Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
ABCDEF		2400609

Shortest Path Algorithms: Taxonomy and Advance in Research

my summary

Contents

1 Introduction

test

1.1 Overview

test

1.2 Restatement of the Problem

First Problem Momentum is “strength or force gained by motion or by a series of events.”It directly shows the player’s current performance. To assess the players’ performance, it is crucial to have a clear understanding of “momentum.” We will focus on the following tasks:

- determine the influencing factors of “momentum”
- Quantify the variations in “momentum” .
- Visualize the process of “momentum” changes.

Second Problem “momentum’s role in the match” means the level of momentum affects the future scores of the match. The coach may subscribe to the idea that each point is an independent event and governed by probability. In this view, consecutive success and momentum changes (swings) are seen as more random than influenced by previous events. To judge this autocorrelation and to use our model, we

- perform autocorrelation test on momentum
- perform correlation test between current momentum and future scores.

1.3 Assumptions

To simplify the problem, we made the following assumptions:

- **Assumption 1:** The `px_unf_err` column of the data only counts those unforced errors that occurred when the player was hitting in baseline.
  Justification: Usually when a player is at net, the point will end in a few strikes, and there’s little probability that the player will hit an unforced error within that few strikes. What’s more, the `px_net_point` and `px_net_point_won` columns of the data can predominantly reflect the player’s ability at net, therefore reducing the impact of counting the unforced errors while at net.
- **Assumption 2:** The current performance on a certain aspect of a player can be reflected by the player’s 3 latest shots of that aspect.
  E.g. P_{ace} can be reflected by the proportion of aces in the 3 latest **serves** of the player, P_{win} can be reflected by the proportion of winners in the 3 latest **shots** of the player, rd can be reflected by the return depth of the 3 latest **returns** of the player, etc.
  Justification: The current performance of a player consists of the average performance and

the status of the player at the moment, which can be comprehensively reflected in the player's performance on recent shots. For convenience, we specified that the 3 latest shots can reflect the player's current performance.

1.4 Our Work

- develop a model to

2 Momentum Evaluation Model

2.1 Model Overview

To determine which player is performing better at a specific time, we create a indicator "Momentum" using the Analytic Hierarchy Process (AHP) to give a quantitative and overall evaluation.

To investigate the reasons behind "momentum", we first need to provide a preliminary definition for "momentum". The magnitude of "momentum" is defined as

$$f_{ijk} = \boldsymbol{\omega} \cdot \boldsymbol{x_{ijk}}$$

where:

1. f_{ijk} represents the "momentum" of player k before the jth point number in the ith match (in the order given by the table).

2. $\boldsymbol{x_{ijk}}$ is an n-dimensional column vector representing some influencing factors at the corresponding moment. Specific details will be provided later.

3. $\boldsymbol{\omega}$ is an n-dimensional row vector indicating the specific weights of the influencing factors, which will be obtained through the Analytic Hierarchy Process (AHP).

4. In this formula, there are two different calculation methods, one representing rounds where the player serves and the other representing rounds where the opponent serves. We can express it as

$$\boldsymbol{\omega} = \boldsymbol{\omega_0} \circ \boldsymbol{\delta} = (\omega_0^{(0)}\delta^{(0)}, \omega_0^{(1)}\delta^{(1)}, \ldots, \omega_0^{(n)}\delta^{(n)})$$

representing a vector formed by element-wise multiplication of two vectors of the same dimension. Here, $\boldsymbol{\delta}$ is a $0, 1$ vector indicating whether it is the player's serving round. In the specific calculation, we will consider two cases separately.

Out of our own interpretation of AHP, we will break down the problems into three parts:

1. Factor Normalization and Data Cleaning

2. Collinearity Detection

3. Analytic Hierarchy Process (AHP)

2.1.1 Notations

Symbols	Description
$player$	the current player we are considering (e.g. while calculating momentum)
$point_i$	the i^{th} point of the match, a vector consists of fields stated in the given dictionary
cur	the current index of the point, i.e. the match is currently at the cur^{th} point
H_i	denotes the set $\{point_{cur}, point_{cur-1}, \ldots, point_{cur-i+1}\}$
S_i	the set of latest i points where $player$ serves
R_i	the set of latest i points where $player$ returns
P_{ace}	current probability of hitting an ace by $player$
P_{df}	current probability of double-faulting by $player$
P_{1st}	current first serve goal rate by $player$
P_{fw}	current probability of $player$ winning a served point within 3 rallies
rd	current return depth of $player$
P_{win}	current probability of hitting a winner by $player$
P_{net}	current net win rate of $player$
$dist$	$player$'s running distance on the point
P_{unf}	current probability of hitting an unforced error by $player$
$scored$	whether $player$ scored the current point
$diff$	the score diffrence for $player$ in the current game (by number of points)
M	the current momentum of $player$ after a point

To access a certain field in a point, we simply use the field name stated in the given dictionary as index, i.e. for a point $point$, we use $point_{ace}$ to denote the binary variable that shows whether $player$ hits an ace ball in the point.

2.1.2 Factor Normalization and Data Cleaning

For the specific definition of $\boldsymbol{x_{ij}^n}$, we believe that, in addition to whether the player is serving, many other factors can have an impact, including the player's skills, fatigue level, and real-time mental state of the games. Based on these three main aspects, we have organized 12 factors as preliminary influencing factors, as follows:

$$P_{ace} = \frac{\sum_{p \in S_3} p_{ace}}{3} \tag{1}$$

$$P_{df} = -\frac{\sum_{p \in S_3} p_{double\_fault}}{3} \tag{2}$$

$$P_{1st} = \frac{\sum_{p \in S_3} [p_{serve\_no} = 1]}{3} \tag{3}$$

$$P_{fw} = \frac{\sum_{p \in S_3} [p_{rally\_count} \le 3][p_{point\_victor} = player]}{3} \tag{4}$$

$$rd = \frac{\sum_{p \in R_3} \begin{cases} 0, & p_{return\_depth} = ND \\ 1, & p_{return\_depth} = D \\ -1, & p_{return\_depth} = NA \end{cases}}{3} \tag{5}$$

$$P_{win} = \frac{\sum_{p \in H_3} p_{winner}}{3} \tag{6}$$

$$P_{net} = \frac{\sum_{p \in H_3} p_{net\_pt\_won}}{\sum_{p \in H_3} p_{net\_pt}} \tag{7}$$

$$dist = \begin{cases} 0, & point_{cur,distance\_run} < 5 \\ -1, & point_{cur,distance\_run} > 45 \\ \frac{5 - point_{cur,distance\_run}}{40}, & otherwise \end{cases} \tag{8}$$

$$P_{unf} = -\frac{\sum_{p \in H_3} p_{unf\_err}}{3} \tag{9}$$

$$scored = [point_{cur,point\_victor} = player] \tag{10}$$

$$diff = \frac{\sum_{p \in point} [p_{set\_no} = point_{cur,set\_no}][p_{game\_no} = point_{cur,game\_no}](2[p_{point\_victor} = player] - 1)}{\min\{3, \sum_{p \in point} [p_{set\_no} = point_{cur,set\_no}][p_{game\_no} = point_{cur,game\_no}]\}} \tag{11}$$

In order to normalize the data processed, we convert the original data to limit them in $[-1, 1]$. For those factors that negatively influence the momentum, such as P_{df}, we made sure it's in $[-1, 0]$. For those factors that positively influence the momentum, such as P_{win}, we made sure it's in $[0, 1]$. For those factors that influence the momentum in both ways, such as $diff$, we made sure it's in $[-1, 1]$.

2.1.3 Collinearity Detection

After processing the data, considering the potential collinearity among factors within the same category, such as serving aces, first-serve scoring rate, and whether the previous point was scored may be correlated, as well as running distance and the number of strokes possibly being related, we conducted collinearity detection using Stata. The results of the detection indicate a significant variance inflation factor between running distance and the number of strokes. Therefore, we decided to exclude one of them, choosing to retain the remaining 11 variables for the Analytic Hierarchy Process (AHP).

2.1.4 Analytic Hierarchy Process

We have previously decomposed the included factors from top to bottom into several levels, where factors within the same level are subordinate to factors in the level above or influence factors in the level above. They also dominate factors in the next level or are influenced by factors in the next level. Starting from the second level of the hierarchy, we construct comparison matrices for each factor influencing the factor in the level above, until reaching the bottom level. Each element in the matrix indicates the preference level between factor i and factor j at the same level. It is essential to note that we have separately established a series of such comparison matrices for two different serving types (serving by oneself and serving by the opponent). Here, we illustrate the matrix using serving by oneself as an example:

influe	ability	degre	manta
ability	1	1	1/3
degre	1/1	1	1
manta	3	1/1	1

ability	serve_	winne	net_wi
serve_	1	5	7
winne	1/5	1	3
net_wi	1/7	1/3	1

Figure 1: Comparison matrix for influencing factors and ability

degre	distan	unforc
distan	1	3
unforc	1/3	1

manta	scored	score_
scored	1	1/5
score_	5	1

Figure 2: Comparison matrix for degree of fatigue and mantality

serve_	ace	doubl	first_s	fast_w
ace	1	1	1/3	1/3
doubl	1/1	1	1/3	1/3
first_s	3	3	1	1/3
fast_w	3	3	3	1

Figure 3: Comparison matrix for serving

We obtain the weights for each component by calculating the maximum eigenvalue and normalizing its corresponding eigenvector. Certainly, for each matrix, we first need to test consistency using the Consistency Ratio (CR), where $CR = \frac{CI}{RI}$, $CI = \frac{\lambda_{max}-n}{n-1}$, $RI = 0.0, 0.58, 0.9$ (for matrices of size $2, 3, 4$). The computed Consistency Ratios for the matrices are 0.076, 0.037, 0.0, 0.0, 0.046. Since they are all less than 0.1, it confirms the consistency of the matrices.

Therefore, the weights for our model are as follows:

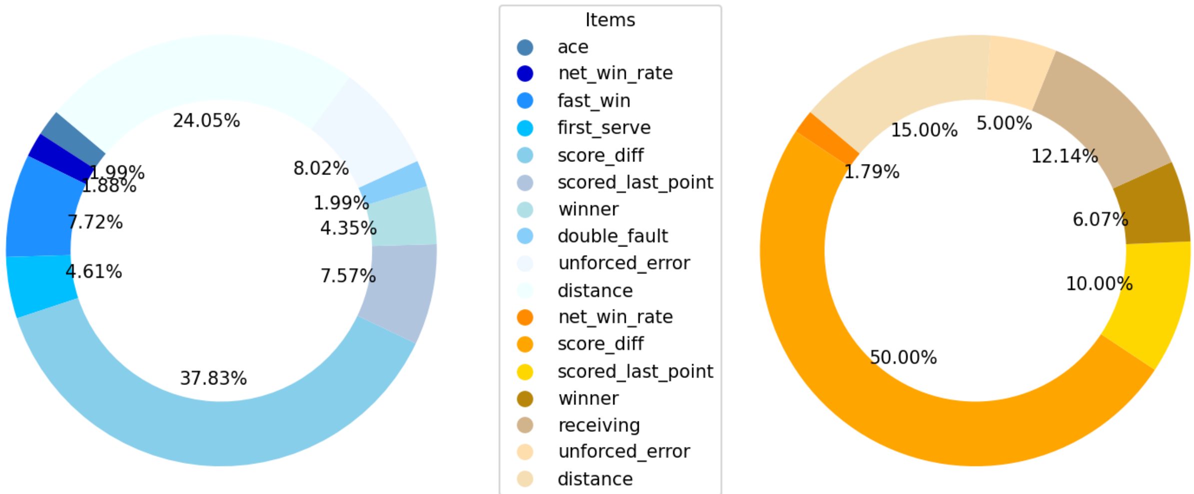


Figure 4: Weights in two different situations

Analyzing the various factors in the chart, it is evident that the most impactful factor is whether the previous point was scored. Following closely is the distance covered during the play, which aligns well with common intuition.

Thus, our final momentum is defined as:(need to change symbol)

$$momentum = \begin{cases} \sum\limits_{n=1, n\neq 5}^{11} \omega_n x_n, & \text{if the player serves} \\ \sum\limits_{n=5}^{11} \omega_n x_n, & \text{if the opponent serves} \end{cases}$$

Where $\omega_n (n = 1, \ldots, 11)$ represent the weight of the factors, which is listed in Figure 4.

2.2 Visualization and Analysis

Now, we illustrate the graph of the ”momentum” in the first match:

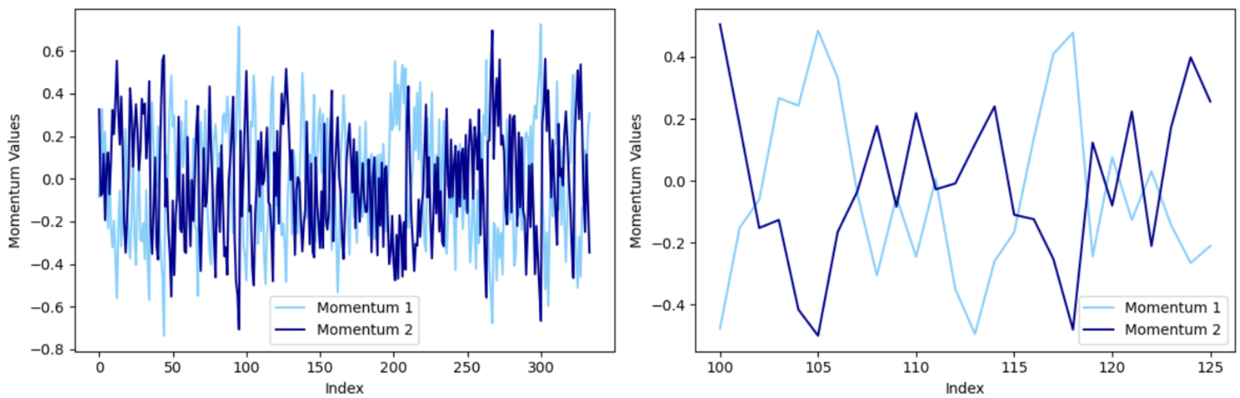


Figure 5: Momentum change in the first competition(global and local)

It can be observed that the variation in “momentum” is a process of give and take. In photo 2, when momentum 1 is above momentum 2, it means that the player 1 is performing better than the opponent.

2.3 momentum autocorrelation and correlation with runs of success

To answer the coach’s doubt, we need to perform autocorrelation test on momentum, and perform correlation test between current momentum and future scores in this section.

If the momentum has a high autocorrelation, it means that the momentum at this moment has a high impact on future performance. And if the correlation between momentum and future scores is high, it means that the player with higher momentum has a higher chance to win the next multiple round.

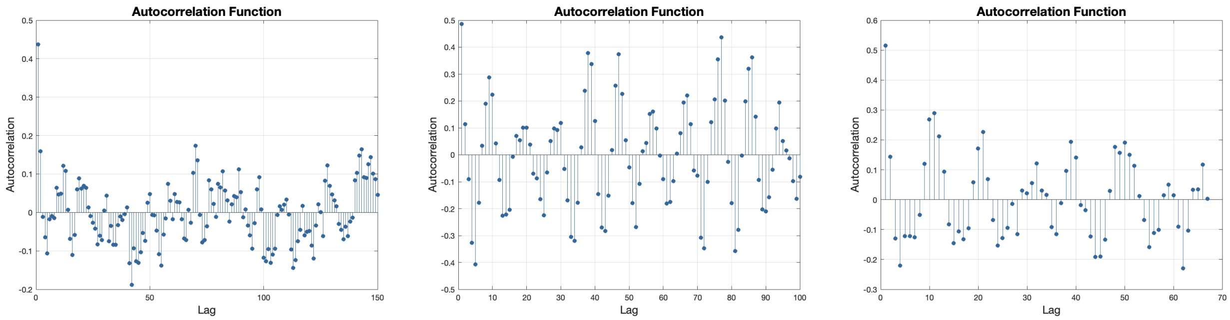
2.3.1 momentum autocorrelation

To check if sequence of momentum is self-related, we calculate the Pearson correlation between momentum and that with a time lag.

Algorithm 1 Calculate autocorrelation function

for i = 1 to 31 **do**
  $time\_series \leftarrow momentum(i^{th}\ match\_index : i+1^{th}\ match\_index - 1)$
  $max\_lag \leftarrow \lfloor length(time\_series)/2 \rfloor$ ▷ Consider lags up to half of the length of the time series
  $autocorrelation \leftarrow \text{zeros}(1, max\_lag)$
  for $shift$ = 1 to max_lag **do**
    $correlation \leftarrow \text{corrcoef}(time\_series(1 : \text{end} - shift), time\_series(shift + 1 : \text{end}))$
    $autocorrelation(shift) \leftarrow correlation(1, 2)$
  end for
  ▷ Further processing or visualization can be performed here
end for

Here we display the autocorrelation of momentum of the first player in first three games. There are similar results for the second player and for the momentum difference.



(a) autocorrelation in match 1 (b) autocorrelation in match 2 (c) autocorrelation in match 3

Figure 6: Momentum autocorrelation

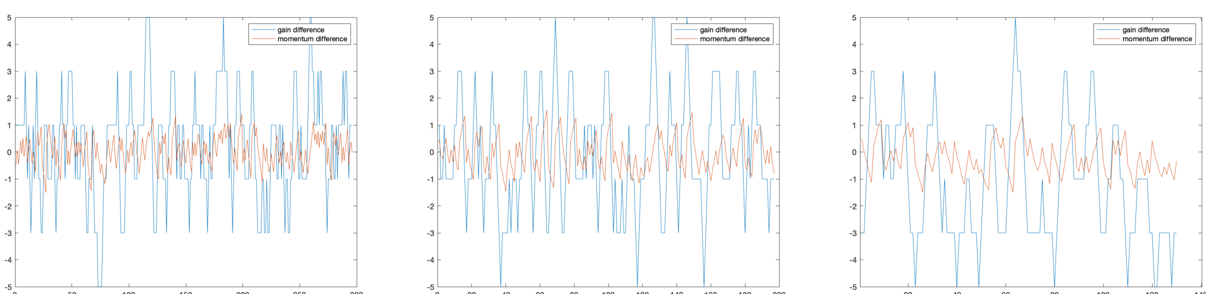
The corrcoef of lag 1 in match one is 0.4546, match two 0.5149,match three 0.5342. It can be seen that the autocorrelation of lag one is high, which means that the momentum at this moment has a strong relationship with the momentum in the next round. And the autocorrelation decreases to random as the lag increases.

2.3.2 correlation with runs of success

To give a quantitative evaluation of "future scores", we count points gain in future multiple rounds, and derive the difference by minus that of the opponent. For example, if the player gains 3 points in the next 5 rounds, and the opponent gains 2 points, the difference is 1. The difference indicates how much better the player is performing than the opponent.

In intuition, the player with higher momentum should have a higher chance to win the next round. And momentum at this moment should have less impact on the future rounds as time extends. The correlation between momentum and future scores verifies our intuition.

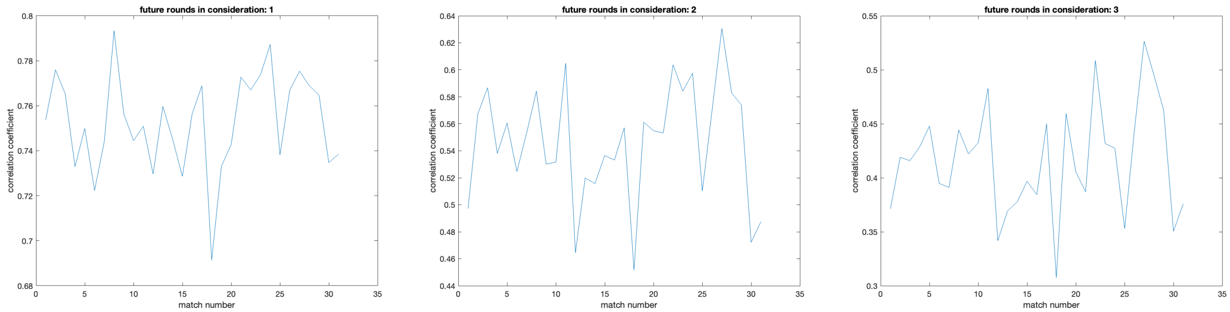
We calculate points gain difference in future one to five rounds at each time of all matches. Here we display five points gain difference and momentum difference in first three games.



(a) 2023-wimbledon-1301 (b) 2023-wimbledon-1302 (c) 2023-wimbledon-1303

Figure 7: Gain Difference and Momentum in First Three Games

And we derive the correlation between gain difference from one to five rounds and momentum difference of all matches. Here we display the three of them.



(a) gain difference in 1 round (b) gain difference in 2 rounds (c) gain difference in 3 rounds

Figure 8: Correlation Between Gain Difference and Momentum Difference

Here we display the max and min correlation in different rounds. (hint. the max and min correlation means maximum and minimum of all matches.)

rounds	1	2	3	4	5
max	0.7934	0.6307	0.5264	0.4678	0.3910
min	0.6914	0.4516	0.3074	0.1824	0.0627

Table 1: max and min Correlation of all matches in different rounds

As we can see from the table, the correlation between momentum and future scores is bigger than 0.5 considering the next 1 round, it implies that momentum has a substantial impact on the next round. And the correlation decreases as the rounds extend, which verifies our intuition, that the momentum at this moment has less impact on the future rounds as time extends.

Now, we have finished problem 1 and 2.

3 Swing Prediction Model

3.1 Model Overview

3.2 Visualization and Analysis

calculate the match percentage of each match between the prediction model and the result from AHP.

3.3 Weight of Factors

If using match data of multiple different players, the factor with bigger weight is effective to all players.

3.4 Advice against Opponent Based on Weight of Factors

If using match data of a single player, the factor with bigger weight is extremely effective to this player.

Now we have finished Problem 3.

3.5 Prediction Error with AHP model

Overlap rate of the prediction model and the result from AHP is?

Factors we may fail to consider First of all, the AHP model is not considering all possible factors to momentum. For example, the. So the overlap rate's effectiveness is limited.

Then, the prediction model is only a classification model,

4 Robustness Analysis

4.1 Generalization ability

Now, we have finished Problem 4.

4.2 Parameters Sensitivity Analysis

5 Strength and Weaknesses

5.1 Strengths

tableofcontents

5.2 Weaknesses

test

References

[1] Alice Axford, Bob Birkin, Charlie Copper, and Danny Dannford. Demostration of bibliography items. *Journal of TEXperts*, 36(7):114–120, Mar 2013.

[2] Timothy I Murphy. Line spacing in latex documents. [Online]. `https://www.popsci.com/article/technology/how-will-drones-change-sports/`.

A Advice to Tennis Coaches and Players on Momentum

[1]

[2]

A.1 1

test

A.1.1 1

test

1 test

1 test

B report on Use of AI

1.

OpenAI ChatGPT (Feb 2,2024 version,ChatGPT3.5)

Query1: *“swings in play and runs of success by one player are random” can this sentence to be interpreted as swings in play and runs of success by one player are independent?*

Output: *Yes, the sentence ”swings in play and runs of success by one player are random” can be interpreted as suggesting that swings in play and runs of success by one player are independent of each other. In statistical terms, independence means that the occurrence or outcome of one event does not influence the occurrence or outcome of another.*

In the context of tennis, this would imply that the fluctuations in the dynamics of the match (swings in play) and a player’s success streaks (runs of success) are not connected or dependent on each other. Each event is considered random and not influenced by the occurrence of the other.