Disaggregating User Evaluations Using the Shapley Value*

Matthias Feldotto
Department of Computer Science
Paderborn University
feldi@mail.upb.de

Claus-Jochen Haake
Department of Economics
Paderborn University
claus-jochen.haake@wiwi.upb.de

Alexander Skopalik
Department of Computer Science
Paderborn University
skopalik@mail.upb.de

Nadja Stroh-Maraun
Department of Economics
Paderborn University
nadja.stroh-maraun@wiwi.upb.de

ABSTRACT

We consider a market where final products or services are compositions of a number of basic services. Users are asked to evaluate the quality of the composed product after purchase. The quality of the basic service influences the performance of the composed services but cannot be observed directly. The question we pose is whether it is possible to use user evaluations on composed services to assess the quality of basic services. We discuss how to combine aggregation of evaluations across users and disaggregation of information on composed services to derive valuations for the single components. As a solution we propose to use the (weighted) average as aggregation device in connection with the Shapley value as disaggregation method, since this combination fulfills natural requirements in our context. In addition, we address some occurring computational issues: We give an approximate solution concept using only a limited number of evaluations which guarantees nearly optimal results with reduced running time. Lastly, we show that a slightly modified Shapley value and the weighted average are still applicable if the evaluation profiles are incomplete.

CCS CONCEPTS

• Information systems → Social recommendation; **• Theory of computation → Solution concepts in game theory**; **• Applied computing → Economics**;

KEYWORDS

Reputation, Disaggregation, Shapley Value, Sampling

ACM Reference Format:
Matthias Feldotto, Claus-Jochen Haake, Alexander Skopalik, and Nadja Stroh-Maraun. 2018. Disaggregating User Evaluations Using the Shapley Value. In *NetEcon '18: Economics of Networks, Systems and Computation, June 18, 2018, Irvine, CA, USA*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3230654.3230659

*This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Centre "On-The-Fly Computing" (SFB 901).

NetEcon '18, June 18, 2018, Irvine, CA, USA

ACM ISBN 978-1-4503-5916-0/18/06…$15.00
https://doi.org/10.1145/3230654.3230659

1 INTRODUCTION

An emerging demand for all-in-one services requires providers to compose their services out of a number of elementary services. For instance, cloud service providers like Google, Amazon, or Microsoft combine hardware resources with (several) software services and sell it as a composed service. Similarly, restaurants sell a composition of service and product (food). However, for the customer or user, it may be difficult to exactly identify the single parts in her good. As a consequence, she may not be able to observe their performances separately, when she experiences her good. For instance a "slow" computation service could result either from an inappropriate hardware or from inefficient algorithms.

Internet platforms offer a possibility to give evaluations for (composed) services or products, so that later users can base their purchase decision on it. But, as the components of a service cannot be identified, the user can only evaluate the composed service as a whole. However, not only from the provider's perspective, it would be worthwhile to know, how we can infer from evaluation data on compositions on an evaluation of single services. The very fact that elementary services can be composed and sold in many different compositions allows us to assess how well a component service fits. Such a disaggregation of evaluations, e.g., helps to sort out inefficient elementary services, so that they should not be considered in the composition process.

As we work with evaluations on compositions from many users, we bring together an *aggregation* operation, that aggregates evaluations across users, and a *disaggregation* process, that evaluates each basic service on the basis of how the different compositions are rated.

In the pursue of finding a plausible combination, we present a set of desirable properties for the interplay of aggregation and disaggregation. Two properties are central in the discussion. A *restricted influence* requirement should forbid users to unilaterally change a service's overall evaluation to the best or worst possible valuation. *Consistency* of our combined aggregation/disaggregation process requires that the order should not matter, i.e., the evaluation of an elementary service should not depend on whether we first aggregate user evaluations of compositions and then disaggregate or whether we first disaggregate each single user's valuation to her valuation of basic services and then aggregate these data.

For the disaggregation process we make use of techniques from cooperative game theory. A complete evaluation of composed services can be represented by a real-valued non-linear set function, which is termed cooperative game with transferable utility, TU game in short, in the realms of game theory. In our analysis we use *the* best-known solution concept for TU games, namely the Shapley value [24]. It allows us to retrieve an (average) influence of a component service on the evaluation of composed services. Hence, it disaggregates evaluation information on compositions to valuations of components.

1.1 Related Work

Aggregation functions are studied in both theoretical and empirical work. The most commonly used operators are the (weighted) mean, the median and the mode. An overview of their theoretical properties can be found in [14] and [6]. [11] compares the three aggregators empirically according to three different criteria, namely informativeness, robustness and strategyproofness. [12] extends this comparison to accuracy. They find evidence that although the mean is widely used it is not always the preferable aggregator regarding these five criteria. In an experiment the aggregators' influence on the users' rating behavior is studied [13]. It is shown that under certain circumstances the mean is preferable in terms of rating behavior. Different aggregators than the mean, median or mode are used by [8] who define an econometric framework to aggregate consumer preferences from online product reviews. They particularly take the heterogeneity of users' opinions, e.g. experts and non-professionals, into account when aggregating the evaluations. Another aspect of aggregation tackles the question how to get the true quality of a (composed) product [22].

In contrast to the widely studied aggregation process, the disaggregation process lacks this variety of studies. Although the problem of common evaluations for combined products or a similar problem with common evaluations of a group of firms selling regional products is known [27], few studies discuss actually the disaggregation of these evaluations. Nevertheless, there are few studies considering the Shapley value to disaggregate users' ratings of component services directly [18]. To decrease computational complexity a variant of the Shapley value might be used [19]. Additionally, researchers investigate the learning of classes of games using the PAC model, i.e. [4].

To solve the disaggregation part of our problem we use the Shapley value, the best-known solution concept for TU games. This concept is widely studied and characterized by some natural properties, namely symmetry, Pareto efficiency, additivity and the dummy player axiom [24]. Alternative characterizations use a fairness [26] or a transfer property [10] instead of additivity. [5] and [1] deal with incomplete information in the context of the Shapley value. Similar concepts have been used by Smets in the context of decision making in the transferable belief model with pignistic transformations [25].

Sampling for the approximation of Shapley values was already introduced by [21], but without any theoretical guarantees. This gap was filled for different classes of games: [3] propose an analysis for simple games in the context of power indices and [2] for matching games. If we restrict the games by the value function instead of any structural properties, [17] show the approximation guarantee for supermodular games and [5] for submodular games with bounded curvature. Additionally, [7] and [20] analyze the approximation in a more general setting, but require a given variance or range of the marginal values. Further approximation methods for the Shapley value have been discussed in [9, 16, 23, 28]

1.2 Our Contribution

In this paper, we propose a model and a solution to handle the aggregation of reputation values from different users and the disaggregation of information on composed services. To the best of our knowledge, this is the first work which handles both steps, aggregation and disaggregation, *together*. Beside the characterization of this evaluation problem through some natural axioms, we show that a combination of the Shapley value as disaggregator and the (weighted) average as aggregator satisfies our normative requirements (Theorem 3.1). In contrast, using classical rules from social choice (maximum, minimum, or median rule) violate them (Propositions 2.5-2.8). Concerning computationally complexity, we give an algorithm approximating the Shapley value in polynomial time (Theorem 4.3). Finally, we address the problem of missing valuations and apply the Data-Dependent Shapley Value [5], which, together with the weighted average as aggregator, still yields a solution that satisfies our requirements (Theorem 5.1). To this end, our combination proves to be a reasonable and robust choice to elicit evaluations of component services.

2 THE MODEL

There are m basic services which can be combined to $2^m - 1$ different service compositions. Furthermore, n users can buy and evaluate these service compositions. We assume that each user i evaluates all compositions once[1], assigning a real number from a predefined scale $[0, u]$ to each service composition $S \subseteq \{1, \ldots, m\}, S \neq \emptyset$. The best evaluation is u, the worst is 0. The matrix E collects all available evaluations, hence $E \in \mathcal{M}\,(n \times 2^m - 1)$ and E_{iS} is user i's evaluation for service composition S $(i = 1, \ldots, n, S \subseteq \{1, \ldots, m\}, S \neq \emptyset)$. A solution (in the wider sense) to the problem of evaluating basic services assigns to each basic service a real number, which can be collected in a real valued matrix $e \in \mathcal{M}\,(1 \times m)$. Note that we do not require these values to lie in the interval $[0, u]$.

To find an appropriate single evaluation for each basic service, we need two different operators. A disaggregator is a mapping $D^{(q)} : \mathcal{M}\,(q \times 2^m - 1) \to \mathcal{M}\,(q \times m)$ that takes each of q (possibly different) evaluations of compositions and computes the corresponding q evaluations of the m basic services. An aggregator is a mapping $A^{(p)} : \mathcal{M}\,(n \times p) \to \mathcal{M}\,(1 \times p)$ that maps the evaluations of n users (either over $p = 2^m - 1$ service compositions or over $p = m$ basic services) to an aggregated evaluation. In the subsequent discussion, we use different specifications of $D^{(q)}$ and $A^{(p)}$. In what follows, we assume that aggregation and disaggregation is anonymous in the sense that for an aggregator A (an instance of $A^{(p)}$) and evaluation matrix E the aggregated value in the i-th coordinate $(A(E))_i$ only depends on the i-th column of E and is calculated using the same function for each of the p coordinate. Similarly, a disaggregator D (an instance of $D^{(q)}$) is assumed to

[1]This is quite a strong assumption. However, a user in our model can also represent a group of similar/identical users who evaluate all compositions together.

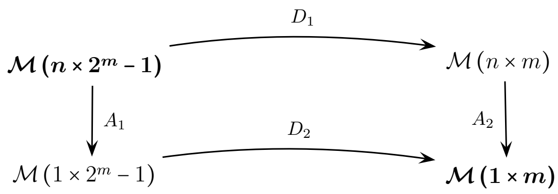

Figure 1: Scheme of the evaluation problem

apply the same function for each of the q users that only depends on the corresponding row in the underlying evaluation matrix.

Two routes can be taken to transform an evaluation matrix E to a final evaluation of basic services (cf. Figure 1). On the one hand, it is possible to start with the aggregation of evaluations for each composed service. Applying the aggregator A_1 (a specification of $A^{(2^m-1)}$) on the input matrix results in the intermediate matrix $A_1(E) \in \mathcal{M}(1 \times 2^m - 1)$. A disaggregator D_2 (a specific $D^{(1)}$) then yields the final evaluation $D_2(A_1(E)) \in \mathcal{M}(1 \times m)$ of basic services.

On the other hand, one could also reverse the order of aggregation and disaggregation. Now, we first disaggregate each user's evaluation to an evaluation over basic services. Then, these evaluations are aggregated across users. Formally, we first apply D_1 (some $D^{(n)}$) to an evaluation matrix E resulting in $D_1(E) \in \mathcal{M}(n \times m)$ before aggregator A_2 (some $A^{(m)}$) yields the final evaluations $A_2(D_1(E)) \in \mathcal{M}(1 \times m)$. Both routes yield overall evaluations for the single services, which should not be restricted to the same range as the initial evaluations. We define $\underline{R}_{D_2,A_1}$ and $\overline{R}_{D_2,A_1}$ so that the range of $(D_2(A_1(\cdot)))_j$ is $[\underline{R}_{D_2,A_1}, \overline{R}_{D_2,A_1}]$. Analogously, $\underline{R}_{A_2,D_1}$ and $\overline{R}_{A_2,D_1}$ are defined such that the range of $(A_2(D_1(\cdot)))_j$ is $[\underline{R}_{A_2,D_1}, \overline{R}_{A_2,D_1}]$ $(j = 1, \ldots, m)$).[2]

2.1 Axioms

A *solution* to the problem of assigning valuations for basic services is a collection of aggregators A_1, A_2 and disaggregators D_1, D_2. In other words, a solution includes both types of operations, aggregation of values for compositions (A_1) and basic services (A_2) and disaggregation of valuations from single users (D_1) and from aggregated valuations (D_2).

We next formulate normative requirements (axioms) that a solution, i.e., the combination of aggregators and disaggregators, should satisfy. The first key axiom is immediate and requires that the solution is independent of the order of aggregation and disaggregation.

Axiom 2.1 (CONSISTENCY). A solution to the evaluation problem is *consistent*, if it does not depend on the order of aggregation and disaggregation, i.e., $A_2(D_1(E)) = D_2(A_1(E))$ for each evaluation matrix E.

Apart from technical considerations (e.g., dynamic updates, etc.), without this requirement the ordering becomes a strategic question. The platform calculating single evaluations could influence the outcome by choosing the order of the aggregation and disaggregation in its favor.

[2]Note that with a continuous aggregator A and disaggregator D, the composed functions $(A(D(\cdot)))_j$ and $(D(A(\cdot)))_j$ are continuous and therefore the image of the compact set of evaluation matrices with entries in $[0, u]$ must be compact.

The next two axioms capture extreme cases with one user or one service only, which makes either aggregation or disaggregation superfluous. Given a single user, $n = 1$, there is no difference between her evaluations and the aggregated ones. Hence, disaggregation should not give different results.

Axiom 2.2 (SINGLE USER). A solution to the evaluation problem fulfills the *single user axiom* if in case $n = 1$ the disaggregation functions are equal, $D_1 \hat{=} D_2$, D in short, in order to yield the same result.

Given only one single service, $m = 1$, the disaggregation step is vacuous.

Axiom 2.3 (SINGLE SERVICE). A solution of the evaluation problem fulfills the *single service axiom* if in case $m = 1$ the aggregation functions are equal, $A_1 \hat{=} A_2$, shortly A, in order to yield the same aggregation result.

Recall that our anonymity requirement from above together with the two previous axioms ensures that aggregation and disaggregation does not vary over coordinates.

We conclude our requirements section with one last aspect. A user might want to manipulate the final evaluation of a basic service with the help of her individual evaluation. Common aims of a manipulation are that a single service receives an extremely low or extremely high final evaluation. Therefore, the aggregation and disaggregation process should rule out these possibilities of manipulation. If a basic service is not already evaluated extremely high or extremely low, a newly evaluating user should not be able to choose her evaluations in such a way that the overall valuation of this basic service becomes extremely high or extremely low.

Axiom 2.4 (RESTRICTED INFLUENCE). Given evaluations E, a range $[0, u]$ and a service j with overall evaluations $(D_2(A_1(E)))_j \neq \overline{R}_{D_2,A_1}$, $(D_2(A_1(E)))_j \neq \underline{R}_{D_2,A_1}$, $(A_2(D_1(E)))_j \neq \overline{R}_{A_2,D_1}$ and $(A_2(D_1(E)))_j \neq \underline{R}_{A_2,D_1}$, a solution fulfills the *restricted influence* axiom, if there exists no $v \in \mathbb{R}^{2^m-1}$ and $\hat{E} = \begin{pmatrix} E \\ v \end{pmatrix}$ such that $(D_2(A_1(\hat{E})))_j = \overline{R}_{D_2,A_1}$, $(D_2(A_1(\hat{E})))_j = \underline{R}_{D_2,A_1}$, $(A_2(D_1(\hat{E})))_j = \overline{R}_{A_2,D_1}$ and $(A_2(D_1(\hat{E})))_j = \underline{R}_{A_2,D_1}$.

2.2 (Dis-)Aggregations with Positional Operators

From a sequence of numbers, positional operators select the number that appears at a prespecified position after ordering the numbers. The most popular examples are the median, maximum and minimum operator, which are frequently used in social choice problems as an aggregation device. We investigate how far they are suitable for our goals.

First, we discuss aggregation and disaggregation with the median. The median is defined as the value dividing the evaluations in the lower and the higher half. It is often used to aggregate values because it is a robust statistical measure as it is not affected by outliers. But is it also a suitable disaggregation tool? It seems convincing that the valuation of a basic service j is described by the "middle value" between the evaluations of all composed services that include j. Taking the median as a tool to aggregate and disaggregate evaluations we can directly see, that the single user and the single service axioms are fulfilled as we use the same type of function for all operations. Unfortunately, we can easily see that using the median as an aggregation and disaggregation function will not always yield a consistent solution. If the number of evaluations is odd, the median is simply the value in the middle when the evaluations are arranged from the lowest to the highest value. If the number of evaluations is even, the median is defined by the mean of the two middle values.

Proposition 2.5. *The solution that uses the median as aggregator and as disaggregator is not consistent.*

Next, we aggregate and disaggregate using the minimum or maximum operator. When choosing the minimum to disaggregate, the evaluation of a basic service j is equal to the evaluation of the worst composed service including j. When choosing the minimum to aggregate, the aggregate evaluation of a basic or composed service is the worst evaluation across users. Analogously, we can define (dis-)aggregation with the maximum operator. However, combining maximum and the minimum contradicts the axioms.

Proposition 2.6. *The solution that uses the minimum to aggregate and the maximum to disaggregate (the minimum to disaggregate and the maximum to aggregate), i.e., $A = \min\{\cdot\}$ and $D = \max\{\cdot\}$, ($D = \min\{\cdot\}$ and $A = \max\{\cdot\}$), is not consistent.*

Let us now consider the cases where $A = D = \min\{\cdot\}$ and $A = D = \max\{\cdot\}$. Economically, the solutions can be interpreted as a lower and upper bound for the evaluations.

Proposition 2.7. *Using the minimum (maximum) to aggregate and to disaggregate, $A = D = \min\{\cdot\}$ ($A = D = \max\{\cdot\}$) yields a consistent solution.*

Although we can achieve consistency using the minimum or maximum, the solutions are manipulable.

Proposition 2.8. *The solution using the minimum (maximum) to aggregate and to disaggregate, $A = D = \min\{\cdot\}$ ($A = D = \max\{\cdot\}$) violates the restricted influence axiom.*

The maximum and minimum solutions can easily be manipulated: Each costumer who evaluates the (composed) services can determine the (dis-)aggregated evaluation of a single service j by simply assigning the best or worst possible evaluation to one of the composed services T with $j \in T$.

3 DISAGGREGATION WITH THE SHAPLEY VALUE AND AGGREGATION WITH THE WEIGHTED AVERAGE

Although minimum and maximum operators are easy to compute, they are proven to generate inappropriate solutions that do not meet all our axioms. Instead, we suggest to follow a different approach to aggregate and disaggregate. For aggregation we use the (weighted) average. As argued above, the weighted average is, besides the median, the most examined aggregation concept. Online reputation systems as on Amazon or eBay provide customers with an average of evaluations as an aggregate signal. For disaggregation we use a well-known TU game solution concept, the Shapley value. This concept is heavily used in game-theoretic models as it is characterized by intuitive properties, namely symmetry, Pareto efficiency, additivity and the dummy player axiom [24]. Applied in our scenario, the Shapley value for basic service j gives the average marginal contribution of j over any possible composition and is formally defined as follows:

$$\Phi_j(\hat{E}, i) = \sum_{S \subseteq \{1,\ldots,m\} \setminus \{j\}} \frac{(m-|S|-1)! \cdot |S|!}{m!} \left(\hat{E}_{iS\cup\{j\}} - \hat{E}_{iS}\right) \tag{1}$$

The Shapley value as a disaggregating function when evaluations $\hat{E}$ are given is defined as follows.

$$D^{(q)}(\hat{E}) = \begin{pmatrix} \Phi(\hat{E}, 1) \\ \vdots \\ \Phi(\hat{E}, q) \end{pmatrix} \tag{2}$$

with $\Phi(\hat{E}, i) = (\Phi_1(\hat{E}, i), \ldots, \Phi_m(\hat{E}, i))$ where $\Phi_j(\hat{E}, i)$ is user i's Shapley value for the basic service j with (or the aggregated Shapley value if $1 = i = q$).

As aggregator, we define the weighted average with user specific weights β_i, $\sum_{i=1}^{n} \beta_i = 1$ over evaluations $\bar{E}$:

$$A^{(p)}(\bar{E}) \text{ with } (A^{(p)}(\bar{E}))_T = \sum_{i=1}^{n} \beta_i \bar{E}_{iT} \quad \forall T \subseteq \{1, \ldots .m\}. \tag{3}$$

Using the weighted mean to aggregate and the Shapley value to disaggregate obviously fulfills the single service and single axioms.

Theorem 3.1. *The solution that uses the Shapley value as disaggregator and the weighted average as aggregator as defined in Equation 1 to 3 is consistent and fulfills the restricted influence axiom.*

Proof. To see consistency, we use the linearity of the Shapley value (in entries of E) and calculate

$$\begin{aligned}
&(D(A(E)))_j \\
&= \sum_{S \subseteq \{1,\ldots,m\} \setminus \{j\}} \frac{(m-|S|-1)! \cdot |S|!}{m!} \left([A(E)]_S - [A(E)]_{S\setminus\{j\}}\right) \\
&= \sum_{S \subseteq \{1,\ldots,m\} \setminus \{j\}} \frac{(m-|S|-1)! \cdot |S|!}{m!} \left(\sum_{i=1}^{n} \beta_i E_{iS\cup\{j\}} - \sum_{i=1}^{n} \beta_i E_{iS}\right) \\
&= \sum_{i=1}^{n} \beta_i \sum_{S \subseteq \{1,\ldots,m\} \setminus \{j\}} \frac{(m-|S|-1)! \cdot |S|!}{m!} \left(E_{iS\cup\{j\}} - E_{iS}\right) \\
&= \sum_{i=1}^{n} \beta_i [D(E)]_{ij} = (A(D(E)))_j
\end{aligned}$$

It is easy to see that this solution also fulfills the restricted influence axiom. If the users' evaluations lie within the range $[0, u]$, taking the mean does not change this range. By using the Shapley value the range changes to $[-u, u]$. Given evaluations E such that after the first disaggregation step the evaluation of a basic service j is $(D(E))_{ij} \neq u$ and $(D(E))_{ij} \neq -u$ for all users i. A new user $n + 1$ gives evaluations v which results in $(D(v))_j$. The new evaluation matrix is given by $\hat{E} = \begin{pmatrix} E \\ v \end{pmatrix}$. By aggregating all $(D(E))_{ij}$ and $(D(v))_j$ to $(A(D(\hat{E})))_j$ with the weighted average, the aggregated value can never be equal to u or $-u$, $(A(D(\hat{E})))_j \neq u$ and $(A(D(\hat{E})))_j \neq -u$. As using the weighted mean and the Shapley value is consistent, we conclude that $(D(A(\hat{E})))_j \neq u$ and $(D(A(\hat{E})))_j \neq -u$ as well, showing that the restricted influence axiom is satisfied. □

Finally, we note that the range of the Shapley values of a basic service is $[-u, u]$. To see this, assume that the users' evaluations have already been aggregated. The highest marginal contribution of a basic service j is u, which is received when j is added to a composition that is evaluated with 0. Therefore the Shapley value for j is only equal to u, when all compositions including j are evaluated with u and all compositions without j are valued 0. Similarly, basic service j receives a Shapley value of $-u$, when it is the other way round (compositions valued 0, when j is included and u, when j is not included. The change of scale from $[0, u]$ to $[-u, u]$ necessitates a different interpretation of overall evaluations. We comment on this in the last section.

4 APPROXIMATION THROUGH SAMPLING

The exact computation of the Shapley value is computationally hard in general since the number of coalitions exponentially grows with the number of basic services. Therefore, computing the solution of our evaluation problem could take take too long for practical scenarios. Nevertheless, by using sampling methods [3, 20, 21] we can achieve at least a reasonable approximation for our reputation values with high probability. Indeed, we can construct a fully polynomial-time randomized approximation scheme (FPRAS) for our evaluation problem. The only assumption needed for this approach are bounds on the possible marginal contribution of a service to a composition. As explained in the previous section, for the setting of reputation values, these bounds are naturally given by $[-u,\ u]$

Algorithm 4.1 (Sampling Algorithm). Given an arbitrary $\mu > 0$ and $\delta > 0$: Generate $\frac{2\ln(8)u^2}{\mu^2}$ random permutations of the services i.i.d. and for each permutation compute the marginal contribution $MC_r = \left(\hat{E}_{iS\cup\{j\}} - \hat{E}_{iS}\right)$ for service j. Let $\overline{MC}$ be the average marginal contribution of service j over these k permutations. Repeat the process $l = \ln(1/\delta)$ times and return the median of all $\overline{MC}$ as $\tilde{\Phi}_j(\hat{E}, i)$.

The following Lemma 4.2 states the approximation guarantee for Algorithm 4.1.

LEMMA 4.2. *Given any constants $\mu > 0$ and $\delta > 0$, Algorithm 4.1 computes a μ-approximation of $\Phi_j(\hat{E}, i)$ for any service j in polynomial time with probability at least $1-\delta$, i. e. $Pr[|\tilde{\Phi}_j(\hat{E}, i) - \Phi_j(\hat{E}, i)| \leq \mu] \geq 1 - \delta$.*

Let $\tilde{\Phi}(\hat{E}, i) = (\tilde{\Phi}_1(\hat{E}, i), \ldots, \tilde{\Phi}_m(\hat{E}, i))$ and

$$\tilde{D}^{(q)}(\hat{E}) = \begin{pmatrix} \tilde{\Phi}(\hat{E}, 1) \\ \vdots \\ \tilde{\Phi}(\hat{E}, q) \end{pmatrix}. \tag{4}$$

THEOREM 4.3. *Given any constant $\mu > 0$, $A(\tilde{D}(E))$ and $\tilde{D}(A(E))$ can be computed in polynomial time and $|A(\tilde{D}(E)) - e| \leq \mu$ and $|\tilde{D}(A(E)) - e| \leq \mu$ with high probability.*

PROOF. Let $\delta = (n^{c+1} \cdot m^{c+1})^{-1}$ for any constant c. First aggregating and then disaggregating only needs m applications of our sampling algorithm as last step (one for each service). With Lemma 4.2 we have for each service j that $Pr\left[|\tilde{\Phi}_j(\hat{E}, 1) - \Phi_j(\hat{E}, 1)| \geq \mu\right] \leq \left(n^{c+1} \cdot m^{c+1}\right)^{-1}$. Using the union bound since all approximations are independent results in $Pr\left[\exists j \in \{1, \ldots, m\} : |\tilde{\Phi}_j(\hat{E}, 1) - \Phi_j(\hat{E}, 1)| \geq \mu\right] \leq m\left(n^{c+1} \cdot m^{c+1}\right)^{-1} = \left(n^{c+1} \cdot m^{c}\right)^{-1}$.

For the other direction, we have to look at the two steps in more detail: In the disaggregation phase, we compute $n \cdot m$ independent approximations of the Shapley value. Again with Lemma 4.2 we can show that: $Pr\left[\exists 1 \leq i \leq n : \exists 1 \leq j \leq m : |\tilde{\Phi}_j(\hat{E}, i) - \Phi_j(\hat{E}, i)| \geq \mu\right] \leq n \cdot m \cdot n^{-c} \cdot m^{-c} = n^{-c} \cdot m^{-c}$. Taking the average over all player i does neither affect the failure probability nor the range of values and the theorem is shown. □

If the range of the marginal contributions is not known, i.e., if we apply our framework for other settings than reputation, we can use a similar analysis. Sampling methods which require the variance of the distribution of marginal contributions [20] or even structural properties like supermodularity [17] or submodularity with bounded curvature [5] can be analyzed similarly resulting in another number of required samples.

5 INCOMPLETE EVALUATION PROFILES

A natural question which arises in our work is the handling of incomplete information. Until now we assumed complete evaluations. But in reality, often each participating user only rates a subset of all possible service compositions. To tackle this problem we follow the line of research recently started by [5] and apply the so called *Data Dependent Shapley Value.*

Here we assume a distribution $\mathcal{D}$ which captures the frequency of the different compositions. The well-know Shapley axioms can be extended with regard to this distribution $\mathcal{D}$ and the value itself can be defined as

$$\Phi_j^{\mathcal{D}}(\hat{E}, i) = \sum_{S: j \in S} Pr[S \sim \mathcal{D}] \cdot \frac{E_{iS}}{|S|}. \tag{5}$$

Let $\Phi^{\mathcal{D}}(\hat{E}, i) = (\Phi_1^{\mathcal{D}}(\hat{E}, i), \ldots, \Phi_m^{\mathcal{D}}(\hat{E}, i))$ and

$$D^{\mathcal{D}(q)}(\hat{E}) = \begin{pmatrix} \Phi^{\mathcal{D}}(\hat{E}, 1) \\ \vdots \\ \Phi^{\mathcal{D}}(\hat{E}, q) \end{pmatrix}. \tag{6}$$

THEOREM 5.1. *Given evaluations E whose frequency follows distribution $\mathcal{D}$, using the Data-Dependent Shapley Value to disaggregate and the weighted average to aggregate as defined in Equation 6 and 3, yields a consistent solution that fulfills the restricted influence axiom.*

The proof follows similar arguments as in the proof of Theorem 3.1. Since the distribution of the evaluations is known in this extension, the sampling algorithm from the previous section can be extended. The samples are now drawn from the underlying distribution $\mathcal{D}$ and again with help of the Hoeffding's inequality [15] the number of samples can be estimated.

COROLLARY 5.2. *Given evaluations E whose frequency follows distribution $\mathcal{D}$ and any constant $\mu > 0$, $A(D^{\mathcal{D}}(E))$ and $D^{\mathcal{D}}(A(E)))$ can be computed in polynomial time and $|A(D^{\mathcal{D}}(E)) - e| \leq \mu$ and $|D^{\mathcal{D}}(A(E)) - e| \leq \mu$ with high probability.*

6 CONCLUSION

We addressed the problem of eliciting information on the quality of components of a composed service (or product). As input we use evaluations of many users on compositions. Therefore, the problem is to find the right combination of aggregating over users and disaggregating evaluations for bundles of services. A reasonable (in the sense of the stated axioms) way to go is to combine the Shapley value from cooperative game theory as disaggregator with the weighed average as aggregator. Moreover, it is efficient and robust as the final valuations can be approximated in polynomial time, even if the evaluation matrix is incomplete.

One should not fail to note that the final valuations for basic services have to be interpreted in a specific way. While initial user evaluations range in $[0, u]$, final values are from the interval $[-u, u]$. Interpreting evaluations as a proxy for quality, a positive Shapley value of service j suggests that it improves the quality of a composition, while a negative one is rather deteriorating quality. In particular, two services that may be used as substitutes can be ranked in terms of quality enhancement by their Shapley values. The actual value in the overall solution has to be considered in relation to the values for other services. For instance, if all user evaluations for all compositions are u, then the final solution attributes u/m to each component service. If valuations for some service and compositions that include it fall, then the final values for all other services increase. Consequently, the interpretation of final values necessarily has to include the whole picture.

We close by mentioning that our method can be applied in other scenarios as well. For instance, instead of using services and service compositions, we can think of workers and teams in a company. Our approach can then be used to measure the effectivity of a worker within a team on the basis of performance values for the different possible team compositions. The design of bonus systems, or employment policies are possible fields of application. Another direction to use our framework is the understanding of learning algorithms. Imagine a complex learning environment, in which we compose the learning tools out of basic components like pre- and postprocessors, different learning algorithms, computing environments and also training data. Here, we can often measure only the performance of the whole tool, but we need to make statements about the contribution of the different involved components to improve the learning tools.

REFERENCES

[1] Néstor E. Aguilera, Silvia C. Di Marco, and Mariana S. Escalante. 2010. The Shapley value for arbitrary families of coalitions. *European Journal of Operational Research* 204, 1 (2010), 125–138.

[2] Haris Aziz and Bart de Keijzer. 2014. Shapley meets Shapley. In *31st International Symposium on Theoretical Aspects of Computer Science (STACS 2014) (LIPIcs)*, Ernst W. Mayr and Natacha Portier (Eds.), Vol. 25. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 99–111.

[3] Yoram Bachrach, Evangelos Markakis, Ezra Resnick, Ariel D. Procaccia, Jeffrey S. Rosenschein, and Amin Saberi. 2010. Approximating power indices: theoretical and empirical analysis. *Autonomous Agents and Multi-Agent Systems* 20, 2 (2010), 105–122.

[4] Maria-Florina Balcan, Ariel D. Procaccia, and Yair Zick. 2015. Learning Cooperative Games. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, Qiang Yang and Michael Wooldridge (Eds.). AAAI Press, 475–481.

[5] Eric Balkanski, Umar Syed, and Sergei Vassilvitskii. 2017. Statistical Cost Sharing. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 6222–6231.

[6] Tomasa Calvo, Anna Kolesárová, Magda Komorníková, and Radko Mesiar. 2002. *Aggregation Operators: Properties, Classes and Construction Methods.* Physica-Verlag HD, Heidelberg, 3–104.

[7] Javier Castro, Daniel Gómez, and Juan Tejada. 2009. Polynomial calculation of the Shapley value based on sampling. *Computers & OR* 36, 5 (2009), 1726–1730.

[8] Reinhold Decker and Michael Trusov. 2010. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing* 27, 4 (2010), 293 – 307.

[9] S. Shaheen Fatima, Michael Wooldridge, and Nicholas R. Jennings. 2008. A linear approximation method for the Shapley value. *Artif. Intell.* 172, 14 (2008), 1673–1699.

[10] Vincent Feltkamp. 1995. Alternative axiomatic characterizations of the Shapley and Banzhaf values. *International Journal of Game Theory* 24, 2 (01 Jun 1995), 179–186.

[11] Florent Garcin, Boi Faltings, and Radu Jurca. 2009. Aggregating reputation feedback. In *Proceedings of the First International Conference on Reputation: Theory and Technology.* 62–74.

[12] Florent Garcin, Boi Faltings, Radu Jurca, and Nadine Joswig. 2009. Rating aggregation in collaborative filtering systems. In *Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys*, Lawrence D. Bergman, Alexander Tuzhilin, Robin D. Burke, Alexander Felfernig, and Lars Schmidt-Thieme (Eds.). ACM, 349–352.

[13] Florent Garcin, Lirong Xia, and Boi Faltings. 2013. How aggregators influence human rater behavior. In *Proceedings Workshop at 14th ACM Conference on Electronic Commerce (EC-13).*

[14] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. 2011. Aggregation functions: Means. *Inf. Sci.* 181, 1 (2011), 1–22.

[15] Wassily Hoeffding. 1963. Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.* 58, 301 (1963), 13–30.

[16] Dennis Leech. 2003. Computing Power Indices for Large Voting Games. *Management Science* 49, 6 (2003), 831–837.

[17] David Liben-Nowell, Alexa Sharp, Tom Wexler, and Kevin Woods. 2012. Computing Shapley Value in Supermodular Coalitional Games. In *Computing and Combinatorics - 18th Annual International Conference, COCOON. Proceedings (Lecture Notes in Computer Science)*, Joachim Gudmundsson, Julián Mestre, and Taso Viglas (Eds.), Vol. 7434. Springer, 568–579.

[18] An Liu, Qing Li, Liusheng Huang, and Shiting Wen. 2012. Shapley Value Based Impression Propagation for Reputation Management in Web Service Composition. In *2012 IEEE 19th International Conference on Web Services, Honolulu, HI, USA, June 24-29, 2012*, Carole A. Goble, Peter P. Chen, and Jia Zhang (Eds.). IEEE Computer Society, 58–65.

[19] An Liu, Qing Li, Xiaofang Zhou, Lu Li, Guanfeng Liu, and Yunjun Gao. 2014. Rating Propagation in Web Services Reputation Systems: A Fast Shapley Value Approach. In *Database Systems for Advanced Applications - 19th International Conference, DASFAA 2014, Bali, Indonesia, April 21-24, 2014. Proceedings, Part I (Lecture Notes in Computer Science)*, Sourav S. Bhowmick, Curtis E. Dyreson, Christian S. Jensen, Mong-Li Lee, Agus Muliantara, and Bernhard Thalheim (Eds.), Vol. 8421. Springer, 466–480.

[20] Sasan Maleki. 2015. *Addressing the computational issues of the Shapley value with applications in the smart grid.* Ph.D. Dissertation. University of Southampton, UK.

[21] Irwin Mann and Lloyd S Shapley. 1962. *Values of Large Games. 6: Evaluating the Electoral College Exactly.* Technical Report. DTIC Document.

[22] Mary McGlohon, Natalie S. Glance, and Zach Reiter. 2010. Star Quality: Aggregating Reviews to Rank Products and Merchants. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM*, William W. Cohen and Samuel Gosling (Eds.). The AAAI Press.

[23] Guillermo Owen. 1972. Multilinear Extensions of Games. *Management Science* 18, 5-part-2 (1972), 64–79.

[24] Lloyd S Shapley. 1953. A value for n-person games. In *Ann. Math. Studies.* Vol. 28. Princeton University Press, Princeton, 307–317.

[25] Philippe Smets. 2005. Decision making in the TBM: the necessity of the pignistic transformation. *Int. J. Approx. Reasoning* 38, 2 (2005), 133–147.

[26] René van den Brink. 2002. An axiomatization of the Shapley value using a fairness property. *Int. J. Game Theory* 30, 3 (2002), 309–319.

[27] Jason A. Winfree and Jill J. McCluskey. 2005. Collective Reputation and Quality. *American Journal of Agricultural Economics* 87, 1 (2005), 206–213.

[28] Gilad Zlotkin and Jeffrey S. Rosenschein. 1994. Coalition, Cryptography, and Stability: Mechanisms for Coalition Formation in Task Oriented Domains. In *Proceedings of the 12th National Conference on Artificial Intelligence, Seattle, WA, USA, July 31 - August 4, 1994, Volume 1.*, Barbara Hayes-Roth and Richard E. Korf (Eds.). AAAI Press / The MIT Press, 432–437.