offers a way to evaluate such criteria through pairwise comparisons.

The methodology can be described as follows: after the definition of an objective (*e.g.,* booking a holiday), one needs to structure a decision hierarchy, *i.e.,* a set of criteria, and sub-criteria, that may be helpful in evaluating the alternatives. Booking a holiday may depend, *e.g.,* on the following set of criteria: *location preference*, *cost*, and *reliability of the travel agency*. A possible set of sub-criteria for cost are, *e.g., hotel cost*, *plane cost*, and *cost of public transportation*.

Some criteria are easy to be evaluated, *e.g.,* everyone knows with confidence if she prefers the seaside or the countryside. However, one could know in advance the cost of the air travel and the accommodation, but costs for public transportations on site may be just estimated. Finally, criteria like reliability of a third party are hard to state *a priori*.

With AHP, decision makers can evaluate the relative importance of each criterion with respect to the others. This is achieved by constructing a *pairwise comparisons matrix*. Each element of this matrix represents the comparison between two criteria of the same level [1]. As explained in (Saaty 2008), comparisons are through a scale of numbers indicating how many times more important one criterion is over another criterion with respect to the objective. Table 1 shows a typical example scale, inherited from (Saaty 2008). For *Intensity of Importance* minor than 2, *i.e.,* between 1.1 and 1.9, it is possible to consider real numbers, by augmenting in such a way the granularity of the comparison.

[1] In this paper, we concentrate only on a first level of criteria.

Table 1: Fundamental Scale for AHP

| Intensity of Importance | Definition |
|---|---|
| 1 | Equal |
| 2 | Weak |
| 3 | Moderate |
| 4 | Moderate plus |
| 5 | Strong |
| 6 | Strong plus |
| 7 | Very strong |
| 8 | Very very strong |
| 9 | Extreme |
| 1.1 - 1.9 | Almost equal |

Table 2 shows how the scale is used to compare the relative importance of criteria in booking a holiday. We need to compare a criterion indicating on the left with one at the top. For example, enter 3 in the cost/reliability position means that the cost of the holiday is *moderately* more important than the reliability of the travel agency.

A pairwise comparisons matrix $M$ has positive entries and it is a reciprocal one, *i.e.,* for each element $a_{ij}$, $a_{ij} = \frac{1}{a_{ji}}$, $i, j = 1, \ldots, n$. Moreover, in order to find *plausible* priorities, $M$ should be consistent (or near consistent). A matrix is consistent if $a_{ij} = a_{ik} * a_{kj}$, $i, j, k = 1, \ldots, n$. The satisfaction of this last property implies that if criterion $x$ is more important then criterion $y$, and criterion $y$ is more important then criterion $z$, then $z$ cannot be more important then $x$. In practice, building a perfectly consistent matrix is not possible, since the judgments are left to humans. As Saaty shows in (Saaty 1990), inconsistency of a reciprocal matrix can be captured by the so called *Consistency Index*: CI = $\frac{\lambda_{max} - n}{n-1}$, where $\lambda_{max}$ is the maximum eigenvalue of $M$ and $rank(M) = n$. In a consistent matrix, $\lambda_{max}$ = $n$, (Saaty 1990). Whereas, in an *almost* consistent matrix, $\lambda_{max}$ is greater and very close to $n$. Thus, Saaty proposes to compare CI with the same index obtained as an average over a large number of reciprocal matrices of the same order, and whose entries are randomly picked among the ones constituting the scale of importance. There exist some studies on the computation of this last index, called *Mean Random Consistency Index* (MRCI), see, *e.g.,* (Tummala and Ling 1998). If the *Consistency Ratio* (CR) of CI to that from MRCI is less than 0.1, then the priorities are plausible.

Table 2: Example Comparisons Matrix

| | Location | Cost | Reliability |
|---|---|---|---|
| Location | 1 | $\frac{1}{2}$ | $\frac{1}{4}$ |
| Cost | 2 | 1 | 3 |
| Reliability | 4 | $\frac{1}{3}$ | 1 |

Table 3: Survey Results

| Pairs of Factors | Number of Users |
|---|---|
| Cleanness *vs* Comfort | 70 - 18 |
| Cleanness *vs* Services | 72 - 16 |
| Cleanness *vs* Staff | 80 - 8 |
| Cleanness *vs* Location | 57 - 31 |
| Cleanness *vs* Value for Money | 44 - 44 |
| Comfort *vs* Services | 56 - 42 |
| Comfort *vs* Staff | 68 - 20 |
| Comfort *vs* Location | 24 - 64 |
| Comfort *vs* Value for Money | 16 - 72 |
| Services *vs* Staff | 59 - 29 |
| Services *vs* Location | 18 - 70 |
| Services *vs* Value for Money | 18 - 70 |
| Staff *vs* Location | 14 - 74 |
| Staff *vs* Value for Money | 9 - 79 |
| Location *vs* Value for Money | 52 - 36 |

Once the matrix has been formed, it is possible to calculate the priorities to assign to each criterion. With perfect consistency, the priorities vector of a reciprocal matrix is the eigenvector associated with the maximum eigenvalue (Saaty 1977). For slightly inconsistent matrices, Saaty himself justifies the computation of the maximum eigenvector with the theory of perturbations, saying that slight variations in a consistent matrix implies slight variations of the eigenvalue and the eigenvector. However, other authors, as described in (Ishizaka and Lusti 2006), suggest to apply the power method, a numerical method to calculate the maximal eigenvector, see *e.g.,* (Bini, M.Capovani, and Menchi 1988).

## 3 Survey on Users Preferences

Here, we describe the survey we carried out in order to establish pairwise comparisons between different factors that may affect the choice of a hotel. Following Booking.com, specialised in online hotel reservations, we consider six factors possibly affecting a hotel review: *clean*, *comfort*, *location*, *services*, *staff*, and *value for money*. We set up an online survey and ask a set of friends and colleagues their feelings about the degree of importance of each of the six factors with respect to all the others. We asked about one hundred persons of varying age, sex, profession, nationality, and social class. They were invited to answer as they were going to choose a high-class hotel with the purpose of going on vacation. We obtained eighty-eight compiled questionnaires. The survey was anonymous. Questions were very elementary and they did not contain any quantitative evaluation, in order to make the level of the questionnaire as simple as possible. In practice, for each pair of factors, the user was asked to express a preference related to which of the two factors she considers more important when choosing a hotel. Results are shown in Table 3.

If the same number of preferences were given to two factors, this would mean that the *Intensity of Importance* for those factors would be equal to 1, *i.e.,* the factors would have an equal importance (see Table 1). Thus, we apply simple proportions to measure how many times one factor is more important than the others. As an example, considering the *Location* factor, we obtain the following ratios $\frac{Location}{Staff} = \frac{74}{14} = 5$; $\frac{Location}{Comfort} = \frac{64}{24} = 3$; $\frac{Location}{Services} = \frac{70}{18} = 4$; *etc...* . Results between 1.1 and 1.9 have been considered as they are. Results equal or higher than 2 have been approximated to the closest integer. Results higher than 9 have been approximated to 9. By proceeding in such a way for all the pairwise comparisons, we obtain the matrix in Table 4. The matrix has a consistency index CI equal to 0.00891. By following (Tummala and Ling 1998), it is possible to estimate the value of the consistency ratio CR as about 0.007, This proves the consistency of the judgments.

Table 4: Pairwise Comparisons Matrix for Hotel Review
CI = 0.00891

| | Clean | Comfort | Services | Staff | Value For Money | Location |
|---|---|---|---|---|---|---|
| Clean | 1 | 4 | 4 | 9 | 1 | 1.8 |
| Comfort | $\frac{1}{4}$ | 1 | 1.3 | 3 | $\frac{1}{4}$ | $\frac{1}{3}$ |
| Services | $\frac{1}{4}$ | $\frac{1}{1.3}$ | 1 | 2 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| Staff | $\frac{1}{9}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | 1 | $\frac{1}{9}$ | $\frac{1}{5}$ |
| ValueForMoney | 1 | 4 | 4 | 9 | 1 | 1.4 |
| Location | $\frac{1}{1.8}$ | 3 | 4 | 5 | $\frac{1}{1.4}$ | 1 |

## 4 Priorities for Hotel Reviews

In this section, we apply AHP to derive the priorities to weigh the evaluation of each factor affecting a hotel review. Table 5 shows the results obtained by computing the eigenvector associated to the maximal eigenvalue. The ones obtained by following the power method (ten iterations) are in practice the same.

Table 5: Priorities

| Clean | Comfort | Services | Staff | V. for M. | Location |
|---|---|---|---|---|---|
| 0.31 | 0.08 | 0.06 | 0.03 | 0.29 | 0.21 |

The *prioritised* review $R^h$ for hotel $h$ is then computed as a weighted mean, where, intuitively: $v_i$ is the score given by the user to factor $i$ and $w_i$ is the priority for factor $i$.

$$R^h = \sum_{i=1}^{6} w_i * v_i \tag{1}$$

We carry out an analysis of existing user reviews on Booking.com. To this aim, we pick a set of eighty hotels in New

York City, ranged over medium and high classes[2]. Hotel reviews are numerical values ranged over $\{1, \ldots, 10\}$. The overall rating is obtained as the plain average of the single scores given by the users to each factor. Following this strategy, Booking.com gives to all the factors the same priority. Instead, we re-compute the overall rating according to equation 1. Table 6 shows a comparison between some rating of hotels computed as a plain average, as reported by Booking.com, and the weighed one obtained adopting AHP.

## 5 Discussion

In the previous sections, we have shown an application of AHP for prioritizing users reviews. A first step in applying AHP regards a pairwise comparisons among a list of criteria helpful for making a decision (in our use case, among a list of factors affecting the evaluation of a service). The scale of importance among criteria is generally left to the judgments of experts, or other kind of measures. Intuitively, there are several ways to extract a dataset to give as input to that process. One alternative is setting up a survey, asking people which factor, between each pair in a set, is more *important* for them. We ran this survey in the specific example of choosing a high class hotel for going on holiday. Obviously, a real world implementation of this model requires to consider more participants. This could be achieved by online providers/vendors with a huge audience.

A set of more representative prioritised reviews could be probably achieved by quantifying the scale of importance between criteria according to *categories of users*, *e.g.,* in case of travel reservation, businessmen, families with children, solo travellers, *etc...* . Also, temporal windows could be used for review computation, to give, *e.g.,* more relevance to recent ratings rather than to old ones (Saaty 2007).

Regarding the particular use case considered in this paper, an alternative to investigate is the following. Looking at the review-set collected from Booking.com, we notice that users tend to rate some factors more strictly than the others, and that this behaviour happens with a regular occurrence. In particular, for that dataset, a significant indicator could be the difference, or *distance*, between each of the mean scores assigned to the single factors affecting the review, and the review itself. More precisely:

1. For each hotel $j$, with $j = 1, \ldots, n$, and for each factor $f_i$, $i = 1, \ldots, 6$, it is possible to calculate the distance $d^j_{f_i}$ (in absolute value) between the score $v^j_{f_i}$ assigned by users to $f_i$ and the overall rating $R^j$. We obtain $n$ distances for each factor. (see 2).

$$d^j_{f_i} = |v^j_{f_i} - R^j \mid \quad i = 1, \ldots, 6 \quad j = 1, \ldots, n \tag{2}$$

2. For each $f_i$, it is possible to compute the mean of all the distances (see 3). $\bar{D}_{f_i}$ represents, on average, how far the score of factor $i$ is from the overall rating.

$$\bar{D}_{f_i} = \frac{\sum_{j=1}^{n} d^j_{f_i}}{j} \quad i = 1, \ldots, 6 \tag{3}$$

[2]Dataset collected from Booking.com in June, 2011.

If there were two, or more, factors, whose scores equally differ, on average, from the overall rating, this could lead to conclude that those factors equally influence the service rating, with respect to a recurrent judgment of users. Thus, by applying simple proportions between the mean distances, it is possible to measure how many times one factor is more distant than the others, with respect to Saaty's scale of importance, and then build the comparisons matrix.

The main idea behind this approach is that those factors that, on average, have assigned a stricter, or larger, judgment, with respect to the overall review, are the ones deserving a higher priority. It is worth noticing that the sketched approach could be easily implemented and maintained by those Internet providers having a huge review set already present, and constantly updated, in their databases.

Finally, other methodologies have been proposed and successfully applied in the area of Multi-Criteria Decision Analysis MCDA. In particular, AHP implies the independence among the individual criteria, which is not always true. In choosing a restaurant for dinner, the decision maker may want to base her decision on cost and food quality. AHP assumes that these two criteria are independent of one another, while it could be the case that higher food quality is achieved by paying more. Examples of techniques that can define interactions among criteria are the Analytic Network Process ANP (Saaty and Vargas 2006), and the 2-additive Choquet integral, see, *e.g.,* (Ceberio and Modave 2004). In this paper, we apply AHP to favour easy of use and low complexity. However, we aim at investigating the application of other MCDA techniques, and we leave this as future work.

## 6 Related Work

Review systems are in-depth related to the concept of reputation and reputation systems for online service provision. According to Dellarocas (Dellarocas 2010), "*reputation* is a summary of one's relevant past actions within the context of a specific community ", and "a reputation system is an information system that mediates and facilitates the process of assessing one's reputation within the context of a specific community". Thus, reputation is a concept related to individuals (and/or to the individual's work), and, in particular, helpful to other individuals in order to make choices.

Reviews posted by users should be considered truthful if supported by a reputation mechanism assessing the trustworthiness of the reviewers. While not our focus, we acknowledge research work in the area of immunising review systems against unfair (or incomplete) ratings, *e.g.,* (Dellarocas 2000; Whitby, Josang, and Indulska 2004; Zhang and Cohen 2006; Feng et al. 2008; Dellarocas and Wood 2008; Gorner, Zhang, and Cohen 2011).

However, assuming fair ratings, services obtaining higher ratings are likely those that will be chosen at most, and this positive trend should enhance the reputation of the service providers. Under this point of view, the reputation of providers could overlap, to some extent, with the notion of *reputation of the service* they provide. As defined by Jøsang et al. in (Jøsang, Ismail, and Boyd 2007), "reputation is what is generally said or believed about a persons or things character or standing", thus including *things* as entities to which

Table 6: NYC Hotel Rating as reported by Booking.com and following our approach: some results

| Clean | Comfort | Services | Staff | V. for M. | Location | **Booking.com Rating** | **Prioritised Rating** |
|---|---|---|---|---|---|---|---|
| 8.5 | 7.9 | 8 | 8.9 | 6.7 | 9 | 8.17 | 7.85 |
| 8.9 | 8.6 | 8 | 8.7 | 7.2 | 9 | 8.40 | 8.17 |
| 9.4 | 8.7 | 8.3 | 8.7 | 7.4 | 8.7 | 8.53 | 8.34 |
| 9.3 | 8.7 | 8.6 | 9 | 8.5 | 7.8 | 8.65 | 8.47 |
| 8.3 | 8.1 | 7.5 | 8.6 | 7.3 | 9 | 8.13 | 7.94 |
| 8.8 | 8.1 | 7.9 | 9.1 | 7.6 | 9.5 | 8.50 | 8.32 |
| 8.5 | 7.8 | 7.8 | 8.0 | 7.1 | 9.0 | 8.03 | 7.92 |
| 9.3 | 9.0 | 9.0 | 9.7 | 8 | 9.0 | 9.0 | 8.64 |
| 9.2 | 8.8 | 8.3 | 7.9 | 7.5 | 9.0 | 8.45 | 8.36 |
| 8.7 | 7.5 | 7.7 | 8.8 | 7.3 | 8.7 | 8.12 | 7.92 |
| 8.8 | 7.5 | 7.8 | 9.1 | 6.6 | 8.8 | 8.27 | 7.91 |
| 9.0 | 7.5 | 8.1 | 6.9 | 6.9 | 5 | 7.23 | 7.93 |

a reputation can be associated with. However, since most of the literature in the area of reputation systems for online provision of services uses to attach reputation values to human beings (or devices operating on their behalf), we have preferred in this paper to refer to the *rating* of a service, rather then its *reputation*.

The rating of a service (or a product) is kept up-to-date according to algorithms generally built on the principle that the new rating is a function of the old ratings and the most recent review(s) (Jøsang, Ismail, and Boyd 2007). In simple models, such the one adopted by Ebay prior to May 2008, past and new ratings about the outcome of online transactions between a buyer and a seller contribute in an equal manner to the calculation of the reputation of the seller. More recently, Ebay started considering only the percentage of positive ratings of the last twelve months. The same temporal window is also used in the Amazon marketplace. Other models combine in a weighted mean the old rating and the newest reviews, that are in such a way *prioritised*. In the literature, there exist proposals to determine the values of the priorities. based on, *e.g.,* the trustworthiness of the reviewer (Buchegger and Le Boudec 2003; Cornelli et al. 2002; Yu and Singh 2003), the evaluation of the users satisfaction for a set of parameters characterising the object (Griffiths 2005), the review freshness, or the distance between the single review and the overall score (as suggested in (Jøsang, Ismail, and Boyd 2007)). Work in (Zacharia, Moukas, and Maes 1999) prioritises reviews by their origin, such that reviews posted by users with an established attendance in the system are weighted more than the reviews given by beginners. Similarly, some models suggest to rate reviewers according to a set of attributes, such as certificates attesting the users expertise on the object of the review. In such a way, different priorities are put on reviews posted by different categories of users, in order for instance to weigh more the reviews given by professionals and give less weight to reviews given by regular users (van Deursen, Koster, and Petkovic 2008; Chen, Zeng, and Wenyin 2006). Work in (Aperjis and Johari 2010a) considers the reputation of sellers in electronic marketplaces as an aggregation of past and recent transactions. In particular, the authors show that, when unweighted ratings are averaged over the entire lifetime of the seller, she may be incentivized to falsely advertise her products. On the other hand, if recent transactions influence the seller's reputation more than the past ones, then the seller is more incentivized to truthful advertisement. To this aim, they propose an optimisation of the *Window Aggregation Mechanism*, in which the seller's score is the average of the last $n$ most recent ratings. Also, work in (Aperjis and Johari 2010b) considers the *Weighted Aggregation Mechanism*, where the seller's score is a weighted average of past ratings, optimal with respect to incentivize the seller to be truthful. Fan et al. (Fan, Tan, and Whinston 2005) propose to achieve a similar goal by adopting exponential smoothing to aggregate ratings, and evaluate it through simulations.

Here, we tackle the issue of prioritising the feedback regarding a set of factors characterising a service. We adopt AHP, a worldwide recognised methodology that was born with the intent of engineering the process of decision making. AHP has been used in several settings to make decisions. A comprehensive list of field of applications, ranging over, *e.g.,* public administration, disaster recovery, allocation of huge sums of money, and military and political systems, can be found in (Saaty 2008), Section 9. Within computer security, two recent proposals suggest the adoption of AHP. Work in (Colantonio 2011) is about prioritising *role engineering*, a discipline strictly related to Role-based Access Control models (Sandhu et al. 1996), and aimed to choose the best way to design a proper set of roles within structured organisations. Instead, the authors of (Rajbhandari and Snekkenes 2011) face the issue of measuring the effectiveness of security controls and metrics, and use AHP to select the most appropriate set of strategies leading to the effectiveness of a control. In our work, we mix the adoption of AHP with a survey whose results serve to fix the values of the pairwise comparisons, that are usually left to the judgments of experts, or to the decision maker.

## 7 Conclusion

Review systems are popular mechanisms exploited by online providers to advertise their services. Users may experience a service and ratify the degree of satisfaction encountered under the form of a textual review, generally summarised in a textual judgement or a numerical value. Often, a review depends on multiple factors, and, being posted by humans, subjectivity and personal attitudes may let the users to give different importance to those factors. In this paper, we investigated the application of the Analytic Hierarchy Process to derive the priorities to attach to the scores assigned to different parameters affecting a global rating. Usually, AHP relies on the judgements of experts to operate a first screening on a set of different criteria useful to reach the ultimate objective. Here, we considered the example of hotel reviews. We have asked about one hundred people to decide, for each pair in a set of factors, which was more important for them when choosing a hotel. On the basis of the results, we have obtained a dataset to give as input to AHP. The derived priorities are consistent, with respect to the input data. Nevertheless, other approaches could be explored in order to scale the factors, as discussed in Section 5. We suggest the use of AHP, combined with an opportune approach for finding input data as a practical application that could be actually incorporated into real multi-factor reviewing systems.

## Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant no 257930 (Aniketos) and under grant no 256980 (NESSoS).

## References

Aperjis, C., and Johari, R. 2010a. Designing aggregation mechanisms for reputation systems in online marketplaces. *SIGecom Exch.* 9:3:1–3:4.

Aperjis, C., and Johari, R. 2010b. Optimal windows for aggregating ratings in electronic marketplaces. *Management Science* 56(5):864–880.

Bini, D.; M.Capovani; and Menchi, O. 1988. *Metodi Numerici per l'Algebra Lineare*. Zanichelli.

Buchegger, S.; and Le Boudec, J. 2003. A robust reputation system for mobile ad-hoc networks. Technical report, IC/2003/50 EPFL-IC-LCA.

Ceberio, M.; and Modave, F. 2004. An interval-valued, 2-additive choquet integral for multicriteria decision making. In *Information Processing and Management of Uncertainty in Knowledge-based Systems*.

Chen, W.; Zeng, Q.; and Wenyin, L. 2006. A User Reputation Model for a User-Interactive Question Answering System. In *Intl. Conference on Semantics, Knowledge and Grid*, 40. IEEE.

Colantonio, A. 2011. Prioritizing role engineering objectives using the analytic hierarchy process. In *Italian Chapter of AIS*, 1–8.

Cornelli, F.; Damiani, E.; di Vimercati, S. D. C.; Paraboschi, S.; and Samarati, P. 2002. Choosing reputable servents in a p2p network. In *World Wide Web*, 376–386. ACM.

Dellarocas, C.; and Wood, C. A. 2008. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science* 54(3):460–476.

Dellarocas, C. 2000. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *ACM Conf. on Electronic Commerce*, 150–157.

Dellarocas, C. 2010. Designing reputation systems for the social web. Technical Report Research Paper No. 2010-18, Boston U. School of Management.

Fan, M.; Tan, Y.; and Whinston, A. B. 2005. Evaluation and design of online cooperative feedback mechanisms for reputation management. *IEEE Trans. Knowl. Data Eng.* 17(2):244–254.

Feng, Q.; Yang, Y.; Y.L., S.; and Dai, Y. 2008. Modeling attack behaviors in rating systems. In *Distributed Computing Systems Workshops*, 241–248.

Figueira, J.; Greco, S.; and Ehrgott, M. 2005. *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer.

Gorner, J.; Zhang, J.; and Cohen, R. 2011. Improving the use of advisor networks for multi-agent trust modelling. In *Privacy, Security and Trust*, 71 –78.

Griffiths, N. 2005. Task delegation using experience-based multi-dimensional trust. In *AAMAS*, 489–496. ACM.

Ishizaka, A.; and Lusti, M. 2006. How to derive priorities in AHP: a comparative study. *Central European Journal of Operations Research* 14(4):387–400.

Jøsang, A.; Ismail, R.; and Boyd, C. 2007. A survey of trust and reputation systems for online service provision. *Decis. Support Syst.* 43:618–644.

Rajbhandari, L.; and Snekkenes, E. A. 2011. An approach to measure effectiveness of control for risk analysis with game theory. In *STAST*, 24 – 29. IEEE.

Saaty, T.; and Vargas, L. 2006. *Decision making with the analytic network process*. Number 95 in Operations research & management science. Springer.

Saaty, T. L. 1977. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 15(3):234–281.

Saaty, T. L. 1980. *The Analytic Hierarchy Process*. McGraw - Hill.

Saaty, T. L. 1990. How to make a decision: The analytic hierarchy process. *European Journal of Operational Research* 48(1):9–26.

Saaty, T. L. 2007. Time dependent decision-making; dynamic priorities in the AHP/ANP. *Mathematical and Computer Modelling* 46(7-8):860–891.

Saaty, T. L. 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences* 1(1).

Sandhu, R. S.; Coyne, E. J.; Feinstein, H. L.; and Youman, C. E. 1996. Role-based access control models. *IEEE Comp.* 29:38–47.

Tummala, V. M. R.; and Ling, H. 1998. A note on the computation of the mean random consistency index of the analytic hierarchy process. *Theory and Decision* 44(3):221–230.

van Deursen, T.; Koster, P.; and Petkovic, M. 2008. Hedaquin: A reputation-based health data quality indicator. *Electr. Notes Theor. Comput. Sci.* 197(2):159–167.

Whitby, A.; Jøsang, A.; and Indulska, J. 2004. Filtering out unfair ratings in bayesian reputation systems. In *Workshop on Trust in Agent Societies*.

Yu, B.; and Singh, M. P. 2003. Detecting deception in reputation management. In *AAMAS*, 73–80. ACM.

Zacharia, G.; Moukas, A.; and Maes, P. 1999. Collaborative reputation mechanisms in electronic marketplaces. In *HICSS*, 8026–. IEEE.

Zhang, J.; and Cohen, R. 2006. Trusting advice from other buyers in e-marketplaces: the problem of unfair ratings. In *ICEC*, 225–234.