Shapley value

The **Shapley value** is a solution concept in cooperative game theory. It was named in honor of Lloyd Shapley, who introduced it in 1951 and won the Nobel Memorial Prize in Economic Sciences for it in 2012.[1][2] To each cooperative game it assigns a unique distribution (among the players) of a total surplus generated by the coalition of all players. The Shapley value is characterized by a collection of desirable properties. Hart (1989) provides a survey of the subject.[3][4]


Lloyd Shapley in 2012

The setup is as follows: a coalition of players cooperates, and obtains a certain overall gain from that cooperation. Since some players may contribute more to the coalition than others or may possess different bargaining power (for example threatening to destroy the whole surplus), what final distribution of generated surplus among the players should arise in any particular game? Or phrased differently: how important is each player to the overall cooperation, and what payoff can he or she reasonably expect? The Shapley value provides one possible answer to this question.

For cost-sharing games with concave cost functions, the optimal cost-sharing rule that optimizes the price of anarchy, followed by the price of stability, is precisely the Shapley value cost-sharing rule.[5] (A symmetrical statement is similarly valid for utility-sharing games with convex utility functions.) In mechanism design, this means that the Shapley value solution concept is optimal for these sets of games.

Contents

Formal definition
- In terms of synergy

Examples
- Business example
- Glove game

Properties
- Efficiency
- Symmetry
- Linearity
- Null player
- Stand-alone test
- Anonymity
- Marginalism

Characterization

Aumann–Shapley value

Generalization to coalitions

Value of a player to another player
In machine learning
See also
References
Further reading
External links

Formal definition

Formally, a **coalitional game** is defined as: There is a set *N* (of *n* players) and a function v that maps subsets of players to the real numbers: $v\colon 2^N \to \mathbb{R}$, with $v(\emptyset) = 0$, where $\emptyset$ denotes the empty set. The function v is called a characteristic function.

The function v has the following meaning: if *S* is a coalition of players, then $v(S)$, called the worth of coalition *S*, describes the total expected sum of payoffs the members of S can obtain by cooperation.

The Shapley value is one way to distribute the total gains to the players, assuming that they all collaborate. It is a "fair" distribution in the sense that it is the only distribution with certain desirable properties listed below. According to the Shapley value,[6] the amount that player *i* is given in a coalitional game (v, N) is

$$\begin{aligned}\varphi_i(v) &= \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\;(n-|S|-1)!}{n!}(v(S \cup \{i\}) - v(S)) \\ &= \sum_{S \subseteq N \setminus \{i\}} \binom{n}{1, |S|, n-|S|-1}^{-1} (v(S \cup \{i\}) - v(S))\end{aligned}$$

where *n* is the total number of players and the sum extends over all subsets *S* of *N* not containing player *i*. Also note that $\binom{n}{a, b, c}$ is the multinomial coefficient. The formula can be interpreted as follows: imagine the coalition being formed one actor at a time, with each actor demanding their contribution $v(S \cup \{i\}) - v(S)$ as a fair compensation, and then for each actor take the average of this contribution over the possible different permutations in which the coalition can be formed.

An alternative equivalent formula for the Shapley value is:

$$\varphi_i(v) = \frac{1}{n!} \sum_R \left[v(P_i^R \cup \{i\}) - v(P_i^R)\right]$$

where the sum ranges over all $n!$ orders R of the players and P_i^R is the set of players in N which precede i in the order R. Finally, it can also be expressed as

$$\varphi_i(v) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \binom{n-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S))$$

which can be interpreted as

$$\varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding } i} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}}$$

In terms of synergy

From the characteristic function v one can compute the *synergy* that each group of players provides. The synergy is the unique function $w\colon 2^N \to \mathbb{R}$, such that

$$v(S) = \sum_{R \subseteq S} w(R)$$

for any subset $S \subseteq N$ of players. In other words, the 'total value' of the coalition S comes from summing up the *synergies* of each possible subset of S.

Given a characteristic function v, the synergy function w is calculated via

$$w(S) = \sum_{R \subseteq S} (-1)^{|S|-|R|} v(R)$$

using the Inclusion exclusion principle. In other words, the synergy of coalition S is the value $v(S)$, which is not already accounted for by its subsets.

The Shapley values are given in terms of the synergy function by[7][8]

$$\varphi_i(v) = \sum_{i \in S \subseteq N} \frac{w(S)}{|S|}$$

where the sum is over all subsets S of N that include player i.

This can be interpreted as

$$\varphi_i(v) = \sum_{\text{coalitions including i}} \frac{\text{synergy of the coalition}}{\text{members in the coalition}}$$

In other words, the synergy of each coalition is divided equally between all members.

Examples

Business example

Consider a simplified description of a business. An owner, o, provides crucial capital in the sense that, without him/her, no gains can be obtained. There are m workers $w_1,...,w_m$, each of whom contributes an amount p to the total profit. Let

$$N = \{o, w_1, \ldots, w_m\}.$$

The value function for this coalitional game is

$$v(S) = \begin{cases} mp, & \text{if } o \in S \\ 0, & \text{otherwise} \end{cases}$$

where m is the cardinality of $S \setminus \{o\}$. Computing the Shapley value for this coalition game leads to a value of $\frac{mp}{2}$ for the owner and $\frac{p}{2}$ for each one of the m workers.

This can be understood from the perspective of synergy. The synergy function w is

$$w(S) = \begin{cases} p, & \text{if } S = \{o, w_i\} \\ 0, & \text{otherwise} \end{cases}$$

so the only coalitions that generate synergy are one-to-one between the owner and any individual worker.

Using the above formula for the Shapley value in terms of w we compute

$$\varphi_{w_i} = \frac{w(\{o, w_i\})}{2} = \frac{p}{2}$$

and

$$\varphi_o = \sum_{i=1}^{m} \frac{w(\{o, w_i\})}{2} = \frac{mp}{2}$$

Glove game

The glove game is a coalitional game where the players have left- and right-hand gloves and the goal is to form pairs. Let

$$N = \{1, 2, 3\},$$

where players 1 and 2 have right-hand gloves and player 3 has a left-hand glove.

The value function for this coalitional game is

$$v(S) = \begin{cases} 1 & \text{if } S \in \{\{1,3\}, \{2,3\}, \{1,2,3\}\}\,; \\ 0 & \text{otherwise.} \end{cases}$$

The formula for calculating the Shapley value is

$$\varphi_i(v) = \frac{1}{|N|!} \sum_R \left[v(P_i^R \cup \{i\}) - v(P_i^R) \right],$$

where R is an ordering of the players and P_i^R is the set of players in N which precede i in the order R.

The following table displays the marginal contributions of Player 1.

Order R	MC_1
$1,2,3$	$v(\{1\}) - v(\varnothing) = 0 - 0 = 0$
$1,3,2$	$v(\{1\}) - v(\varnothing) = 0 - 0 = 0$
$2,1,3$	$v(\{1,2\}) - v(\{2\}) = 0 - 0 = 0$
$2,3,1$	$v(\{1,2,3\}) - v(\{2,3\}) = 1 - 1 = 0$
$3,1,2$	$v(\{1,3\}) - v(\{3\}) = 1 - 0 = 1$
$3,2,1$	$v(\{1,3,2\}) - v(\{3,2\}) = 1 - 1 = 0$

Observe

$$\varphi_1(v) = \left(\frac{1}{6}\right)(1) = \frac{1}{6}.$$

By a symmetry argument it can be shown that

$$\varphi_2(v) = \varphi_1(v) = \frac{1}{6}.$$

Due to the efficiency axiom, the sum of all the Shapley values is equal to 1, which means that

$$\varphi_3(v) = \frac{4}{6} = \frac{2}{3}.$$

Properties

The Shapley value has many desirable properties.

Efficiency

The sum of the Shapley values of all agents equals the value of the grand coalition, so that all the gain is distributed among the agents:

$$\sum_{i \in N} \varphi_i(v) = v(N)$$

Proof:

$$\sum_{i \in N} \varphi_i(v) = \frac{1}{|N|!} \sum_R \sum_{i \in N} v(P_i^R \cup \{i\}) - v(P_i^R) = \frac{1}{|N|!} \sum_R v(N) = \frac{1}{|N|!} |N|! \cdot v(N) = v(N)$$

since $\sum_{i \in N} v(P_i^R \cup \{i\}) - v(P_i^R)$ is a telescoping sum and there are $|N|!$ different orderings R.

Symmetry

If i and j are two actors who are equivalent in the sense that

$$v(S \cup \{i\}) = v(S \cup \{j\})$$

for every subset S of N which contains neither i nor j, then $\varphi_i(v) = \varphi_j(v)$.

This property is also called *equal treatment of equals*.

Linearity

If two coalition games described by gain functions v and w are combined, then the distributed gains should correspond to the gains derived from v and the gains derived from w:

$$\varphi_i(v + w) = \varphi_i(v) + \varphi_i(w)$$

for every i in N. Also, for any real number a,

$$\varphi_i(av) = a\varphi_i(v)$$

for every i in N.

Null player

The Shapley value $\varphi_i(v)$ of a null player i in a game v is zero. A player i is *null* in v if $v(S \cup \{i\}) = v(S)$ for all coalitions S that do not contain i.

Given a player set N, the Shapley value is the only map from the set of all games to payoff vectors that satisfies *all four* properties: Efficiency, Symmetry, Linearity, Null player.

Stand-alone test

If v is a subadditive set function, i.e., $v(S \sqcup T) \leq v(S) + v(T)$, then for each agent i: $\varphi_i(v) \leq v(\{i\})$.

Similarly, if v is a superadditive set function, i.e., $v(S \sqcup T) \geq v(S) + v(T)$, then for each agent i: $\varphi_i(v) \geq v(\{i\})$.

So, if the cooperation has positive externalities, all agents (weakly) gain, and if it has negative externalities, all agents (weakly) lose.[9]: 147–156

Anonymity

If i and j are two agents, and w is a gain function that is identical to v except that the roles of i and j have been exchanged, then $\varphi_i(v) = \varphi_j(w)$. This means that the labeling of the agents doesn't play a role in the assignment of their gains.

Marginalism

The Shapley value can be defined as a function which uses only the marginal contributions of player i as the arguments.

Characterization

The Shapley value not only has desirable properties, it is also the *only* payment rule satisfying some subset of these properties. For example, it is the only payment rule satisfying the four properties of Efficiency, Symmetry, Linearity and Null player.[10] See[9]: 147–156 for more characterizations.

Aumann–Shapley value

In their 1974 book, Lloyd Shapley and Robert Aumann extended the concept of the Shapley value to infinite games (defined with respect to a non-atomic measure), creating the diagonal formula.[11] This was later extended by Jean-François Mertens and Abraham Neyman.

As seen above, the value of an n-person game associates to each player the expectation of his contribution to the worth or the coalition or players before him in a random ordering of all the players. When there are many players and each individual plays only a minor role, the set of all players preceding a given one is heuristically thought as a good sample of the players so that the value of a given infinitesimal player ds around as "his" contribution to the worth of a "perfect" sample of the population of all players.

Symbolically, if v is the coalitional worth function associating to each coalition c measured subset of a measurable set I that can be thought as $I = [0,1]$ without loss of generality.

$$(Sv)(ds) = \int_0^1 (\, v(tI + ds) - v(tI)\,)\, dt.$$

where $(Sv)(ds)$denotes the Shapley value of the infinitesimal player ds in the game, tI is a perfect sample of the all-player set I containing a proportion t of all the players, and $tI + ds$ is the coalition obtained after ds joins tI. This is the heuristic form of the diagonal formula.

Assuming some regularity of the worth function, for example assuming v can be represented as differentiable function of a non-atomic measure on I, μ, $v(c) = f(\mu(c))$ with density function φ, with $\mu(c) = \int 1_c(u)\varphi(u)\, du,$ ($1_c()$ the characteristic function of c). Under such conditions

$$\mu(tI) = t\mu(I),$$

as can be shown by approximating the density by a step function and keeping the proportion t for each level of the density function, and

$$v(tI + ds) = f(t\mu(I)) + f'(t\mu(I))\mu(ds).$$

The diagonal formula has then the form developed by Aumann and Shapley (1974)

$$(Sv)(ds) = \int_0^1 f'_{t\mu(I)}(\mu(ds))\, dt$$

Above μ can be vector valued (as long as the function is defined and differentiable on the range of μ, the above formula makes sense).

In the argument above if the measure contains atoms $\mu(tI) = t\mu(I)$ is no longer true—this is why the diagonal formula mostly applies to non-atomic games.

Two approaches were deployed to extend this diagonal formula when the function f is no longer differentiable. Mertens goes back to the original formula and takes the derivative after the integral thereby benefiting from the smoothing effect. Neyman took a different approach. Going back to an elementary application of Mertens's approach from Mertens (1980):[12]

$$(Sv)(ds) = \lim_{\varepsilon\to 0, \varepsilon>0} \frac{1}{\varepsilon} \int_0^{1-\varepsilon} (f(t + \varepsilon\mu(ds)) - f(t))\, dt$$

This works for example for majority games—while the original diagonal formula cannot be used directly. How Mertens further extends this by identifying symmetries that the Shapley value should be invariant upon, and averaging over such symmetries to create further smoothing effect commuting averages with the derivative operation as above.[13] A survey for non atomic value is found in Neyman (2002)[14]

Generalization to coalitions

The Shapley value only assigns values to the individual agents. It has been generalized[15] to apply to a group of agents C as,

$$\varphi_C(v) = \sum_{T\subseteq N\setminus C} \frac{(n - |T| - |C|)!\, |T|!}{(n - |C| + 1)!} \sum_{S\subseteq C} (-1)^{|C|-|S|} v(S \cup T)\ .$$

In terms of the synergy function w above, this reads[7][8]

$$\varphi_C(v) = \sum_{C\subseteq T\subseteq N} \frac{w(T)}{|T| - |C| + 1}$$

where the sum goes over all subsets T of N that contain C.

This formula suggests the interpretation that the Shapley value of a coalition is to be thought of as the standard Shapley value of a single player, if the coalition C is treated like a single player.

Value of a player to another player

The Shapley value $\varphi_i(v)$ was decomposed in[16] into a matrix of values

$$\varphi_{ij}(v) = \sum_{S\subseteq N} \frac{(|S| - 1)!\, (n - |S|)!}{n!} (v(S) - v(S \setminus \{i\}) - v(S \setminus \{j\}) + v(S \setminus \{i, j\})) \sum_{t=|S|}^{n} \frac{1}{t}$$

Each value $\varphi_{ij}(v)$ represents the value of player i to player j. This matrix satisfies

$$\varphi_i(v) = \sum_{j\in N} \varphi_{ij}(v)$$

i.e. the value of player i to the whole game is the sum of their value to all individual players.

In terms of the synergy w defined above, this reads

$$\varphi_{ij}(v) = \sum_{\{i,j\} \subseteq S \subseteq N} \frac{w(S)}{|S|^2}$$

where the sum goes over all subsets S of N that contain i and j.

This can be interpreted as sum over all subsets that contain players i and j, where for each subset S you

- take the synergy $w(S)$ of that subset
- divide it by the number of players in the subset $|S|$. Interpret that as the surplus value player i gains from this coalition
- further divide this by $|S|$ to get the part of player i's value that's attributed to player j

In other words, the synergy value of each coalition is evenly divided among all $|S|^2$ *pairs* (i, j) of players in that coalition, where i generates surplus for j.

In machine learning

The Shapley value provides a principled way to explain the predictions of nonlinear models common in the field of machine learning. By interpreting a model trained on a set of features as a value function on a coalition of players, Shapley values provide a natural way to compute which features contribute to a prediction.[17] This unifies several other methods including Locally Interpretable Model-Agnostic Explanations (LIME),[18] DeepLIFT,[19] and Layer-Wise Relevance Propagation.[20]

See also

- Airport problem
- Banzhaf power index
- Shapley–Shubik power index

References

1. Shapley, Lloyd S. (August 21, 1951). "Notes on the n-Person Game -- II: The Value of an n-Person Game" (https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf) (PDF). Santa Monica, Calif.: RAND Corporation.
2. Roth, Alvin E., ed. (1988). *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511528446 (https://doi.org/10.1017%2FCBO9780511528446). ISBN 0-521-36177-X.
3. Hart, Sergiu (1989). "Shapley Value". In Eatwell, J.; Milgate, M.; Newman, P. (eds.). *The New Palgrave: Game Theory*. Norton. pp. 210–216. doi:10.1007/978-1-349-20181-5_25 (https://doi.org/10.1007%2F978-1-349-20181-5_25). ISBN 978-0-333-49537-7.
4. Hart, Sergiu (May 12, 2016). "A Bibliography of Cooperative Games: Value Theory" (http://www.ma.huji.ac.il/~hart/value.html).
5. Phillips, Matthew; Marden, Jason R. (July 2018). "Design Tradeoffs in Concave Cost-Sharing Games". *IEEE Transactions on Automatic Control*. **63** (7): 2242–2247. doi:10.1109/tac.2017.2765299 (https://doi.org/10.1109%2Ftac.2017.2765299). ISSN 0018-9286 (https://www.worldcat.org/issn/0018-9286). S2CID 45923961 (https://api.semanticscholar.org/CorpusID:45923961).

6. For a proof of unique existence, see Ichiishi, Tatsuro (1983). *Game Theory for Economic Analysis* (https://books.google.com/books?id=zFm7AAAAIAAJ&pg=PA118). New York: Academic Press. pp. 118–120. ISBN 0-12-370180-5.
7. Grabisch, Michel (October 1997). "Alternative Representations of Discrete Fuzzy Measures for Decision Making". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. **5** (5): 587–607. doi:10.1142/S0218488597000440 (https://doi.org/10.1142%2FS0218488597000440). ISSN 0218-4885 (https://www.worldcat.org/issn/0218-4885).
8. Grabisch, Michel (1 December 1997). "k-order additive discrete fuzzy measures and their representation". *Fuzzy Sets and Systems*. **92** (2): 167–189. doi:10.1016/S0165-0114(97)00168-1 (https://doi.org/10.1016%2FS0165-0114%2897%2900168-1). ISSN 0165-0114 (https://www.worldcat.org/issn/0165-0114).
9. Herve Moulin (2004). *Fair Division and Collective Welfare*. Cambridge, Massachusetts: MIT Press. ISBN 9780262134231.
10. Shapley, Lloyd S. (1953). "A Value for n-person Games". In Kuhn, H. W.; Tucker, A. W. (eds.). *Contributions to the Theory of Games*. Annals of Mathematical Studies. Vol. 28. Princeton University Press. pp. 307–317. doi:10.1515/9781400881970-018 (https://doi.org/10.1515%2F9781400881970-018). ISBN 9781400881970.
11. Aumann, Robert J.; Shapley, Lloyd S. (1974). *Values of Non-Atomic Games*. Princeton: Princeton Univ. Press. ISBN 0-691-08103-4.
12. Mertens, Jean-François (1980). "Values and Derivatives". *Mathematics of Operations Research*. **5** (4): 523–552. doi:10.1287/moor.5.4.523 (https://doi.org/10.1287%2Fmoor.5.4.523). JSTOR 3689325 (https://www.jstor.org/stable/3689325).
13. Mertens, Jean-François (1988). "The Shapley Value in the Non Differentiable Case". *International Journal of Game Theory*. **17** (1): 1–65. doi:10.1007/BF01240834 (https://doi.org/10.1007%2FBF01240834). S2CID 118017018 (https://api.semanticscholar.org/CorpusID:118017018).
14. Neyman, A., 2002. Value of Games with infinitely many Players, "Handbook of Game Theory with Economic Applications," Handbook of Game Theory with Economic Applications, Elsevier, edition 1, volume 3, number 3, 00. R.J. Aumann & S. Hart (ed.).[1] (http://ratio.huji.ac.il/dp/neyman/values.pdf)
15. Grabisch, Michel; Roubens, Marc (1999). "An axiomatic approach to the concept of interaction among players in cooperative games". *International Journal of Game Theory*. **28** (4): 547-565. doi:10.1007/s001820050125 (https://doi.org/10.1007%2Fs001820050125). S2CID 18033890 (https://api.semanticscholar.org/CorpusID:18033890).
16. Hausken, Kjell; Mohr, Matthias (2001). "The Value of a Player in n-Person Games" (http://www.jstor.org/stable/41060209). *Social Choice and Welfare*. **18** (3): 465–83. doi:10.1007/s003550000070 (https://doi.org/10.1007%2Fs003550000070). JSTOR 41060209 (https://www.jstor.org/stable/41060209). S2CID 27089088 (https://api.semanticscholar.org/CorpusID:27089088).
17. Lundberg, Scott M.; Lee, Su-In (2017). "A Unified Approach to Interpreting Model Predictions" (https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html). *Advances in Neural Information Processing Systems*. **30**: 4765–4774. arXiv:1705.07874 (https://arxiv.org/abs/1705.07874). Retrieved 2021-01-30.
18. Ribeiro, Marco Tulio; Singh, Sameer; Guestrin, Carlos (2016-08-13). *"Why Should I Trust You?"*. New York, NY, USA: ACM. doi:10.1145/2939672.2939778 (https://doi.org/10.1145%2F2939672.2939778). ISBN 978-1-4503-4232-2.
19. Shrikumar, Avanti; Greenside, Peyton; Kundaje, Anshul (2017-07-17). "Learning Important Features Through Propagating Activation Differences" (http://proceedings.mlr.press/v70/shrikumar17a.html). *PMLR*. pp. 3145–3153. ISSN 2640-3498 (https://www.worldcat.org/issn/2640-3498). Retrieved 2021-01-30.

20. Bach, Sebastian; Binder, Alexander; Montavon, Grégoire; Klauschen, Frederick; Müller, Klaus-Robert; Samek, Wojciech (2015-07-10). Suarez, Oscar Deniz (ed.). "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498753). *PLOS ONE*. Public Library of Science (PLoS). **10** (7): e0130140. Bibcode:2015PLoSO..1030140B (https://ui.adsabs.harvard.edu/abs/2015PLoSO..1030140B). doi:10.1371/journal.pone.0130140 (https://doi.org/10.1371%2Fjournal.pone.0130140). ISSN 1932-6203 (https://www.worldcat.org/issn/1932-6203). PMC 4498753 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498753). PMID 26161953 (https://pubmed.ncbi.nlm.nih.gov/26161953).

Further reading

- Friedman, James W. (1986). *Game Theory with Applications to Economics* (https://archive.org/details/gametheorywithap0000frie). New York: Oxford University Press. pp. 209 (https://archive.org/details/gametheorywithap0000frie/page/209)–215. ISBN 0-19-503660-3.

External links

- "Shapley value" (https://www.encyclopediaofmath.org/index.php?title=Shapley_value), *Encyclopedia of Mathematics*, EMS Press, 2001 [1994]
- Shapley Value Calculator (http://shapleyvalue.com/)
- Calculating a Taxi Fare using the Shapley Value (https://www.youtube.com/watch?v=aThG4YAFErw)