Understanding Shapley Values

@AIinMinutes

2025

Presentation Covers:

- Cooperative Games & Coalitions
- Fairness Axioms
- Shapley Value Formulations
- Three-Player Example
- Linear Regression Application

#1 Cooperative Games & Coalitions

Foundation of Shapley values

Cooperative Games:

- Players work together toward shared goals
- Team success/failure affects all players

Coalitions:

- Players: $N = \{1, 2, ..., n\}$
- Coalition (S): Any subset $S \subseteq N$
- Grand Coalition: All players (N)
- Number of possible coalitions: 2^n

Characteristic Function:

- $v : 2^N \to \mathbb{R}$ maps coalitions to payoffs
- $v(S) =$ total payoff coalition S can achieve

@AIinMinutes

#2 Fairness Axioms I

Mathematical principles of fair distribution

1. **Symmetry:** Equal contributors receive equal payoffs

$$\text{If } v(S \cup \{i\}) = v(S \cup \{j\}) \text{ for all } S \subseteq N \setminus \{i, j\}$$
$$\implies \phi_i(v) = \phi_j(v)$$

2. **Null Player:** Players adding no value receive nothing

$$\text{If } v(S \cup \{i\}) = v(S) \text{ for all } S \subseteq N \setminus \{i\}$$
$$\implies \phi_i(v) = 0$$

@AIinMinutes

#3 Fairness Axioms II

Mathematical principles of fair distribution

3. Efficiency: The entire value is fully distributed

$$\sum_{i=1}^{n} \phi_i(v) = v(N)$$

4. Additivity: Payoffs are additive across different games

For any two characteristic functions v and w:

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

@AIinMinutes

#4 Shapley Value Definitions

Two equivalent mathematical formulations

Permutation Form:

$$\phi_i(v) = \frac{1}{N!} \sum_{\pi \in \Pi} [v(P_i^\pi \cup \{i\}) - v(P_i^\pi)]$$

- Π: Set of all $N!$ permutations of players
- P_i^π: Players preceding i in ordering π

Combinatorial Form:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)]$$

- $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ = Probability weight
- $[v(S \cup \{i\}) - v(S)]$ = Marginal contribution

@AIinMinutes

#5 From Ordered to Unordered Sets

Converting permutations to combinations

Key Insights:

- Players after i don't affect i's marginal contribution
- For same unordered set preceding i, contribution is identical
- For coalition S before i (excluding i):

  $|S|!$ ways to arrange players before i

  $(|N| - |S| - 1)!$ ways to arrange players after i
- Conversion factor from permutations to combinations:

  $\frac{|S|!(|N|-|S|-1)!}{|N|!}$

@AIinMinutes

#6 Three-Player Example: Setup

Simple example illustrating Shapley value calculation

Game Parameters:

- Players: A, B, C ($N = 3$)
- Characteristic Function: $v(S) = |S|$

  (Coalition value = number of players in it)

Coalition Values:

$$\begin{aligned} v(\emptyset) &= 0 \\ v(\{A\}) = v(\{B\}) = v(\{C\}) &= 1 \\ v(\{A, B\}) = v(\{A, C\}) = v(\{B, C\}) &= 2 \\ v(\{A, B, C\}) &= 3 \end{aligned}$$

@AIinMinutes

#7 Example: Combinatorial Calculation

Computing Shapley value for Player A

For each subset $S \subseteq N \setminus \{A\}$:

1. $S = \emptyset$:

   Weight: $\frac{0!(3-0-1)!}{3!} = \frac{2!}{6} = \frac{1}{3}$

   Marginal: $v(\{A\}) - v(\emptyset) = 1 - 0 = 1$

   Term: $\frac{1}{3} \cdot 1 = \frac{1}{3}$

2. $S = \{B\}$ and $S = \{C\}$:

   Weight: $\frac{1!(3-1-1)!}{3!} = \frac{1}{6}$ (each)

   Marginal: $v(S \cup \{A\}) - v(S) = 2 - 1 = 1$

   Terms: $\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 = \frac{1}{3}$

3. $S = \{B, C\}$:

   Weight: $\frac{2!(3-2-1)!}{3!} = \frac{1}{3}$

   Marginal: $v(\{A, B, C\}) - v(\{B, C\}) = 3 - 2 = 1$

   Term: $\frac{1}{3} \cdot 1 = \frac{1}{3}$

Result: $\phi_A = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$

@AIinMinutes

#8 Linear Regression Application

Shapley values for model explanation

Multiple Linear Regression Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$$

Intercept Relationship:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \hat{\beta}_3 \bar{x}_3$$

Prediction Difference from Mean:

$$\hat{y}_i - \bar{y} = \hat{\beta}_1 (x_{1i} - \bar{x}_1) + \hat{\beta}_2 (x_{2i} - \bar{x}_2) + \hat{\beta}_3 (x_{3i} - \bar{x}_3)$$

#9 SHAP Values for Features

Feature importance as Shapley values

Feature Shapley Value:

$$\mathrm{SHAP}_j(i) = \hat{\beta}_j(x_{ji} - \bar{x}_j)$$

Properties:

- Each feature's marginal contribution (in all ordered sets) is constant

- Contribution depends on deviation from feature's mean

- Total prediction difference equals sum of SHAP values:

$$\hat{y}_i - \bar{y} = \textstyle\sum_{j=1}^{m} \mathrm{SHAP}_j(i)$$

@AIinMinutes