A Formal Cost-Effectiveness Analysis Model for Product Evaluation in E-Commerce

Ran Wei and Haiping Xu
Computer and Information Science Department
University of Massachusetts Dartmouth, North Dartmouth, MA 02747, USA
{rwei, hxu}@umassd.edu

***Abstract*—Due to the inherent nature of e-commerce, customers usually have to take certain level of risks while shopping online. To deal with such risks and their associated uncertainty, most of the e-commerce websites provide product review ranking services to help customers to make purchase decisions. However, such services are typically not reliable because the ranking results are usually based on the averages of review scores given by different reviewers without considering their reliability. In this paper, we propose a formal cost-effectiveness analysis model for product evaluation in e-commerce, which takes the reliability of each review into consideration. We define four pieces of evidence, namely positive reviews, the number of positive reviews, negative reviews, and the number of negative reviews, and combine them using the Dempster-Shafer (D-S) theory. Based on the belief values about the product, we can calculate its effectiveness, and further derive its cost-effectiveness value by considering its minimal price. By ranking various products sold by different vendors based on their cost-effectiveness values, our approach can greatly help customers to make decisions on selecting the most cost-effective products for online purchasing.**

Keywords-E-commerce; product reviews; cost-effectiveness; reasoning under uncertainty; Dempster-Shafer (D-S) theory.

I. INTRODUCTION

As e-commerce techniques are growing rapidly, people today increasingly shop online instead of directly shopping in physical stores. However, because of the inherent nature and complexity of e-commerce environments, the evaluation and selection techniques for purchasing favorable products that fit a customer's needs could be very sophisticated. Customers typically lack the technical knowledge about the product to be purchased. Furthermore, decision making on selecting online products has become more complex due to the variety of brands and tremendous number of similar products available on the electronic market.

To help customers to select the favorable products, many companies, such as Amazon, have attempted to develop suitable and effective product evaluation mechanisms [1]. However, such mechanisms are typically not well employed as expected because the information redundancy and complexity on the review pages usually make customers lose patient or even get confused. Due to this shortcoming, customers often only check the average ratings rather than reading through all the product reviews across multiple web pages.

We noticed that some review ranking services such as the average star ratings were not always reliable. This is because much information related to a product review (e.g., the helpfulness of the review rated by other customers and the qualification of the reviewer) was usually ignored by users, which otherwise could be used as evidential knowledge for evaluating the reliability of the product review. Thus, in our research, we consider such information as hidden knowledge that can be defined as multiple attributes. We first set up the evaluation criteria for each attribute quantified using certain scales. Then we propose a cost-effectiveness analysis model based on the Dempster-Shafer (D-S) theory to rank product alternatives. Note that the D-S theory is a mathematical theory of evidence, which is a powerful tool to support reasoning under uncertainty [2]. Using the Dempster's combination rules, we are allowed to combine various pieces of independent evidence and reach a high-level degree of belief for specific hypotheses. In this paper, we consider the hypotheses whether a product is a favorable one that is worth buying or it is an unfavorable one that is not worth buying. To verify these hypotheses, we divide the available product reviews into two sets, namely the positive reviews and the negative reviews. We calculate the belief values for each set by combining the weighted average review score of a set and its number of reviews as independent pieces of evidence using the D-S theory. Then the two sets' belief values are combined again as independent pieces of evidence to calculate the effectiveness of the product, which can be used to further derive its cost-effectiveness value by taking the product's minimal cost into consideration. By ranking various products sold by different vendors based on their cost-effectiveness values, our approach can greatly help customers to make decisions on selecting the most cost-effective products in online shopping.

II. RELATED WORK

The D-S theory has been used in various areas to support reasoning under uncertainty. Dong *et al.* proposed a practical shill detection mechanism in online auctions using the D-S theory of evidence [3]. The approach takes multiple pieces of evidence from different information layers into account, detects shilling behaviors and assists decision making on shill bidders. Panigrahi *et al.* developed a fraud detection system in mobile communication networks [4]. They utilized the D-S theory to combine multiple pieces of evidence from the rule-based component and compute an overall suspicion score to help users filter suspicious incoming calls. Yang *et al.* presented an evidential reasoning approach that could be used to solve uncertain decision problem with both quantitative and qualitative attributes [5]. They proposed an alternative way to deal with hybrid multiple-attribute decision-making problems with uncertainty. Different from the above approaches, in this paper, we adopt the D-S theory to develop a cost-effectiveness analysis model. Our approach can be used to combine multiple pieces of evidence to evaluate the quality of a product by calculating its effectiveness value based on the review ratings and their associated information.

There are many previous research efforts related to our approach for supporting decision making under uncertainty. Li *et al.* proposed a grey-based decision-making approach to the supplier selection problem [6]. Their approach employed the grey theory, which was one of the methods for mathematical analysis of systems with uncertain information. Denguir-Rekik *et al.* developed a choquet integral-based decision-making method for propagating possibility distributions using generalized weighted mean aggregation operators [7]. They emphasized on possibility distributions rather than precise quantitative evaluations, and used uncertainty indicators to give a user some idea about other people's variability of the evaluations. Herrera *et al.* proposed a fusion approach for managing information evaluated in different linguistic term sets [8]. The aim of their approach is to manage information assessed in different linguistic term sets together in a decision-making problem with multiple information sources. Huynh *et al.* reanalyzed the evidential reasoning (ER) approach, and proposed a general scheme of attribute aggregation in multiple attributes decision-making problem under uncertainty [9]. They showed that new aggregation schemes satisfied the synthesis axioms, for which any rational aggregation process should grant. Most of the above approaches are based on calculating probabilities of certain events, thus they are not readily scalable for decision making with newly acquired evidence. In contrast, we use the D-S theory that is an evidence-based approach to calculate the brief values about a product, which can be easily refined and updated using Dempster's rule of combination when new pieces of evidence about the product are acquired.

In addition, there are some previous research efforts on product analysis. Cho *et al.* developed a product taxonomy for collaborative recommendation in e-commerce [10]. In their approach, they used web usage mining technique to enhance the quality recommendation and system performance. Sarwar *et al.* proposed an analysis recommendation algorithm that could produce useful recommendations to customers [11]. They used traditional approaches such as data mining and dimensionality reduction techniques to handle large-scale purchase and preference data. Although the above approaches are useful in deriving product recommendations, they require analysis of a large amount of data sets. In contrast, our approach emphasizes on analyzing the review information related to a specific product, thus it is much more efficient than data mining based approaches.

III. Dempster-Shafer Theory

The D-S theory is a probabilistic reasoning method, which was developed to solve problems with uncertainty and incompleteness of available information [2, 3]. Let Θ be a finite set of mutually exclusive possible hypotheses, called the *frame of discernment*. For example, when we consider the domain of product evaluation, each product is considered either *favorable* or *unfavorable* for buying, depending on the nature of the evaluated properties and the quantified values of the review evidence. Thus, the frame of discernment for a product can be defined as $\Theta = \{favorable, unfavorable\}$. The power set of Θ that contains all subsets of Θ is defined as $2^\Theta = \{\varnothing, \{favorable\}, \{unfavorable\}, \Theta\}$.

In the D-S theory, a belief mass is assigned to each element of the power set 2^Θ in the interval between 0 and 1. Thus, the basic mass assignment (*BMA*) function *m* is defined as

$$m : 2^\Theta \rightarrow [0, 1],$$

which satisfies the following two requirements:

$$m(\varnothing) = 0 \quad (1)$$

$$\sum_{A \in 2^\Theta} m(A) = 1 \quad (2)$$

In Eq. (1), the mass of the empty set $\varnothing$ represents the measurement for zero state, thus it is defined as 0. Eq. (2) represents that the sum of masses of the elements in the power set equals 1. For example, in our product review example, since $m(\varnothing) = 0$, we have $m(\{favorable\}) + m(\{unfavorable\}) + m(\Theta) = 1$. Note that $m(\Theta)$ represents the mass for conflicting states (both favorable and unfavorable in our example, i.e., a hypothesis says that a product is both favorable and unfavorable), thus it can be interpreted as the measurement for uncertainty. For clarification purpose, in the rest of the paper, we use the notation $m(U)$ to represent $m(\Theta)$, where U represents uncertainty.

Another important function for a set of states (or a hypothesis) *A* is called the *belief* function, which is defined as the sum of the masses of all subsets of *A*.

$$belief\,(A) = \sum_{B \subseteq A} m(B) \quad (3)$$

Intuitively, any portion of the belief committed to the hypothesis *A* must also be committed to any hypothesis that it implies. To obtain the total belief in *A*, one must therefore add to $m(A)$ the quantities $m(B)$ for $\forall\ B \subseteq A$. In our example, we have two hypotheses, namely 1) the product is a favorable one; and 2) the product is an unfavorable one, both of which do not have any proper subset except for $\varnothing$. Thus, according to Eq. (3), we have $belief(\{favorable\}) = m(\{favorable\})$ and $belief(\{unfavorable\}) = m(\{unfavorable\})$.

The Dempster's rule of combination is a critical concept to the original idea of the D-S theory. Given two masses m_1 and m_2 for a hypothesis, the combination rule computes a *joint* mass for the two pieces of evidence under the same hypothesis, which can be calculated as follows,

$$m_{1,2}(\varnothing) = 0 \quad (4)$$

$$m_{1,2}(A) = m_1(A) \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A \neq \varnothing} m_1(B) m_2(C) \quad (5)$$

where $K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C)$.

Eq. (4) says that the combined mass for the empty set $\varnothing$ is zero. In Eq. (5), K represents the measure of the amount of conflicts between the two mass sets. This is determined by summing the masses of any pair of sets *B* and *C*, where $B \subseteq \Theta$, $C \subseteq \Theta$, and the intersection of them is empty. Note that in Eq. (5), $1\text{-}K$ is used as a normalization factor that has the effect of ignoring conflicts between any pairs of states.

IV. A COST-EFFECTIVENESS ANALYSIS MODEL

A. A Conceptual Model

Our proposed formal cost-effectiveness analysis model for product evaluation in e-commerce can help customers verify the quality of a product, which is reflected by its effectiveness value based on the information collected from an e-commerce website, such as Amazon. We notice that Amazon offered a flexible e-commerce platform, which contains a large amount of useful information that can be used to evaluate the quality of a product. More specifically, not only a customer who has purchased a product online is allowed to give a review star rating as well as review comments to the product, but also his review can be further rated by other customers. Due to the page size limitation, for most of the products, the review information has to be distributed across multiple web pages, which are typically ignored by most users except for the first few pages. In order to support automatic analysis of such useful information for decision making on online purchasing, we treat all reviews and their associated properties as evidence that supports a product as either favorable or unfavorable, and derive our cost-effectiveness analysis model. The conceptual model for cost-effectiveness analysis in e-commerce can be formally defined as a 3-tuple (*P*, *Bel*, *MinC*), where

1. $P = \{p_1, p_2, \ldots, p_n\}$ is a set of product alternatives to be evaluated and ranked, which should have very similar functionality and are within the same price range;
2. $Bel: P \rightarrow [0, 1]$ is a belief function employed in our model. Each product alternative has a degree of belief quantifying that the product is worth buying or not;
3. $MinC: P \rightarrow R^+$ is a cost function that maps a product alternative to its minimal price, defined as a positive real number. Note that for a particular product *p*, we can calculate its cost-effectiveness value using $Bel(p)$ and $MinC(p)$, which can then be used to rank the product alternatives in *P*.

Each product $p \in P$ can be further formally defined as a 6-tuple (*REV*, *S*, *PROP*, *Rel*, *EV*, *M*), where

1. $REV = \{r_1, r_2, \ldots, r_n\}$ is a set of product reviews for product *p*, given by different reviewers;
2. $S: REV \rightarrow \{0.2, 0.4, 0.6, 0.8, 1\}$ is the star ranking function for product reviews, where the star rankings of 1 to 5 has been normalized to a value between [0.2, 1];
3. $PROP = \{pr_1, pr_2, \ldots, pr_k\}$ is a set of review properties, which contribute to calculating the reliability of each review;
4. $Rel: REV \rightarrow [0, 1]$ is a reliability function for product reviews, which represents the importance and accuracy of each review;
5. $EV = \{ev_1, ev_2, \ldots, ev_l\}$ is a set of evidence used to justify a product as either favorable or unfavorable;
6. $M = \{m: EV \rightarrow [0, 1]\}$ is a set of mass assignment functions, which quantify and assess each piece of evidence into a mass that supports a product as either favorable or unfavorable.

In our research, we classify all available reviews for a product *p* into two groups: 1) a supportive group with a set of positive reviews that support the product as a favorable one (i.e., worth buying); and 2) a non-supportive group with a set of negative reviews that do not support the product as a favorable one (i.e., not worth buying). For example, when the 5-star rating mechanism is used, we would consider a review with 4 or 5 stars as a positive review; while a review with 1, 2, or 3 stars as a negative one. Furthermore, we consider the number of reviews in each group as separate independent pieces of evidence. Thus, we have four pieces of evidence in total that can be used to justify a product is either favorable or unfavorable. The four pieces of evidence are denoted as {*PR*, *NP*, *NR*, *NN*}, where *PR* is a set of positive reviews, *NP* is the number of positive reviews, *NR* is a set of negative reviews, and *NN* is the number of negative reviews. Note that since each group has two pieces of evidence to support its committed hypotheses, we first combine the evidence in each group separately (i.e., *PR* and *NP* for positive reviews, and *NR* and *NN* for negative reviews, respectively), then the mass values for the two groups of reviews are combined again to calculate the belief values about the product.

Fig. 1 shows the framework for processing the review data of a particular product. Once the review ratings and their associated review properties are extracted, the reliability of each review can be calculated. We classify the review data into two groups of evidence, namely the supportive group and the non-supportive group. The two pieces of evidence in the same group are combined using Dempster's rule of combination, and derive the two sets of mass values for supportive evidence and non-supportive evidence, respectively. We consider the two pieces of combined evidence as conflicting evidence, and use Dempster's rule of combination again to calculate the *belief values* about the product, and further derive its *effectiveness* as defined in Section IV.C of this paper. Note that our approach does *not* involve analyzing the actual review comments, but it is envisioned as our more ambitious future research direction.

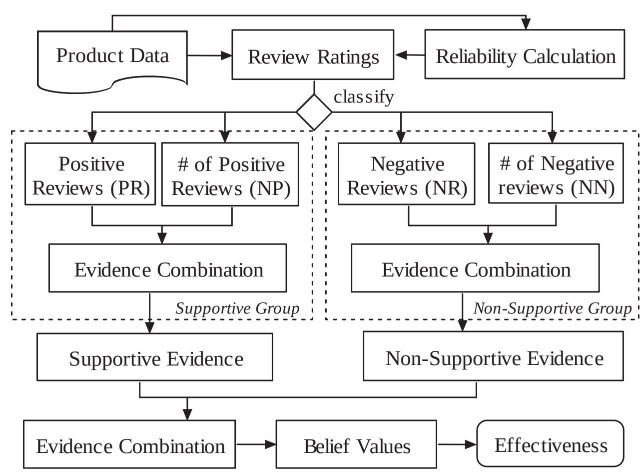

Figure 1. A framework for processing product review data

B. Calculation of Basic Mass Assignments

The first step to calculate the basic mass assignmnets for product reviews is to compute the reliability of each review. The reliability of a review is determined by a number of factors, called *review properties*, which indicate the trustworthiness of a review. We now use Amazon website as an example to demonstrate how to calculate the reliability of a review. Fig. 2 shows a snapshot of a review. As shown in the figure, the Amazon's website allows a review to be voted as a *helpful* review by its customers. Such information is enclosed in a box, which is labeled as *Helpful Rate* (pr_1) in the figure. The more votes as helpful reviews over the total number of votes, the more reliabile the review is. Thus, we consider the number of helpful votes and the number of total votes as major factors to calculate the reliability of each review. Let the number of total votes be *total_votes* and the number of helpful votes be *helpful_votes*, we calculate the *Helpful Rate* (pr_1) of the review as *help_votes* / *total_votes*. Note that if *total_votes* equals 0, we set $pr_1 = 0$.

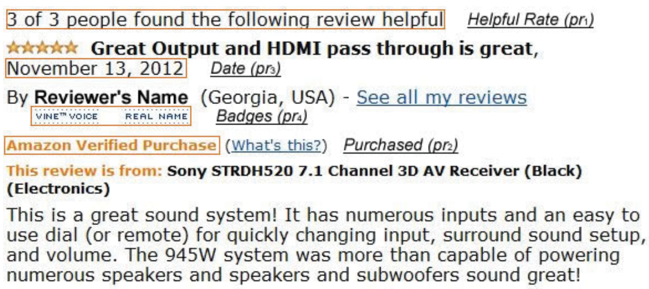

Figure 2. A snapshot of a product review with review properties

We further consider four additional review properties as factors that contribute to computing the reliability of a review. Those factors are *Purchased* (pr_2), *Date* (pr_3), and *Badges* (pr_4) as shown in Fig. 2, and *Top Reviewer Ranking* (pr_5) as shown in Fig. 3. Note that the properties pr_2 , pr_4 , and pr_5 are actually properties of the reviewer who wrote the review. Since the reliability of a review is closely related to the reliability of the person who wrote the review, in this paper, we also call them review properties.

Figure 3. A snapshot of a reviewer's personal profile

We now provide the detailed descriptions of the review properties pr_2 to pr_5 as follows.

Purchased (pr_2) is a label of a reviewer indicating that the e-commerce company has verified the reviewer has purchased the product. A reviewer who has bought and had a real experience with the product can surely write more reliable reviews than those who do not.

Date (pr_3) is the date when the review was posted. For simplicity, we convert the date into the number of months that have passed since the review was posted. The more recent a review was written, the more useful and reliable the review is.

Badges (pr_4) is the number of badges that a reviewer has been awarded. At Amazon website, there are totally nine types of badges. For example, the `REAL NAME` badge indicates that the customer used his real name from his credit card. The more badges a reviewer owns, the better review history the reviewer should have. Note that in Fig. 2, the reviewer has a `REAL NAME` badge, but for privacy purpose, we have removed the reviewer's real name from the figure.

Top Reviewer Ranking (pr_5) of a reviewer reflects the opinions of other customers about the reviewer. A reviewer's *Top Reviewer Ranking* is determined by the overall helpfulness of the reviewer's reviews, factoring in the number of reviews the reviewer has written.

Before calculating the reliability of a review, we first normalize the property values into ones in the range [0, 1]. Table 1 shows the value ranges and the normalized values for the five review properties pr_1 to pr_5.

Table 1. Review properties used to determine review reliability

Property	Description	Value Range	Normalized Value
pr_1	Helpful Rate	[0, 1]	helpful_votes / total_votes 0 (if total_votes = 0)
pr_2	Purchased	{0, 1}	0 → 0 (not purchased) 1 → 1.0 (purchased)
pr_3	Date	[0, +∞)	0~3 months → 1.0 3~6 months → 0.7 6~12months → 0.4 > 1year → 0.1
pr_4	Badges	[0, 9]	no_of_badges / 9
pr_5	Top Reviewer Ranking	[1, +∞)	< 1,000 → 1.0 1,000~10,000 → 0.7 10,000~100,000 → 0.4 >100,000 → 0.1

The reliability *Rel*(r) of a product review r can be calculated as in Eq. (6).

$$Rel(r) = w_1 \times pr_1 + w_2 \times (pr_2 + pr_3 + pr_4 + pr_5) \tag{6}$$

where the weights w_1 and w_2 indicate the importance of the review property pr_1 and the other four review properties pr_2 to pr_5, respectively. Since the review property *Helpful Rate* represents the most important factor to determine the reliability of the review, based on our experience, we set $w_1 = 0.6$ and $w_2 = 0.1$, which leads to a reliability in the range [0, 1].

To calculate the *BMA*s for both of the positive and negative reviews, we compute the weighted average star (*WAS*) for the two groups of reviews as in Eq. (7).

$$WAS = \frac{S(r_1) \times Rel(r_1) + S(r_2) \times Rel(r_2) + ... + S(r_k) \times Rel(r_k)}{k} \tag{7}$$

where $S(r)$ and *Rel* (r) are the normalized star ranking and the calculated review reliability for review r, respectively, and k is the total number of reviews in the corresponding group.

Let F = {*favorable*} and ~F = {*unfavorable*}, we have $U = F \cup \sim F$ = {*favorable, unfavorable*}. Let WAS_{PR} and WAS_{NR} be the *WAS* values for the groups of positive and negative reviews, respectively, we can calculate the *BMAs* for both groups as in Eqs. (8-9). Note that $m_{PR}(U)$ and $m_{NR}(U)$ refer to the mass values of uncertainty for positive reviews and negative reviews, respectively.

$$\begin{cases} m_{PR}(F) = WAS_{PR} \\ m_{PR}(\sim F) = 0 \\ m_{PR}(U) = 1 - WAS_{PR} \end{cases} \tag{8}$$

$$\begin{cases} m_{NR}(F) = 0 \\ m_{NR}(\sim F) = WAS_{NR} \\ m_{NR}(U) = 1 - WAS_{NR} \end{cases} \tag{9}$$

Since the number of reviews can serve as a good indicator of the quality and popularity of a product, we consider the number of reviews in each group of reviews as independent evidence. To assess how the number of reviews has an impact on either supporting that the product is a favorable one or an unfavorable one, we first identify the maximum numbers of reviews in both groups (denoted as N_{NP_max} and N_{NN_max}) among the set of product alternatives P. Then we compare the number of reviews in a given group with the corresponding maximum number of reviews in order to assess its impact on the belief value of the product. We realize that some popular product may have an extremely large number of reviews comparing to others. In this case, the result will be dominated by such maximum number. To avoid this situation, we use logarithm function to narrow down the gaps between the numbers of reviews for different product alternatives. We use the following simple example to illustrate the basic idea. Suppose N_{NP_max} = 2,000 among a set of product alternatives Ω, and for a certain product $\alpha \in \Omega$, n_{NP} = 100. When we compare n_{NP} with N_{NP_max}, the impact of n_{NP} becomes very small, although 100 positive reviews represent a considerable amount of reviews. Now if we try to compare $\log_{10} n_{NP}$ with $\log_{10} N_{NP\_max}$, the gap between them can be significantly narrowed down, and the number of positive reviews, $n_{NP} = 100$ in this case, can be properly taken into account as a piece of evidence to support product α as a favorable one.

The *BMA*s for the number of reviews in the supportive group can be calculated as in Eqs. (10-12), where $N_{NP\_max} > 0$.

$$m_{NP}(F) = \begin{cases} \dfrac{2\times\log_{10}(n_{NP}+1)}{\log_{10}(N_{NP\_\max}+1)} - 1 & if\ \log_{10}(n_{NP}+1) \geq (\log_{10}(N_{NP\_\max}+1))/2 \\ 0 & otherwise \end{cases} \quad (10)$$

$$m_{NP}(\sim F) = \begin{cases} 1 - \dfrac{2\times\log_{10}(n_{NP}+1)}{\log_{10}(N_{NP\_\max}+1)} & if\ \log_{10}(n_{NP}+1) < (\log_{10}(N_{NP\_\max}+1))/2 \\ 0 & otherwise \end{cases} \quad (11)$$

$$m_{NP}(U) = 1 - m_{NP}(F) - m_{NP}(\sim F) \quad (12)$$

Note that in order to deal with the special case when n_{NP} = 1, in the above equations, we replace n_{NP} and N_{NP_max} with $(n_{NP}+1)$ and $(N_{NP\_max}+1)$, respectively. When $\log_{10}(n_{NP}+1) \geq (\log_{10}(N_{NP\_max}+1))/2$, we consider it as a piece of evidence that supports the product is a favorable one. As special cases, when $n_{NP} = N_{NP\_max}$, the mass equals 1, which means the evidence fully support that the product is a favorable one. When $\log_{10}(n_{NP}+1) = (\log_{10}(N_{NP\_max}+1))/2$, the mass equals 0, which means we do not consider the insufficient positive reviews as a piece of evidence to support the product as a favorable one. On the other hand, when $\log_{10}(n_{NP}+1) < (\log_{10}(N_{NP\_max}+1))/2$, we consider it as a piece of evidence that supports the product is an unfavorable one rather than a favorable one due to a lack of positive reviews. As a special case, when $n_{NP} = 0$, the mass equals 1. This means that when there is no positive reviews, the product must be a low-quality one, which should not be suggested to customers for purchasing.

Similarly, the *BMA*s for the number of reviews in the non-supportive group can be calculated as in Eqs. (13-15), where $N_{NN\_max} > 0$.

$$m_{NN}(F) = 0 \quad (13)$$

$$m_{NN}(\sim F) = \frac{\log_{10}(n_{NN}+1)}{\log_{10}(N_{NN\_\max}+1)} \quad (14)$$

$$m_{NN}(U) = 1 - m_{NN}(\sim F) \quad (15)$$

Note that different from the positive reviews, the negative reviews are always considered as a piece of evidence that supports the product is an unfavorable one. As two special cases, when $n_{NN} = N_{NN\_max}$, the mass equals 1, which means the evidence fully support the product is an unfavorable one; on the other hand, when $n_{NN} = 0$, the mass value equals 0, which means that there is no evidence (in terms of negative reviews) to show that the product is an unfavorable one.

C. *Combination of Evidence*

Once the basic mass assignments for each piece of evidence are calculated, they can be combined in a systematic manner to provide a more complete assessment on product quality by reducing the uncertainty involved in individual evidence. The evidence fusion procedure is carried out using the Dempster's rule of combination. As shown in Fig. 1, we first combine the evidence of positive reviews (*PR*) and the number of positive reviews (*NP*) into masses m_{SG} for the supportive group, and the evidence of negative reviews (*NR*) and the number of negative reviews (*NN*) into masses m_{NSG} for the non-supportive group. The corresponding rules of combining evidence for F and $\sim F$ are listed as in Eqs. (16-18) and Eqs. (19-21) for the supportive group and the non-supportive group, respectively.

$$m_{SG}(F) = m_{PR}(F) \oplus m_{NP}(F) \quad (16)$$

$$m_{SG}(\sim F) = m_{PR}(\sim F) \oplus m_{NP}(\sim F) \quad (17)$$

$$m_{SG}(U) = m_{PR}(U) \oplus m_{NP}(U) \quad (18)$$

$$m_{NSG}(F) = m_{NR}(F) \oplus m_{NN}(F) \quad (19)$$

$$m_{NSG}(\sim F) = m_{NR}(\sim F) \oplus m_{NN}(\sim F) \quad (20)$$

$$m_{NSG}(U) = m_{NR}(U) \oplus m_{NN}(U) \quad (21)$$

When the masses for both of the supportive group and non-supportive group have been calculated, we can use Dempster's rule of combination again to combine them into masses $m_{PRODUCT}$ for the product as in Eqs. (22-24).

$$m_{PRODUCT}(F) = m_{SG}(F) \oplus m_{NSG}(F) \quad (22)$$

$$m_{PRODUCT}(\sim F) = m_{SG}(\sim F) \oplus m_{NSG}(\sim F) \quad (23)$$

$$m_{PRODUCT}(U) = m_{SG}(U) \oplus m_{NSG}(U) \quad (24)$$

According to Eq. (3), the belief values for the product hypotheses can be calculated as in the following Eqs. (25-27).

$$belief\,(F) = m(F) \quad (25)$$

$$belief\,(\sim F) = m(\sim F) \quad (26)$$

$$belief\,(U) = m(U) \quad (27)$$

We now use an example to show how the masses for combined evidence can be calculated. Suppose we want to calculate the mass values for the supportive group. According to Eq. (5), we can calculate $m_{SG}(F)$, $m_{SG}(\sim F)$ and $m_{SG}(U)$ as in Eqs. (28-30).

$$m_{SG}(F) = m_{PR}(F) \oplus m_{NP}(F) = \frac{m_{PR}(F)\times m_{NP}(F) + m_{PR}(F)\times m_{NP}(U) + m_{PR}(U)\times m_{NP}(F)}{1-K} \quad (28)$$

$$m_{SG}(\sim F) = m_{PR}(\sim F) \oplus m_{NP}(\sim F) = \frac{m_{PR}(\sim F)\times m_{NP}(\sim F) + m_{PR}(\sim F)\times m_{NP}(U) + m_{PR}(U)\times m_{NP}(\sim F)}{1-K} \quad (29)$$

$$m_{SG}(U) = m_{PR}(U) \oplus m_{NP}(U) = \frac{m_{PR}(U)\times m_{NP}(U)}{1-K} \quad (30)$$

where $K = m_{PR}(F)\times m_{NP}(\sim F) + m_{PR}(\sim F)\times m_{NP}(F)$

Note that since $U \cap F = F$ and $U \cap \sim F = \sim F$, we have $U \cap F \neq \emptyset$ and $U \cap \sim F \neq \emptyset$ as in Eqs. (28-29).

The other two sets of mass values $\{m_{NSG}(F), m_{NSG}(\sim F), m_{NSG}(U)\}$ and $\{m_{PRODUCT}(F), m_{PRODUCT}(\sim F), m_{PRODUCT}(U)\}$ can be calculated in the same way.

According to Eqs. (25-27), the belief value that indicates a product α is a favorable one equals $m_\alpha(F)$, and the value of $m_\alpha(U)$ quantifies the uncertainty that the product α is both favorable and unfavorable. By taking the uncertainty into consideration, we calculate the effectiveness of product α by summing the belief value for a favorable product and 50% of the uncertainty value as in Eq. (31).

$$\begin{aligned} Effectiveness(\alpha) &= Bel(\alpha) = belief(F) + 0.5 \times belief(U) \\ &= m_\alpha(F) + 0.5 \times m_\alpha(U) \end{aligned} \quad (31)$$

By further taking the price factor into consideration, we can calculate the cost-effectiveness value (i.e., *E/C Ratio*) of product α as in Eq. (32).

$$E/C\ Ratio\ (\alpha) = Effectiveness(\alpha) / Cost(\alpha) \quad (32)$$

where $Cost(\alpha)$ is the normalized cost of product α. Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of product alternatives to be evaluated and ranked, which have similar functionality and are within the same price range. For $\forall\ \alpha \in P$, $Cost(\alpha)$ can be calculated as in Eq. (33).

$$Cost(\alpha) = MinC(\alpha) / Max(MinC(p_1), MinC(p_2), ..., MinC(p_n)) \quad (33)$$

where $MinC(\alpha)$ is the cost function (defined in Section IV.A) that maps product α to its lowest price offered by one of the online sellers. With the *E/C Ratio* for each product in set *P*, we can rank the product alternatives, and provide users useful insights about the products in online shopping.

V. Case Study

In this section, we demonstrate how our D-S theory based analysis model can be used to analyze data sets collected from Amazon. We use a case study to show how our analysis model can provide more reliable and accurate results than the typical product ranking based on average star ratings (*ASR*).

A. Data Collection

The data used in our case study was collected from recent product records at the Amazon website, but note that the product data such as star ratings, minimal price, and the number of reviews may have changed by the time of this publication. The product review information as well as the reviewer's profile information used in this case study (as demonstrated in Fig. 2 and Fig. 3) can be accessed directly from the product web pages. To ease our data collection task, we developed a Java program that can automatically collect the needed data items from the product web pages, and used them as inputs to our cost-effectiveness analysis model.

Table 2 gives a few examples of our collected raw data, where each row contains the star rating as well as the five review properties. The normalized values for the star rating and the review properties are shown inside the parentheses along with the raw data, which can be calculated according to the normalization rules described in Section IV.B as well as the conversion rules defined in Table 1.

Table 2. Examples of collected raw data and the normalized values

Star Rating	Helpful Votes / Total Votes	Purchased	Date	Badges	Top Ranking
5(1)	118/124(0.952)	1(1)	17(0.1)	1(0.11)	63,027(0.4)
5(1)	83/86(0.965)	1(1)	5(0.7)	2(0.22)	556(1.0)
2(0.4)	71/79(0.899)	0(0)	16(0.1)	0(0.00)	81,258(0.4)
4(0.8)	26/27(0.963)	1(1)	11(0.4)	3(0.33)	292,053(0.1)
3(0.6)	98/116(0.845)	1(1)	16(0.1)	0(0.00)	458,571(0.1)

Note that besides the data items listed in Table 2, we also need to collect additional evidence, such as the number of reviews in each category (positive reviews or negative reviews) and the prices of the product offered by different online sellers, which are all required in our analysis approach.

B. Case Study: Audio/Video Receiver

In this case study, we collected 10 A/V receiver products in the price range \$200~\$300, which are different in brands, series, star ratings and the number of reviews, but having their average star ratings at least 4. Table 3 lists the 10 products along with some raw data and the analysis results. Among the raw data, "*ASR*" refers to the average star rating of a product that is posted at its corresponding product page; "*# of Reviews*" is the total number of reviews including both positive (4 and 5 stars) and negative (1, 2, and 3 stars) reviews; and "*Price*" refers to the minimal price of a product that is offered by one of the online sellers. As shown in Table 3, the 10 A/V receiver products are sorted according to *ASR*. Based on *ASR*, the first three product alternatives (No.1-3) look like the best choices for purchasing. By further looking into the number of reviews and prices, a customer may select one out of the three options accordingly (e.g., if the customer does not care about the price too much, he may choose product No. 2 for purchasing).

Now with our analysis model, we can calculate the values of *Effectiveness* and the *E/C Ratio* for all 10 product alternatives, which are listed at the last two columns of Table 3. Since the effectiveness value quantifies the quality level of the products, a customer who cares only about quality may select products with the highest effectiveness values. The top three choices are No. 8 with *Effectiveness* 0.845, No. 3 with *Effectiveness* 0.789, and No. 6 with *Effectiveness* 0.768. On the other hand, if the customer cares about both the quality and cost, he may select products with highest *E/C Ratio* values. In this case, the top three choices are No. 8 with *E/C Ratio* 1.056, No. 10 with *E/C Ratio* 1.032, and No. 3 with *E/C Ratio* 0.913. Note that the ranking results calculated using our analysis model are different from the ranking results based on *ASR*, but our ranking results are more accurate and reliable because our model considers more evidential information before the ranking results are calculated.

Table 3. Product information of ten A/V receivers and the analysis results

No.	ASR	# of Reviews	Price	Product & Brand	Effective-ness	E/C Ratio
1	**5.0**	1	199.99	Yamaha HTR-3064	0.609	0.761
2	**4.5**	101	240.81	Harman Kardon HK 3390	0.767	0.796
3	**4.3**	73	215.99	Onkyo HT-S3500	**0.789**	**0.913**
4	4.3	62	228	Onkyo CS-445	0.642	0.704
5	4.2	18	199.99	Onkyo TX-SR313	0.692	0.865
6	4.2	57	269	Onkyo TX-8050	**0.768**	0.714
7	4.2	66	247.99	Yamaha RX-V471BL	0.764	0.770
8	4.1	153	200	Sony STRDH 520	**0.845**	**1.056**
9	4.1	48	249.95	Yamaha RX-V373	0.758	0.758
10	4.1	81	179.95	Yamaha RX-v371BL	0.743	**1.032**

To verify our analysis results, we look into the raw data collected for our case study. For product No. 1, although it has the highest *ASR* (5.0), but since it has only one positive review, its effectiveness value becomes not high enough comparing to the other product alternatives. For product No. 2, although it has a very good *ASR* as well as a decent number of product reviews, when looking into the review properties of those reviews, we found that many of the reviews are not reliable (e.g., having a very low *Helpful Rate* or even no *Helpful* votes). On the other hand, product No. 8 has relative lower *ASR*; however, it has the highest number of reviews, and most of its reviews are ones with high *Helpful Rates*. Consequently, product No. 8 has the highest *effectiveness* value among the 10 product alternatives.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a formal cost-effectiveness analysis model for product evaluation in e-commerce, which was developed using the D-S theory. In our approach, we consider the product reviews as well as their review properties as pieces of evidence to justify whether a product is a favorable one or not. Product data from e-commerce websites such as Amazon is quantified and evaluated using our formal approach. By applying Dempster's rule of combination, we can combine difference pieces of evidence to derive more reliable belief values about the hypotheses on the quality of the product. Due to the nature of the D-S theory, our analysis model can handle uncertain information and reduce the degree of uncertainty appropriately. Thus, our approach produces more reliable and accurate results than conventional ranking mechanisms such as the one based on average star ratings. By ranking the product alternatives properly, our approach can be very effective in assisting customers to evaluate various products, and make purchase decisions on the most cost-effective ones.

In future research, we plan to develop a trustworthy e-commerce platform based on our formal cost-effectiveness analysis model. In the trustworthy e-commerce model, product reviews can be classified into more meaningful groups using data mining approaches as we did in our previous work [12]. The groups of product reviews can then be used as independent evidence for evidence combination using the D-S theory. With more evidence, our approach can significantly reduce the level of uncertainty, and lead to more accurate and reliable product ranking results. In addition, we will consider deploying our trustworthy e-commerce platform into cloud so that it would work in a flexible way and can be more conveniently accessed through the Internet. Finally, we plan to implement our approach on mobile e-commerce platforms such as tablets and smart phones, and provide customers with more friendly and flexible interfaces for mobile commerce.

REFERENCES

[1] Amazon, *Managing Reviews*, Amazon.com Site Features, 2013. Retrieved on January 1, 2013 from http://www.amazon.com/gp/help/customer/display.html?nodeId=16465311

[2] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[3] F. Dong, S. M. Shatz, and H. Xu, "Reasoning Under Uncertainty for Shill Detection in Online Auctions Using Dempster-Shafer Theory," *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, Vol. 20, No. 7, Nov. 2010, pp. 943-973.

[4] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majunmdar, "Use of Dempster-Shafer Theory and Bayesian Inferencing for Fraud Detection in Mobile Communication Networks," *Lecture Notes in Computer Science*, Spring Berlin Heidelberg, Vol. 4586, 2007, pp. 446-460.

[5] J.-B. Yang and M. G. Singh, "An Evidential Reasoning Approach for Multiple-Attribute Decision Making with Uncertainty," *IEEE Transactions on Systems, Man, and Cybernetics (IEEE TSMC)*, Vol. 24, No. 1, Jan. 1994, pp. 1-17.

[6] G.-D. Li, D. Yamaguchi, and M. Nagai, "A Grey-Based Decision-Making Approach to the Supplier Selection Problem," *Mathematical and Computer Modelling*, Vol. 46, No. 3-4, 2007, pp573–581.

[7] A. Denguir-Rekik, G. Mauris, and J. Montmain, "Propagation of Uncertainty by the Possibility Theory in Choquet Integral-Based Decision Making: Application to an E-Commerce Website Choice Support," *IEEE Transactions on Instrumentation and Measurement*, Vol. 55, No. 3, June 2006, pp. 721-728.

[8] F. Herrera, E. Herrera-Viedma, and L. Martínez, "A Fusion Approach for Managing Multi-Granularity Linguistic Terms Sets in Decision Making," *Fuzzy Sets and Systems*, Vol. 114, No. 1, 2000, pp. 43-58.

[9] V. N. Huynh, Y. Nakamori, T. B. Ho, and T. Murai, "Multiple Attribute Decision Making Under Uncertainty: the Evidential Reasoning Approach Revisited," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, Vol. 36, No. 4, 2006, pp. 804-822.

[10] Y. H. Cho and J. K. Kim, "Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-Commerce," *Expert Systems with Applications*, Vol. 26, No. 3, 2004, pp. 234-246.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommendation Algorithms for E-Commerce," *Proceedings of the 2nd ACM Conference on Electronic Commerce*, 2000, pp.158-167.

[12] B. J. Ford, H. Xu, and I. Valova, "A Real-Time Self-Adaptive Classifier for Identifying Suspicious Bidders in Online Auctions," *The Computer Journal (COMPJ)*, Vol. 56, No. 5, 2013, pp. 646-663.