Minimax - An Application of MWU Method

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Outline

Zero-Sum Games

MWU Method

Proof

Zero-Sum Games

2-Player Zero Sum Game Example

Matching Pennies

2-Players: Row and Column. Toss coin simultaneously. If the outcome on both the coins is the same, Row player wins $1 from Column player, otherwise loses a dollar to the column player.

Payoff matrix of the row player :

	Head	Tail
Head	$+1$	-1
Tail	-1	$+1$

Payoff matrix of the column player:

	Head	Tail
Head	-1	$+1$
Tail	$+1$	-1

Zero-Sum: Loss of one player = Gain of the other player.

Matching Pennies Game

Payoff matrix of the row player :

	Head	Tail
Head	$+1$	-1
Tail	-1	$+1$

Pure Strategies: In each round of the game, Row player decides to always play Head.

Mixed Strategy: Row player decides to play Heads/Tails based on some probability distribution.

Think of the following questions:

1. What will be the revenue of row player if it employs the pure strategy of always playing Heads?
2. What will be the best strategy for the column player if the row player plays the mixed strategy of playing Heads and Tails with equal probability? What will be its expected revenue?
3. What will be the best strategy and the expected revenue of column player if row player plays Heads with probability 0.7 and Tails with probability 0.3.

Matching Pennies Game

Payoff matrix of the row player :

	Head	Tail
Head	$+1$	-1
Tail	-1	$+1$

4. Is the best strategy for both the players is to choose heads and tails with equal probability? What is the expected payoff?
   Evaluate $\sum_{i=1}^{2}\sum_{j=1}^{2} p_i q_j A[i,j]$ and show that it equals to 0, where $p_1 = p_2 = \frac{1}{2}$ and $q_1 = q_2 = \frac{1}{2}$.
5. What is the expected payoff of row player if it chooses each row with probability $\frac{1}{2}$, and column player can choose any possible mixed strategy?

Minimax for Matching Pennies Game

Row player goes first

Assume that the row player plays first and chooses the rows 1 and 2 with probabilities p_1 and $p_2 = 1 - p_1$. Assume that it announces its mixed strategy vector $p = (p_1, p_2)$. The following holds:

1. The optimum payoff of the column player is at least $\max_p \left(\min_q p^T A q \right)$. First let the row player choose a strategy p, and then the column player minimizes over the various choices for q.
2. The best strategy for the column player is to deterministically play one of the columns - the column that minimizes the value $p^T A q$, where $q = (1, 0)$ or $q = (0, 1)$.

Column player goes first

If the column player plays first and the row player plays second, then the optimum payoff for the row player is at least $\min_q \left(\max_p p^T A q \right)$.

Minimax for Matching Pennies Game

Minmax for matching pennies

For the matching pennies game

$$\max_p \left(\min_q p^T A q \right) = \min_q \left(\max_p p^T A q \right) = 0$$

Counterintutive: Though it seems that a player who plays first has a disadvantage, the above statement shows that it doesn't matter who goes first. The value of the game is the same.

2-player zero sum game example

Rock-Paper-Scissors

2-Players: Row and Column. Each plays one of the three possibilities (Rock, Paper, Scissors) simultaneously. Rock beats Scissors, Scissors beats Paper, and Paper beats Rock. Outcome of the game is either a draw (when both the players choose the same), or a gain of $1 for one player and loss to the other depending on the their choice.

Payoff matrix of the row player:

	Rock	Paper	Scissors
Rock	0	-1	$+1$
Paper	$+1$	0	-1
Scissors	-1	$+1$	0

2-player zero sum game example

Consider the following payoff matrix for row player:

	Head	Tail
Head	$+3$	-1
Tail	-2	1

Row player plays first with mixed strategy $(p, 1-p)$.
Column player plays $\min(3p - 2(1-p), -p + (1-p))$.
Knowing this, the row player should play $\max_p \min(3p - 2(1-p), -p + (1-p))$
$\implies$ $\max$ is achieved when $3p - 2(1-p) = -p + (1-p)$, i.e.,
$p = 3/7 \in [0, 1]$. The value of the game is $-p + (1-p) = 1/7$.

Column player plays first with mixed strategy $(q, 1-q)$.
Row player plays $\max(3q - (1-q), -2q + (1-q))$.
$\implies$ column player should play $\min_q \max(3q - (1-q), -2q + (1-q))$.
$\min$ is achieved when $3q - (1-q) = -2q + (1-q)$, or, $q = 2/7 \in [0, 1]$. Value of the game $= 1 - 3q = 1/7 =$ same value when the row player played first!

Conclusion: Row and column players mixed strategies $(3/7, 4/7)$ and $(2/7, 5/7)$, respectively, yield the same (optimal) value of the game ($= 1/7$), irrespective of who plays first.

Generic 2-player zero sum games

2-player zero sum games

Row player plays one of the possible n strategies and the column player plays one of the m strategies. The payoff of the row player is specified by an $n \times m$ matrix A, where A_{ij} is the payoff/reward of the row player if it plays strategy i and the column player plays strategy j. In the zero-sum games, the payoff of the column player is $-A_{ij}$.

For the mixed strategies $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_m)$ of the row and column players, respectively, the payoff of the row player is $\sum_{j=1}^{m} \sum_{i=1}^{n} p_i q_j A[i,j] = p^T A q$, and the payoff to the column player is $-p^T A q$.

Minimax for generic 2-player zero sum games

Row player plays first

Suppose the row player commits to the mixed strategy p first. Best strategy for the column player is to optimize the function $\min_q p^T Aq$.

Column player plays first

Suppose the column player commits to the mixed strategy q first. Best strategy for the row player is to optimize the function $\max_p p^T Aq$.

Minimax theorem

In two-player zero sum games, if both players play rationally than $\max_p \left(\min_q p^T Aq \right) = \min_q \left(\max_p p^T Aq \right)$. This quantity is the **value** of the game.

Remark: The value of the game remains the same irrespective of which player plays first provided they use optimal mixed strategies.

Minimax for generic 2-player zero sum games

Row player plays first

Suppose the row player commits to the mixed strategy p first. Best strategy for the column player is to optimize the function $\min_q p^T Aq$.

Column player plays first

Suppose the column player commits to the mixed strategy q first. Best strategy for the row player is to optimize the function $\max_p p^T Aq$.

Minimax theorem

In two-player zero sum games, if both players play rationally than $\max_p \left(\min_q p^T Aq \right) = \min_q \left(\max_p p^T Aq \right)$. This quantity is the **value** of the game.

Remark: The value of the game remains the same irrespective of which player plays first provided they use optimal mixed strategies.

The interpretation is as follows. Column player wants to ensure that the row player gets the smallest possible reward once row player fixes its strategy p. Row player wants to choose that p which achieves $\max_p \left(\min_q p^T Aq \right)$.

Now consider the scenario when the column player chooses the mixed strategy q first. Now the row player, using a similar reasoning, will like to optimize $\max_p p^T Aq$. Column player wants to choose that q which achieves $\min_q \left(\max_p p^T Aq \right)$.

MWU Method

MWU with costs in $[-1, 1]$

Set of experts $E = \{1, \ldots, n\}$.
Let η be any real number in $[0, \frac{1}{2}]$
For each expert i, set its initial weight $w_i^1 = 1$

For each day $t := 1$ to T do:

Step 1: Define $\Phi^t = \sum_{i=1}^{n} w_i^t$

For each expert i, compute $p_i^t = \frac{w_i^t}{\Phi^t}$

Step 2: Choose experts based on their probabilities and follow their advise for day t.

Step 3: Update Weights: For day $t + 1$, for each expert i, set its weight $w_i^{t+1} = w_i^t(1 - \eta m_i^t)$.

Results on MWU with costs in $[-1, 1]$

Cost of MWU

The cost of MWU algorithm is off by an additive factor that is proportional to the square root of the product of the number of days and the number of experts as compared to the best expert. I.e., by setting $\eta = \sqrt{\frac{\ln n}{T}}$ in $\sum_{t=1}^{T} M^t \leq \frac{\ln n}{\eta} + \eta T + \sum_{t=1}^{T} m_i^t$, we obtain $\sum_{t=1}^{T} M^t \leq 2\sqrt{T \ln n} + \sum_{t=1}^{T} m_i^t$. Recall that M^t is the expected loss that the algorithm incurs on day t and is given by $M^t = \sum_{i=1}^{n} p_i^t m_i^t$, where $p^t = (p_1^t, p_2^t, \ldots, p_n^t)$ and $m^t = (m_1^t, m_2^t, \ldots, m_n^t)$.

Average Error: Consider the average error on each day (divide by T):

$$\frac{1}{T} \sum_{t=1}^{T} M^t \leq 2\sqrt{\frac{\ln n}{T}} + \frac{1}{T} \sum_{t=1}^{T} m_i^t$$

Observe that as T increases the average error drops down.
Thus, the MWU method is able to learn from the experts reasonably well when executed over a number of days.

Proof

Proof of minimax theorem using MWU

Let A be the zero-sum $n \times m$ game matrix, where each $A[i, j] \in [-1, 1]$, and assume $n \geq m$; otherwise work with A^T.

Assume row player goes first. Consider the following adaptation of the MWU method:

For each time step $t = 1, \ldots, T = \frac{4 \log n}{\epsilon^2}$ do:

1. Consider each row strategy as an expert. Row player chooses a mixed strategy p^t as in the MWU method.
2. Given p^t, the column player plays the column that gives the best expected value of the payoff. Let q^t be the play of the column player for a given p^t.
3. If column j is chosen, $\forall i \in \{1, \ldots, n\}$, set $m_i^t = A_{ij}$, and apply the MWU update rule for the experts corresponding to each row ($w_i^{t+1} = w_i^t(1 - \eta m_i^t)$).

Analysis for column player

Column player's expected reward

Time averaged (negative) expected reward for the column player is at most $\max_p \left(\min_q p^T A q \right)$.

Proof: Let $\hat{p} = \frac{1}{T} \sum_{t=1}^{T} p^t$. Let q^* be the optimal response of column player to $\hat{p}$.

$$\begin{aligned} \max_p \left(\min_q p^T A q \right) &\geq \min_q \hat{p}^T A q \\ &= \hat{p}^T A q^* \\ &= \frac{1}{T} \sum_{t=1}^{T} (p^t)^T A q^* \\ &\geq \frac{1}{T} \sum_{t=1}^{T} (p^t)^T A q^t \qquad \square \end{aligned}$$

Note: q^t is the best response by column player to p^t on day t.

$\frac{1}{T} \sum_{t=1}^{T} (p^t)^T A q^t$ is the column player's reward returned by the MWU method.

Analysis for row player

Row player's expected reward

Time averaged expected reward for the row player is at least $\min_q \left(\max_p p^T A q \right) - \epsilon$.

Proof: Let $\hat{q} = \frac{1}{T} \sum_{t=1}^{T} q^t$.

MWU method gurantees that the time-averaged expected reward of the row player is within ϵ of the best it can obtain using any fixed (mixed) strategy.

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} (p^t)^T A q^t &\geq \max_p \left(\frac{1}{T} \sum_{t=1}^{T} p^T A q^t \right) - \epsilon \\
&= \max_p p^T A \left(\frac{1}{T} \sum_{t=1}^{T} q^t \right) - \epsilon \\
&= \max_p p^T A \hat{q} - \epsilon \\
&\geq \min_q \left(\max_p p^T A q \right) - \epsilon \qquad \square
\end{aligned}$$

Easy direction of minimax

2nd Player Advantage

The player who plays second can't perform worse than the first player in a zero-sum game. Since row player plays first, we have $\max_p \left(\min_q p^T A q \right) \leq \min_q \left(\max_p p^T A q \right)$.

Proof: Row player, knowing that if it chooses mixed strategy p, the column will choose that strategy q that minimizes the row players reward.

Thus, the row player finds the best p, say p^*, that maximizes its worst-case reward, i.e.,

$$\max_p \left(\min_q p^T A q \right) = \min_q (p^*)^T A q = V_R^* \tag{1}$$

If the row player chooses p^*, let the column's player best response be $\hat{q}$.

Question: If the column player plays $\hat{q}$, is the best response of row player p^*?

Easy direction of minimax (contd.)

Similarly, column player chooses best q, say q^*, that minimizes row players payoff, i.e.,

$$\min_q \left(\max_p p^T A q \right) = \max_p p^T A q^* = V_C^* \tag{2}$$

We want to show that $V_R^* \leq V_C^* \implies \max_p \left(\min_q p^T A q \right) \leq \min_q \left(\max_p p^T A q \right)$.

What happens if row and column players play p^* and q^*, respectively?

The reward of row player is at least V_R^* by Equation 1.

The reward of row player is at most V_C^* by Equation 2.

Thus, $V_R^* \leq V_C^*$ $\square$

Minimax from MWU

Minimax Theorem

From column and row player's expected reward after $T = \frac{4 \ln n}{\epsilon^2}$ days using the MWU adaptation, we have

$$\min_q \left(\max_p p^T A q \right) - \epsilon \leq \frac{1}{T} \sum_{t=1}^{T} (p^t)^T A q^t \leq \max_p \left(\min_q p^T A q \right).$$

By setting $\epsilon \to 0$, and using the 2nd players advantage, we obtain

$$\min_q \left(\max_p p^T A q \right) = \max_p \left(\min_q p^T A q \right).$$

Remarks

1. The pair (p^*, q^*) discussed in the 'Easy direction' ensures that $V_R^* = V_C^* =$ Value of the game. This pair forms the Nash equilibrium, i.e., none of the players will get a better reward by changing their mixed strategy when the other players strategy remains fixed.
2. MWU method assumed that the rewards of row and column players are in $[-1, 1]$. We can extend this easily to rewards in the range $[-\rho, \rho]$ at the cost of running the MWU method for $T = \frac{4\rho^2 \ln n}{\epsilon^2}$ days and modifying the update weight function to $w_i^{t+1} = w_i^t \left(1 - \eta \frac{m_i^t}{\rho}\right)$.

References

For Minmax Theorem, refer to books in game theory and linear programming duality.

For applications of MWU, see the research article by
Sanjeev Arora, Elad Hazan, and Satyen Kale: The Multiplicative Weights Update Method: a Meta Algorithm and Applications, Theory of Computing 8 (6): 121-164, 2012.