國立政治大學統計學系

碩士學位論文

網球雙打選手的技能評分模型探討

A Study on Skill-Rating Models for Doubles Tennis Players

國立政治大學
National Chengchi University

指導教授：翁久幸 博士

研究生：林澤慶 撰

中華民國 110 年 7 月

DOI:10.6814/NCCU202101154

致謝

即將邁向畢業的我，首先感謝指導教授翁久幸老師的指導，讓我對於學術上的研究有初步的認識，在研究的過程中，老師給我研究方向與實驗過程很多的幫助，也在論文寫作上給予了指導。感謝研究所的同學，在這兩年的期間，一起修課與討論課業，讓我們之間彼此學習與成長，感謝家人一路的支持，讓我能專心於課業，感謝一路上幫助過我的人，讓我能渡過低潮與克服難關。

林澤慶

2021 年 7 月

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

摘要

技能評分系統是根據參賽隊伍的比賽戰績與其他相關資料，以評估隊伍實力與預測獲勝機率的一套計算方法。以職業網球比賽為例，選手的實力發揮往往受到比賽場地因素的影響，近年有學者針對網球單打提出一個將場地因素納入考量的模型，研究結果顯示納入場地因素的模型可以更準確地預測獲勝機率。本論文參考文獻中的相關討論，將此模型推廣至職業網球雙打比賽。本研究結果顯示納入場地因素於雙打比賽可以提高預測準確率，此外更新技能變數的變異程度也可提高預測準確率。

關鍵字：

技能評分系統、職業網球雙打、場地

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

Abstract

A skill-rating system is a system that evaluates players' skills and predict the winning probability based on competition records and relevant information. Take Association of Tennis Professionals as an example. Player's performance is often affected by the surfaces. In recent years, some researchers have proposed models that take surfaces into consideration for tennis singles. Research results show that models incorporating surfaces can more accurately predict the probability of winning. The present thesis aims to extend this model to tennis doubles matches. The experimental results show that the inclusion of surfaces in doubles matches can improve the accuracy of prediction.

Keyword: skill-rating system, professional tennis doubles, surfaces

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

目 錄

第 1 章 緒論 7
1.1 研究動機 7
1.2 研究目的 8
第 2 章 文獻回顧 9
2.1 Bradley-Terry Model 9
2.2 Elo 10
2.3 Glicko 11
第 3 章 研究方法 14
3.1 GenElo Surface Model 14
3.2 多組別多組員之比賽 16
3.3 結合場地於雙打比賽 18
第 4 章 資料介紹 23
4.1 ATP 雙打資料集 23
第 5 章 分析結果 25
5.1 參數設定 25
5.2 評估準則 28
5.3 模型預測結果 28
5.3.1 多組別多隊員之比賽 28
5.3.2 結合場地於雙打比賽 29
5.3.3 整體 32
第 6 章 結論與建議 33
參考文獻 34

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202101154

圖次

4.1.1 ATP 雙打-部分比賽資料 24
5.3.1 在測試集上，使用 3.2 節的模型評估每場比賽中勝隊與敗隊的實力。 29
5.3.2 在測試集上，用本論文提出的模型分別在紅土、草地與硬地的比賽場地中，評估每場比賽中勝隊與敗隊的實力。 31

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

表次

4.1.1 訓練集與測試集之場地與場數 23
5.1.1 選手在不同場地上實力的變異數與相關係數 27
5.3.1 在測試集中，評估模型預測的能力 32

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

第一章緒論

1.1 研究動機

網球一直被視為一種熱門的運動，從基層的學校網球社團、網球校隊、業餘的網球俱樂部，甚至到職業的網球比賽，皆有網球的活動。隨著網球運動的興起，越來越多的網球好手參加職業網球比賽，其中又以四大網球公開賽為主，分別是澳洲網球公開賽 (Australian Open)、法國網球公開賽 (French Open)、溫布敦網球公開賽 (The Championships, Wimbledon)、美國網球公開賽 (U.S. Open)，此外，我國好手也曾屢次獲得亮眼的表現，例如：詹皓晴與詹詠然在泰國芭達雅公開賽 (Thailand Open) 拿下冠軍，盧彥勳曾在馬哈拉施特拉邦公開賽 (Maharashtra Open) 獲得雙打冠軍。

隨著越多的好手在球場上發揮精湛的技術，除了有線電視的直播與網路直播的興起，搭配運動彩券的帶動下，網球比賽的討論度也越來越高，一些網球愛好者除了討論球員發揮的技術以外，亦會討論球員之間的比賽結果，甚至會投入資金於運動彩券，並且支持他們喜愛的球員，倘若能建立一個預測模型，並能有效預測球員間的比賽結果，因此，人們可以透過投注運動彩券，將此模型預測球員間的比賽結果，並在運動彩券上獲利。

過去討論職業網球比賽模型的文獻主要是以單打居多，例如：Kovalchik (2020) 撰寫的 Extension of the elo rating system to margin of victory、Ingram (2021) 撰寫的 How to extend Elo: a Bayesian perspective 等文獻，這些文獻皆有建立技能 (skill) 評分模型，以衡量各選手所發揮的技能或實力，並且預測網球選手之間的比賽結果，然而討論職業網球雙打的文獻偏少，因此本論文想要研究職業網球雙打的技能評分模型。

DOI:10.6814/NCCU202101154

1.2 研究目的

過去在常見評分系統有 Elo rating system(Elo, 1978)、Glicko rating system(Glickman, 1999) 等，這些評分系統皆可以評估參賽隊伍的實力，然而，這些評分系統可以在配對比較模型的幫助下，協助調整參賽隊伍的實力。職業網球選手的實力發揮往往會受到比賽場地的因素而受到影響，本論文參考Ingram (2021) 的研究，作者發現在網球單打上，場地的因素會影響到參賽選手實際表現。本論文參考Weng and Lin (2011) 的研究，作者提出多組別多隊員之比賽模型，因此，本研究基於多組別多隊員的模型，加入比賽場地的因素，讓模型能運用在職業網球雙打比賽上。

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

第二章 文獻回顧

配對比較 (Paired comparison) 或稱成對比較，簡言之就是兩兩之間的比較，也就是兩組隊伍進行比賽，常見於運動競賽、遊戲比賽等。常用的配對比較模型有 Thurstone-Mosteller 模型 (Thurstone, 1927) 與 Bradley-Terry 模型 (Bradley and Terry, 1952)，而 Elo 評分系統與 Glicko 評分系統就是分別使用這兩種模型。我們在下一章會介紹此二模型。起初，Elo 評分系統用在職業西洋棋比賽，用來評估選手的實力，後來也有運用於其他成隊的比賽，如足球賽等。Glicko 評分系統亦運用在西洋棋比賽上，至今，運動賽事也常採用配對比較方法來評估參賽者的實力，本節會介紹 Bradley-Terry 模型、Elo 評分系統與 Glicko 評分系統。

2.1 Bradley-Terry Model

Bradley and Terry (1952) 提出 Bradley-Terry Model，此為一種成對比賽的機率模型，它透過參賽者的實力計算獲勝機率。令 X_i 與 X_q 分別代表第 i 隊與第 q 隊在比賽中的實際表現，比方在球場上的實際表現會因為天候、場地、球員的身體狀況而有所差別，所以可將 X_i 與 X_q 視為兩個隨機變數，Bradley-Terry 模型如下：

$$P(X_i > X_q) = \frac{\pi_i}{\pi_i + \pi_q} \tag{2.1.1}$$

其中 $X_i > X_q$ 代表第 i 隊隊伍擊敗第 q 隊隊伍，$P(X_i > X_q)$ 代表第 i 隊隊伍擊敗第 q 隊隊伍的機率，π_i 代表第 i 隊隊伍的實力，π_q 代表第 q 隊隊伍的實力，此外，仍須滿足每一隊隊伍的評分 π 皆須大於等於 0 且 $\Sigma\pi_i = 1$，在 $\Sigma\pi_i = 1$ 的條件限制下，可以避免 (2.1.1) 式有無限多解。

假設每一場比賽結果是互相獨立的，第 i 隊隊伍與第 q 隊隊伍在第 k

DOI:10.6814/NCCU202101154

次對戰中，其發生對戰結果機率如下：

$$(\frac{\pi_i}{\pi_i+\pi_q})^{r_{iqk}}(\frac{\pi_q}{\pi_i+\pi_q})^{1-r_{iqk}} \tag{2.1.2}$$

在 (2.1.2) 式中，r_{iqk} 代表第 i 隊隊伍與第 q 隊隊伍在第 k 次對戰中的比賽結果，當比賽結果是第 i 隊隊伍擊敗第 q 隊隊伍時，則 $r_{iqk}=1$，反之，當第 q 隊隊伍擊敗第 i 隊隊伍時，則 $r_{iqk}=0$。

經過了 n 場比賽後，根據每一場比賽的結果資訊，我們可以獲得其概似函數，其概式函數呈現如下：

$$L(\pi)=\prod_{i<j}\prod_{k}(\frac{\pi_i}{\pi_i+\pi_q})^{r_{iqk}}(\frac{\pi_q}{\pi_i+\pi_q})^{1-r_{iqk}} \tag{2.1.3}$$

我們可以透過 (2.1.3) 式估計每一隊隊伍的實力 π。Bradley-Terry 模型 (2.1.1) 可以用若干不同方式推得，其中一種是令 X_i 服從 Gumbel 分配，其累積密度函數為

$$P(X_i\leq x)=exp(-exp(-(x-\theta_i))),\theta_i=log\pi_i$$

則 X_i-X_q 服從邏輯斯分配，其累積密度函數為

$$P(X_i-X_q\leq x)=\frac{e^{\theta_q}}{e^{\theta_i-x}+e^{\theta_q}}$$

當 $x=0$，則

$$P(X_i>X_q)=1-P(X_i\leq X_q)=\frac{e^{\theta_i}}{e^{\theta_i}+e^{\theta_q}}=\frac{\pi_i}{\pi_i+\pi_q},$$

即為 (2.1.1) 式。

2.2 Elo

Elo (1978) 提出 Elo 評分系統 (Elo rating system)，其對於參賽者實力的調整公式為：

$$new\ winner\ rating=old\ winner\ rating+k*(1-P(win))$$

DOI:10.6814/NCCU202101154

$$new\ loser\ rating = old\ loser\ rating - k * (1 - P(win))$$

其中獲勝機率 $P(win)$ 是對於該比賽獲勝者，我們在比賽前對他獲勝機率而估計。此機率可藉由配對比較的機率模型來計算。若使用 Bradley-Terry 模型，則此機率可以透過邏輯函數 (logistic function) 估計運動員獲勝的機率，k 代表調整參賽者實力的權重，當 k 值越大，則參賽者的實力越容易受到近期的表現影響，反之，當 k 值越小，則參賽者的實力越會參考更早期的表現，參考Elo (1978) 的設定，一般而言，可將 k 設定介於 10 至 32 之間，每一位參賽者的評分起始值設定為 1500，透過上述的方式調整每位參賽選手的評分。

根據 Elo 評分系統，第 i 位參賽者的獲勝機率為 (2.2.1) 式：

$$p(y = 1|\hat{\theta}_i, \hat{\theta}_j) = \frac{1}{1 + 10^{-(\hat{\theta}_i - \hat{\theta}_j)/400}} \tag{2.2.1}$$

參賽者的評分更新方式為 (2.2.2) 式和 (2.2.3) 式：

$$\hat{\theta}'_w = \hat{\theta}_w + k \times (1 - p(y = 1|\hat{\theta}_w, \hat{\theta}_l)) \tag{2.2.2}$$

$$\hat{\theta}'_l = \hat{\theta}_l - k \times (1 - p(y = 1|\hat{\theta}_w, \hat{\theta}_l)) \tag{2.2.3}$$

(2.2.2) 式和 (2.2.3) 式中 $\hat{\theta}_w$ 與 $\hat{\theta}_l$ 分別代表更新前贏家的技能評分與輸家的技能評分，$\hat{\theta}'_w$ 與 $\hat{\theta}'_l$ 分別代表更新後贏家的技能評分與輸家的技能評分，並顯示贏家的技能評分會被向上調整 $k \times (1 - p(y = 1|\hat{\theta}_w, \hat{\theta}_l))$，輸家的技能評分會被向下調整 $k \times (1 - p(y = 1|\hat{\theta}_w, \hat{\theta}_l))$，更新後的贏家技能評分與輸家技能評分之和與更新前的贏家評分與輸家評分之和相同，故贏家技能評分與輸家技能評分之和維持一定。

2.3 Glicko

Glickman (1999) 提出的 Glicko 評分系統是一個基於 Bradley-Terry 模型的

DOI:10.6814/NCCU202101154

貝氏技能評分系統。此方法之主要貢獻是納入技能評分的可靠性 (Ratings Reliability)，也就是除了點估計之外，還考慮其變異程度。Glicko 評分系統會根據每個時間點賦予每位參賽者實力參數 θ，其隨著時間變化而受到波動。假設參賽者實力參數 θ 為一隨機變數，且其事前分配服從常態分配，表示如下：

$$\theta|\sigma_0^2 \sim N(1500, \sigma_0^2) \tag{2.3.1}$$

此系統還假設隨著時間的推移，參賽者實力的變異程度也隨著增加，且與間隔時間成正比。所以在時間 t_0 時，若參賽者實力參數分布如下：

$$\theta^{(t_0)}|\mu, \sigma^2 \sim N(\mu, \sigma^2) \tag{2.3.2}$$

則經過時間 t 個單位後，其實力參數分布如下：

$$\theta^{(t_0+t)}|\theta^{(t_0)}, \nu^2, t \sim N(\mu, \nu^2 t) \tag{2.3.3}$$

根據 (2.3.2) 與 (2.3.3) 可以得到下式：

$$\theta^{(t_0+t)}|\mu, \sigma^2, \nu^2, t \sim N(\mu, \sigma^2 + \nu^2 t) \tag{2.3.4}$$

接下來作者透過近似理論，估算 θ 的事後分配之期望值 μ' 與變異數 σ'^2，作為 μ、σ^2 的更新方法，其更新式子如下：

$$\mu' = \mu + \frac{b}{1/\sigma^2 + 1/\delta^2} \Sigma_{j=1}^{m} \Sigma_{k=1}^{n_j} g(\sigma_j^2)[s_{jk} - E(s|\mu, \mu_j, \sigma_j^2)] \tag{2.3.5}$$

$$\sigma'^2 = (\frac{1}{\sigma^2} + \frac{1}{\delta^2})^{-1} \tag{2.3.6}$$

其中

$$b = log(10)/400$$

$$g(\sigma^2) = \frac{1}{\sqrt{1 + 3b^2\sigma^2/\pi^2}}$$

DOI:10.6814/NCCU202101154

$$E(s|\mu,\mu_j,\sigma_j^2) = \frac{1}{1+10^{-g(\sigma_j^2)(\mu-\mu_j)/400}} \tag{2.3.7}$$

$$\delta^2 = [b^2\Sigma_{j=1}^m n_j g(\sigma_j^2)^2 E(s|\mu,\mu_j,\sigma_j^2)[1-E(s|\mu,\mu_j,\sigma_j^2)]]^{-1}$$

上述 (2.3.7) 式代表參賽者碰到第 j 位參賽者預期會獲勝的機率，b 為常數，n_j 代表與第 j 位參賽者比賽的場數，s 代表比賽結果，m 代表遭遇參賽者的個數，(2.3.5) 式為事後評分的更新方式，其更新的方法與 Elo 評分系統更新的方式相似，會透過比賽結果與期望會獲勝的機率來調整參賽者的實力，若預期獲勝的機率與實際比賽結果差異過大，則參賽者實力會有較大幅調整。

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154

第三章研究方法

3.1 GenElo Surface Model

職業網球選手會因為場地的不同材質，而影響到參賽者的實力發揮，網球場地的材質基本上可以分成 3 類，分別是紅土（Clay）、草地（Grass）與硬地（Hard）。紅土球場的特性在於球速較為緩慢，有比較多的時間追球，對於防禦型底線型球員有較好的發揮空間，草地球場的特性在於球落地時有較小的摩擦力，且不時會有不規則的彈跳，對於發球上網型球員有較佳的發揮空間，硬地球場的特性在於球的彈跳速度快，因此讓對手有較少的反應時間，對於侵略型底線型球員較具有優勢。

GenElo Surface Model 由Ingram (2021) 將 Elo 評分模型進行擴增，此模型是將 Elo 評分模型考慮了比賽場地的因素，讓模型除了參考雙方選手的實力與比賽的結果之外，亦會將與場地資訊當作參考依據。

為了讓模型能配適場地的變數，首先令 i 與 j 選手的實力參數為 $\boldsymbol{\theta_i}$ 與 $\boldsymbol{\theta_j}$，假設其事前分配服從常態：

$$\boldsymbol{\theta_i} \sim \boldsymbol{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma}); \boldsymbol{\theta_j} \sim \boldsymbol{N}(\boldsymbol{\mu_j}, \boldsymbol{\Sigma}) \tag{3.1.1}$$

在 (3.1.1) 式中，$\boldsymbol{\theta_i}$ 與 $\boldsymbol{\theta_j}$ 皆為長度為 n 的向量，分別代表贏家與輸家在 n 種不同場地上的實力，$\boldsymbol{\mu_i}$ 與 $\boldsymbol{\mu_j}$ 皆為長度為 n 的向量，Σ 為 $n*n$ 的矩陣，其代表在不同場地之間，參賽者實力的共變異數矩陣，且假設所有參賽者皆有共同的共變異數矩陣 Σ。比方，以網球比賽而言，場地可能是紅土、草皮、硬地等三種，則 $n = 3$，且 $\boldsymbol{\theta_i} = (\theta_{i1}, \theta_{i2}, \theta_{i3})$，其中 θ_{ik} 代表勝方在第 k 種場地之實力。

DOI:10.6814/NCCU202101154

在假設 i 與 j 的實力彼此獨立的情況下，$\boldsymbol{\theta_i}$ 與 $\boldsymbol{\theta_j}$ 的聯合分配如下：

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta_i} \\ \boldsymbol{\theta_j} \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{\mu_i} \\ \boldsymbol{\mu_j} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma} \end{bmatrix} \right) = \boldsymbol{N}(\boldsymbol{\mu_\theta}, \boldsymbol{\Sigma_\theta}) \tag{3.1.2}$$

我們介紹 $\boldsymbol{a}$，其定義為：

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{a_i} \\ -\boldsymbol{a_j} \end{bmatrix} \tag{3.1.3}$$

在 (3.1.3) 式中，$\boldsymbol{a_i} = \boldsymbol{a_j}$，且皆為長度為 n 的 dummy vector，在該比賽的場地指派為 1，其餘則指派為 0，比方，若該場地為第 1 種場地，則 $\boldsymbol{a_i} = \boldsymbol{a_j} = (1, \underbrace{0, ..., 0}_{n-1})$，所以一場比賽由 $\boldsymbol{i}$ 獲勝的機率可以寫成 $P(i\ win|\boldsymbol{\theta_i}, \boldsymbol{\theta_j}) = \gamma(b\boldsymbol{a^T\theta})$，其中

$$\gamma(x) = logit^{-1}(x)$$

$$b = log(10)/400$$

現在考慮 $\boldsymbol{\theta}$ 在給定比賽結果之下的後驗分配，將其取對數後可得出式子如下：

$$t(\boldsymbol{\theta}) \propto -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu_\theta})^{\boldsymbol{T}}\boldsymbol{\Sigma_\theta^{-1}}(\boldsymbol{\theta} - \boldsymbol{\mu_\theta}) + log\gamma(b\boldsymbol{a^T\theta}) \tag{3.1.4}$$

將 (3.1.4) 式利用牛頓法 (Newton's method) 求出最大值，對應之 $\boldsymbol{\theta}$ 即為事後分配之眾數 (posterior mode)，以此作為實力的後驗分配之期望值，為簡化更新式，作者考慮一次迭代，其更新的方法如下：

$$\boldsymbol{\mu'_\theta} = \boldsymbol{\mu_\theta} - H^{\boldsymbol{-1}}(\boldsymbol{\mu_\theta})\boldsymbol{J}(\boldsymbol{\mu_\theta}) \tag{3.1.5}$$

在 (3.1.5) 式中 H 和 $\boldsymbol{J}$ 分別代表 $t(\boldsymbol{\theta})$ 的 Hessian matrix 與 Jacobian，其中 $H(\boldsymbol{\mu_\theta})$ 與 $\boldsymbol{J}(\boldsymbol{\mu_\theta})$ 可以表示如下：

$$\boldsymbol{J}(\boldsymbol{\mu_\theta}) = (1 - \gamma(b\boldsymbol{a^T\mu_\theta}))b\boldsymbol{a} \tag{3.1.6}$$

DOI:10.6814/NCCU202101154

$$H(\boldsymbol{\mu_\theta}) = -\boldsymbol{\Sigma_\theta^{-1}} - \gamma(b\boldsymbol{a^T}\boldsymbol{\mu_\theta})(1-\gamma(b\boldsymbol{a^T}\boldsymbol{\mu_\theta}))b^2\boldsymbol{aa^T} \tag{3.1.7}$$

我們可以透過 (3.1.5)、(3.1.6) 與 (3.1.7) 式的更新方式，參賽選手的實力依照比賽結果進行調整。

3.2 多組別多組員之比賽

Weng and Lin (2011) 提出 Bradley-Terry Model with full-pair 與 partial pair 評分模型，該模型可以適用於 $k \geq 2$ 組隊伍且每隊隊伍有多位選手的比賽。當 $k = 2$ 且每隊有 2 人的情況，就是常見的雙打比賽。當 $k > 2$，比方 $k = 4$，代表有 4 組隊伍 A,B,C,D。若一場比賽之結果，A,B,C,D 依序為第 1, 第 2, 第 3, 第 4，此結果可以用 A 贏 B，B 贏 C，C 贏 D 這 3 個兩兩比賽的結果代表；也可以用 A 贏 B，B 贏 C，C 贏 D，A 贏 C，A 贏 D，B 贏 D 這 6 個兩兩比賽的結果代表之。前者以 $(k-1)$ 個兩兩比賽結果代表一場 k 隊比賽結果，稱之為 partial-pair；後者以 C_2^k 個兩兩比賽結果代表一場 k 隊比賽結果，稱之為 full-pair。不管是 partial-pair 或 full-pair，在網球比賽 $(k = 2)$ 得 $k - 1 = C_2^k = 1$，結果相同。每經過一場比賽後，可以將組內參賽選手的實力進行調整，令 X_i 代表第 i 隊的實際表現。比方在球場上的實際表現會因為天候、場地、球員的身體狀況而有所差別，所以可將 X_i 視為隨機變數。此模型是將 Bradley-Terry Model 進行擴充，讓模型能多考慮到變異數的部分，此模型為：

$$P(X_i > X_q) = \frac{e^{\theta_i/c_{iq}}}{e^{\theta_i/c_{iq}} + e^{\theta_q/c_{iq}}} \tag{3.2.1}$$

其中

$$c_{iq}^2 = \beta_i^2 + \beta_q^2$$

其中 θ_i 代表第 i 隊的實力，β_i 代表第 i 隊隊伍實際表現的不確定性，c_{iq}^2 代表第 i 隊隊伍與第 q 隊隊伍之不確定性的加總。

DOI:10.6814/NCCU202101154

假設第 i 隊的實力 θ_i 亦為一隨機變數，且在比賽前其事前分配為 $N(\mu_i, \sigma_i^2)$。當比賽結束後，可以透過貝式方法估計第 i 隊實力的事後期望值與變異數，亦即更新隊伍實力的平均數與變異數。此模型假設各隊之 μ_i，σ_i^2 分別為其隊員之事前平均數，事前變異數之加總，且假設 $\beta_i^2 = \sigma_i^2 + \beta^2$，其中 β^2 是事先給定之正數。該論文的 Algorithm 1 詳述當 k 隊伍同時參與一場比賽且每隊伍人數不盡相同的情況下，各選手實力更新之演算法。當 $k = 2$，亦即每場比賽僅有 2 組，且每組各有 2 位選手，亦即網球雙打的情況，其更新的演算法簡化如下：

步驟 1：給定每一場比賽的結果 s 與每一隊隊伍中隊員的 $\mu_{ij}, \sigma_{ij}^2 (i = 1, 2, j = 1, 2)$ 與 $\beta^2, \kappa > 0$，並決定 γ_q

步驟 2：針對一場比賽中每一隊計算 $\mu_i, \sigma_i^2 (i = 1, 2)$

$$\mu_i = \Sigma_{j=1}^2 \mu_{ij}, \sigma_i^2 = \Sigma_{j=1}^2 \sigma_{ij}^2$$

步驟 3：針對一場比賽中每一隊團隊評分與個人實力進行調整

步驟 3.1：透過計算 Ω_i 與 Δ_i 來調整團隊實力

For $i = 1, 2$

令 $c^2 = (\sigma_1^2 + \sigma_2^2 + 2\beta^2), \hat{p}_i = \dfrac{e^{\mu_i/c}}{e^{\mu_1/c} + e^{\mu_2/c}}$

$$\Omega_i = \frac{\sigma_i^2}{c}(s - \hat{p}_i), \Delta_i = \gamma_q (\frac{\sigma_i}{c})^2 \hat{p}_1 \hat{p}_2, \text{ 其中 } s = \begin{cases} 1 & if\ i\ wins \\ 1/2 & if\ tie \\ 0 & if\ i\ loses \end{cases}$$

步驟 3.2：針對每隊隊伍中 2 位參賽選手實力進行調整

For $j = 1, 2$

$$\mu_{ij} \leftarrow \mu_{ij} + \frac{\sigma_{ij}^2}{\sigma_i^2}\Omega_i, \sigma_{ij}^2 \leftarrow \sigma_{ij}^2 max(1 - \frac{\sigma_{ij}^2}{\sigma_i^2}\Delta_i, \kappa)$$

其中 κ 避免在更新參數的階段時，當 $1 - \dfrac{\sigma_{ij}^2}{\sigma_i^2}\Delta_i$ 太接近 0，導致 σ_{ij}^2 會趨近

DOI:10.6814/NCCU202101154

於 0，使得失去個人實力的變異性，γ_q 可以控制 σ_i^2 下降的幅度，通常建議將 γ_q 設置為 $1/k$，其中 k 代表一場比賽中隊伍數。

透過上述的演算法，可以將網球雙打比賽中每一位運動員的實力，依照比賽的最終結果，進行個別的調整。若模型只更新隊伍實力的平均數，卻不更新隊伍實力的變異數，其更新過程同上，只需將步驟 3.2 的 σ_{ij}^2 的更新部份省略即可。此簡化模型相較於上述的完整模型，每完成一場比賽，簡化模型只考慮隊伍實力的平均數調整，不考慮隊伍實力的變異數變動。

3.3 結合場地於雙打比賽

第 3.1 節中介紹Ingram (2021) 將 Elo 模型考慮了比賽場地的因素，第 3.2 節中介紹Weng and Lin (2011) 將 Bradley-Terry 模型擴充成可以讓組內隊員同時進行調整實力，本論文提出的想法是結合此兩者方法，並應用於網球雙打之比賽資料。亦即結合第 3.1 節考慮比賽場地的因素與第 3.2 節可以估計組內每一位選參賽者的實力。

首先，假設第 i 隊隊伍中第 j 位參賽者的實力如下：

$$\varphi_{ij} = \begin{bmatrix} \varphi_{ij1} \\ \varphi_{ij2} \\ \varphi_{ij3} \end{bmatrix} \sim N(\mu_{ij} = \begin{bmatrix} \mu_{ij1} \\ \mu_{ij2} \\ \mu_{ij3} \end{bmatrix}, \Lambda_{ij}) \tag{3.3.1}$$

在 (3.3.1) 式中，$\varphi_{ij1}, \varphi_{ij2}, \varphi_{ij3}$ 分別對應第 i 隊隊伍中第 j 位參賽者在紅土、草地與硬地的實力。假設每位選手實力的變異數皆相同，所以不管 i，j 如何，Λ_{ij} 都相同。將第 i 隊隊伍中每位參賽者的實力進行加總以獲得第 i 隊隊伍的實力，得出第 i 隊隊伍實力如下：

$$\theta_i = \Sigma_{j=1}^2 \varphi_{ij} = \begin{bmatrix} \theta_{i1} \\ \theta_{i2} \\ \theta_{i3} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \end{bmatrix}, \Lambda \right) \tag{3.3.2}$$

DOI:10.6814/NCCU202101154

$$\Lambda = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}$$

其中在 (3.3.2) 式中 Λ 裡的參數 $\sigma_m^2, m = 1, 2, 3$ 分別代表有關於紅土、草地與硬地的變異數，ρ_{ij} 代表隊伍評分在不同場地間的相關係數，將 θ_{i1} 與 θ_{i2} 的變異數矩陣相加後記做 Λ。

由於 $\boldsymbol{\theta_i}$ 和 $\boldsymbol{\theta_q}$ 相互獨立，聯合先驗分配如下：

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta_i} \\ \boldsymbol{\theta_q} \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{\mu_i} \\ \boldsymbol{\mu_q} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda} \end{bmatrix} \right) = \boldsymbol{N}(\boldsymbol{\mu_\theta}, \boldsymbol{\Sigma}) \tag{3.3.3}$$

令 s_{iqk} 代表參賽隊伍 i 碰上參賽隊伍 q 在場地 k 的比賽結果，則 $\boldsymbol{\theta}$ 的概似函數如下：

$$L(\boldsymbol{\theta_i}, \boldsymbol{\theta_q}; s) = p(s|\boldsymbol{\theta_i}, \boldsymbol{\theta_q}) = (\frac{e^{\theta_{ik}/c_k}}{e^{\theta_{ik}/c_k} + e^{\theta_{qk}/c_k}})^{s_{iqk}} (\frac{e^{\theta_{qk}/c_k}}{e^{\theta_{ik}/c_k} + e^{\theta_{qk}/c_k}})^{1-s_{iqk}}$$

其中

$$c_k = (\sigma_k^2 + \sigma_k^2 + 2\beta^2)^{0.5}$$

c_k^2 代表此兩隊伍實際表現的不確定性，其由選手實力在該場地的變異性 σ_k^2 與比賽中不可控制的變異性 β^2 所組合而成的。

倘若第 i 隊隊伍獲勝，則比賽結果紀錄為 $s_{iqk} = 1$，反之，則比賽結果紀錄為 $s_{iqk} = 0$。接著 loglikelihood 表示如下：

$$l(\boldsymbol{\theta_i}, \boldsymbol{\theta_q}; s) = logL(\boldsymbol{\theta_i}, \boldsymbol{\theta_q}) = s_{iqk}\frac{\theta_{ik}}{c_k} + (1 - s_{iqk})\frac{\theta_{qk}}{c_k} - log(e^{\theta_{ik}/c_k} + e^{\theta_{qk}/c_k}) \tag{3.3.4}$$

DOI:10.6814/NCCU202101154

透過式 (3.3.3) 與式 (3.3.4) 的推導，$\boldsymbol{\theta}$ 的後驗分配取對數後可表示如下：

$$
\begin{aligned}
t(\boldsymbol{\theta}) &= logp(\boldsymbol{\theta}|s) \\
&\propto -\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu_\theta})^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}}(\boldsymbol{\theta}-\boldsymbol{\mu_\theta}) \\
&\quad + s_{iqk}\frac{\theta_{ik}}{c_k} + (1-s_{iqk})\frac{\theta_{qk}}{c_k} - log(e^{\theta_{ik}/c_k}+e^{\theta_{qk}/c_k})
\end{aligned}
\tag{3.3.5}
$$

參考 3.1 章節的作法，引進 $\boldsymbol{\alpha}$，其中 $\boldsymbol{\alpha}$ 為：

$$
\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha_i} \\ -\boldsymbol{\alpha_q} \end{bmatrix} \tag{3.3.6}
$$

在 (3.3.6) 式中，$\boldsymbol{\alpha_i} = \boldsymbol{\alpha_q}$ 皆為長度為 3 的 dummy vector，在該比賽的場地指派為 1，其餘則指派為 0。

以比賽場地為硬地為例，此時 $k=3$，將 $t(\boldsymbol{\theta})$ 進行一階偏微分後，結果如下：

$$
\nabla t(\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{\theta}) = -\boldsymbol{\Sigma^{-1}}(\boldsymbol{\theta}-\boldsymbol{\mu_\theta}) + \frac{1}{c_3}\begin{bmatrix} 0 \\ 0 \\ s_{iq3} - \dfrac{e^{\theta_{i3}/c_3}}{e^{\theta_{i3}/c_3}+e^{\theta_{q3}/c_3}} \\ 0 \\ 0 \\ (1-s_{iq3}) - \dfrac{e^{\theta_{q3}/c_3}}{e^{\theta_{i3}/c_3}+e^{\theta_{q3}/c_3}} \end{bmatrix}
$$

DOI:10.6814/NCCU202101154

將 $t(\boldsymbol{\theta})$ 進行二階偏微分，結果如下：

$$\nabla^2 t(\boldsymbol{\theta}) = H(\boldsymbol{\theta}) = -\Sigma^{-1} + \frac{1}{c_3^2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\dfrac{e^{(\theta_{i3}+\theta_{q3})/c_3}}{(e^{\theta_{i3}/c_3}+e^{\theta_{q3}/c_3})^2} & 0 & 0 & \dfrac{e^{(\theta_{i3}+\theta_{q3})/c_3}}{(e^{\theta_{i3}/c_3}+e^{\theta_{q3}/c_3})^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{e^{(\theta_{i3}+\theta_{q3})/c_3}}{(e^{\theta_{i3}/c_3}+e^{\theta_{q3}/c_3})^2} & 0 & 0 & -\dfrac{e^{(\theta_{i3}+\theta_{q3})/c_3}}{(e^{\theta_{i3}/c_3}+e^{\theta_{q3}/c_3})^2} \end{bmatrix}$$

利用牛頓法更新 $\boldsymbol{\mu_\theta}$，其更新的過程如下：

$$\boldsymbol{\mu'_\theta} = \boldsymbol{\mu_\theta} - H^{-1}(\boldsymbol{\mu_\theta})\boldsymbol{J}(\boldsymbol{\mu_\theta}) = \boldsymbol{\mu_\theta} + \boldsymbol{\delta} \tag{3.3.7}$$

在 (3.3.7) 式中 H 與 $\boldsymbol{J}$ 分別代表 $\boldsymbol{t(\theta)}$ 的 Hessian matrix 與 Jacobian，其中 $H(\mu_\theta)$ 與 $\boldsymbol{J(\mu_\theta)}$ 可以表示如下：

$$J(\mu_\theta) = \frac{1}{c_3}\begin{bmatrix} 0 \\ 0 \\ s_{iq3} - \dfrac{e^{\mu_{i3}/c_3}}{e^{\mu_{i3}/c_3}+e^{\mu_{q3}/c_3}} \\ 0 \\ 0 \\ 1 - s_{iq3} - \dfrac{e^{\mu_{q3}/c_3}}{e^{\mu_{i3}/c_3}+e^{\mu_{q3}/c_3}} \end{bmatrix} \tag{3.3.8}$$

DOI:10.6814/NCCU202101154

$$H(\boldsymbol{\mu_\theta}) = -\Sigma^{-1} + \frac{1}{c_3^2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{e^{(\mu_{i3}+\mu_{q3})/c_3}}{(e^{\mu_{i3}/c_3}+e^{\mu_{q3}/c_3})^2} & 0 & 0 & \frac{e^{(\mu_{i3}+\mu_{q3})/c_3}}{(e^{\mu_{i3}/c_3}+e^{\mu_{q3}/c_3})^2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{e^{(\mu_{i3}+\mu_{q3})/c_3}}{(e^{\mu_{i3}/c_3}+e^{\mu_{q3}/c_3})^2} & 0 & 0 & -\frac{e^{(\mu_{i3}+\mu_{q3})/c_3}}{(e^{\mu_{i3}/c_3}+e^{\mu_{q3}/c_3})^2} \end{bmatrix} \tag{3.3.9}$$

透過 (3.3.7) 式，每經過一場比賽即可更新隊伍的實力。隊伍中隊員的實力則可以透過隊伍實力的變化量進行更新，其更新的分法如下：

$$\mu'_{ij} = \mu_{ij} + \frac{1}{2}\delta_i \tag{3.3.10}$$

其中 δ_i 為第 i 隊隊伍評分更新的變化量，隨著更新隊伍實力的同時，亦可以同時隊員的實力。

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202101154

第四章 資料介紹

本次實驗資料來自男子職業網球協會 (Association of Tennis Professionals，ATP) (https://www.flashscore.com/tennis/atp-doubles/)，ATP 為男子職業網球協會，其宗旨在於維護男性職業網球運動員的利益，並舉辦全球網球巡迴賽。

4.1 ATP 雙打資料集

資料主要記載每場比賽的情形，每場比賽包含球場場地、盃賽名稱、比賽級別、比賽日期、贏家姓名、輸家姓名、比賽分數等資料，時間從 2010 年至 2019 年，總共有 13043 場比賽。本研究不考慮中場自願退賽 (retirement)、失格 (default)、不戰而勝 (walkover) 的比賽，此外，本研究也不納入台維斯盃（Davis Cup）的比賽。由於中場自願退賽、失格與不戰而勝的比賽皆未完成完整的比賽，故本研究不採用，而台維斯杯的比賽是以國家為單位的團體比賽，不是本研究想考慮的對象，資料經過篩選後，總共有 12430 場比賽。我們將 2010 年至 2017 年間的比賽做為訓練集，2018 年與 2019 年間的比賽則為測試集，訓練集與測試集所包含之場地數與場地如表 4.1.1：

表 4.1.1: 訓練集與測試集之場地與場數

資料集	訓練集	測試集	總數
比賽場數	9859	2571	12430
Hard	5548	1464	7012
Clay	3112	767	3879
Grass	1199	340	1539

DOI:10.6814/NCCU202101154

圖4.1.1為部分資料，可觀察每場比賽資料有比賽編號 tourney id、比賽場地 surface、比賽日期 tourney date、第一位勝方姓名 winner1 name、第二位勝方姓名 winner2 name、第一位負方姓名 loser1 name、第二位負方姓名 loser2 name、比賽分數 score 等。

	tourney_id	surface	tourney_date	winner1_name	winner2_name	loser1_name	loser2_name	score
12560	2010-339	Hard	2010-01-03	Lukas Dlouhy	Leander Paes	Carsten Ball	Sam Querrey	7-5 6-3
12561	2010-339	Hard	2010-01-03	Jeremy Chardy	Marc Gicquel	Julian Knowle	Robert Lindstedt	3-6 7-6 10-4
12562	2010-339	Hard	2010-01-03	Michael Llodra	Andy Ram	Peter Luczak	Joseph Sirianni	3-6 6-4 10-7
12563	2010-339	Hard	2010-01-03	Marcelo Melo	Bruno Soares	Ross Hutchins	Jordan Kerr	6-2 6-4
12564	2010-339	Hard	2010-01-03	Tomas Berdych	Radek Stepanek	Thomaz Bellucci	Paul Hanley	6-1 6-4

圖 4.1.1: ATP 雙打-部分比賽資料

男子職業雙打比賽可以分成四大網球公開事、ATP 世界巡迴賽總決賽、ATP 世界巡迴賽 1000 大師系列賽、ATP 世界巡迴賽 500 系列賽事、ATP 世界巡迴賽 250 系列賽等，而每一個系列賽或是公開賽會有對應的網球比賽制度，有些系列賽會採取小組循環賽，小組前幾名則晉級下一輪，例如：ATP 世界巡迴賽總決賽，有些則是採取單敗淘汰賽制，比賽獲勝者則晉級下一輪，例如：四大網球公開賽，這些賽制皆有共通點，越靠近系列賽的尾部，代表參賽隊伍皆是贏得前面的比賽，也意味著強隊遇到強隊，彼此之間的實力差距應差異不大。

DOI:10.6814/NCCU202101154

第五章分析結果

我們首先討論在不同模型下，各別設置不同的參數來配適模型，並在不同的評估準則下，判斷不同的參數設定下，模型預測的表現能力。

5.1 參數設定

本篇研究結合了Ingram (2021) 的 GenElo surface model 與Weng and Lin (2011) 的多組別多組員之實力更新方法，預測網球雙打比賽結果。兩者模型皆有一些超參數需要被設定，透過優化 (optimize) 的方式，找出較佳的超參數設定，讓模型在預測能力上有更好的表現。

在第 3.2 節中 Bradley-Terry Model(Full-pair) 模型需要設定初始參數有每隊中參賽者實力 μ_{ij}、參賽者實力的變異數 σ_{ij}^2、比賽中不可控制的變異數 β^2、控制因子 κ 與 γ_q，參考第 2.1 節中的參數設定，本研究將新進選手的實力 μ_{ij} 初始值設為 1500，比賽中不可控制的變異數設為 9，控制因子 κ 設為 0.0001，參考Weng and Lin (2011) 的設定，將 γ_q 設置為 $\frac{1}{2}$，參賽者實力的變異數 σ_{ij}^2，則透過以下的方式進行優化。

本研究參考Glickman (1999) 採用最小化交叉熵 (cross-entropy) 的方式進行優化，交叉熵為一種在分類問題中常見的損失函數，透過此優化方式，將訓練集的比賽資料進行訓練，找出適當的初始參賽者實力的變異數 σ_{ij}^2，讓預測結果越能貼近比賽真實結果。每一場比賽中的交叉熵如下：

$$d_{iq} = -s_{iq}log(p_{iq}) - (1 - s_{iq})log(1 - p_{iq}) \tag{5.1.1}$$

其中 s_{iq} 代表第 i 隊遇上第 q 隊的比賽結果，p_{iq} 代表第 i 隊隊伍擊敗第 q 隊隊伍的機率。

$$p_{iq} = \frac{e^{\mu_i/c_{iq}}}{e^{\mu_i/c_{iq}} + e^{\mu_q/c_{iq}}}$$

DOI:10.6814/NCCU202101154

$$c_{iq} = (\sigma_i^2 + \sigma_q^2 + 2\beta^2)^{0.5}$$

$$\sigma_i^2 = \Sigma_{j=1}^2 \sigma_{ij}^2$$

目標函數則由每場比賽中的交叉熵進行加總而來，即為 Σd_{iq}，亦即為負對數概似函數 (negative loglikelihood)，此目標函數會隨著每一場參賽者實力變動而影響，參賽者實力的調整亦會收到先前比賽的結果有所更動。因為此目標函數難以對 σ_{ij}^2 微分，故本研究採用由Nelder and Mead (1965) 所提出的 Nelder-Mead 方法，其能找尋極小值，並找尋到較佳的 σ_{ij}^2，本研究假設 σ_{ij}^2 不因 i，j 而改變，亦即 σ_{ij}^2 =const，經由交叉熵最小化之結果。得到此估計值為 3.55。在只更新隊伍實力的平均數，卻不更新隊伍實力變異數的模型裡，採取優化的方式與上述相同，得到 σ_{ij}^2 估計值為 1.75。

由第 3.3 節的 (3.3.7) 式可知結合場地資訊於雙打比賽的模型需要設定的初始參數有每隊中參賽者實力 μ_{ijk}、比賽中不可控制的變異數 β^2、參賽選手在各場地發揮實力的變異數 σ_k^2 與在各場地間的相關係數 ρ_{12}、ρ_{23} 與 ρ_{13}，此研究亦將每隊中參賽者實力 μ_{ijk} 設置為 1500，比賽中不可控制的變異數 β^2 設為 9，參賽選手在各場地發揮實力的變異數 σ_k^2 與在各場地間的相關係數 ρ_{12}、ρ_{23} 與 ρ_{13}，則透過優化的方式來取得。

此處優化方式亦採取最小化交叉熵，每一場比賽中的交叉熵 d_{iqk} 如下：

$$d_{iqk} = -s_{iqk} log(p_{iqk}) - (1 - s_{iqk}) log(1 - p_{iqk}) \tag{5.1.2}$$

其中 s_{iqk} 表示第 i 隊與第 q 隊在場地 k 的比賽結果，p_{iqk} 代表第 i 隊隊伍在場地 k 擊敗第 q 隊隊伍的機率。

$$p_{iqk} = \frac{e^{\mu_{ik}/c_k}}{e^{\mu_{ik}/c_k} + e^{\mu_{qk}/c_k}}$$

$$c_k = (\sigma_k^2 + \sigma_k^2 + 2\beta^2)^{0.5}$$

$$\boldsymbol{\mu'_\theta = \mu_\theta - H^{-1}(\mu_\theta) J(\mu_\theta)}$$

DOI:10.6814/NCCU202101154

將每一場比賽的交叉熵進行加總形成目標函數，即為 Σd_{iqk}，亦即為負對數概似函數 (negative loglikelihood)，此目標函數會隨著每一場參賽者實力的變動受到影響，參賽者實力的調整會受到先前比賽的結果而有所變動，而當該場地比完賽時，除了參賽者在該場地發揮的實力會做到調整以外，參賽者在其他場地發揮的實力亦會受到調整。因為目標函數難以對 σ_k^2、ρ_{12}、ρ_{23} 與 ρ_{13} 偏微分，故選擇用由Nelder and Mead (1965) 所提出的 Nelder-Mead 方法，能找到 $(\sigma_1^2,\sigma_2^2,\sigma_3^2,\rho_{12},\rho_{13},\rho_{23})$ 的組合，並使交叉熵達到最小。

表 5.1.1: 選手在不同場地上實力的變異數與相關係數

σ_{clay}^2	σ_{grass}^2	σ_{hard}^2	$\rho_{clay-grass}$	$\rho_{clay-hard}$	$\rho_{grass-hard}$
41.00	250.37	29.79	0.57	0.92	0.88

表5.1.1為選手在不同場地上的實力變數的變異數與相關係數，透過 Nelder-Mead 方法，將使 Σd_{iqk} 達最小的 $(\sigma_1^2,\sigma_2^2,\sigma_3^2,\rho_{12},\rho_{13},\rho_{23})$ 組合。由此表可知，相關係數皆為正，代表當選手在比賽場地發揮的實力有正向的調整時，在其他場地發揮的實力也會有正向的調整，草地的變異數相較於其他場地的變異數還要大，這種情況也是合理的，草地的場地會因為天氣的狀況不同，影響球速也有所不同，此外，相較於其他場地，草地較容易出現一些不規則彈跳，會影響到比賽選手實力的發揮。此外，紅土與硬地的相關係數最高，與草地的相關係數最低。

在測試集中，將先前於訓練集優化完的參數配適於模型中，即可獲得每位參賽者於比賽前的實力的平均數以及變異數，將隊伍中隊員的實力相加即可獲得隊伍的實力，並依照比賽前隊伍實力高低而預測，將隊伍實力較高的預測為勝方，反之，隊伍實力較低的預測為負方。

DOI:10.6814/NCCU202101154

5.2 評估準則

為了能夠有效比較出不同模型的預測能力，我們這次採用統計學與機器學習上常用的指標，分別為準確度 (Accuracy) 與 mean log-likelihood 這兩種指標當作模型的評估準則。首先，透過模型預測比賽結果與實際比賽結果比較，計算出此模型能成功預測比賽結果的比例，代表模型的準確度，若準確度越高，代表模型有愈好的預測能力。其次，我們計算 mean log-likelihood 這個數值也可以做為模型配適優劣的參考，數值越大代表配適越好。

5.3 模型預測結果

5.3.1 多組別多隊員之比賽

首先，我們使用 3.2 節之模型建模並預測比賽結果，過程中必須評估每場比賽勝隊與敗隊的實力，我們將測試集中關於各隊實力之估計與對應之比賽結果呈現於圖5.3.1。圖5.3.1的左半部為每一場比賽勝隊與敗隊的實力分布圖，其中橘點代表勝隊的實力大於敗隊的實力，亦即代表此模型預測成功的部份，反之，藍點代表勝隊的實力小於敗隊的實力，亦即代表此模型預測錯誤的部份，由此圖可知，由於大部分的點為橘色的，代表大致而言，模型預測結果是正確的。圖5.3.1的右半部為兩隊平均實力差距的直方圖，橘色與藍色分別代表勝隊的實力大於敗隊的實力與勝隊的實力小於敗隊的實力，由圖可知，隨著兩隊平均的實力差距越大，可看出橘色的部分佔的比例增加，代表模型預測比賽結果更加準確，換而言之，當兩隊的平均實力差異不大時，比較難預測比賽結果。

DOI:10.6814/NCCU202101154

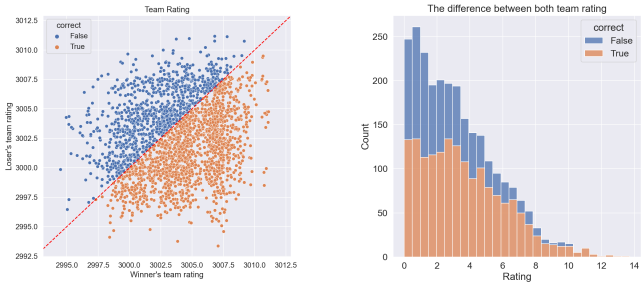
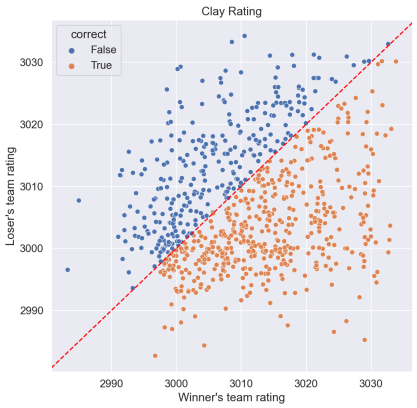


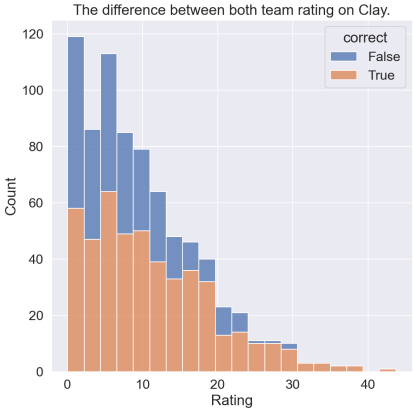
圖 5.3.1: 在測試集上，使用 3.2 節的模型評估每場比賽中勝隊與敗隊的實力。

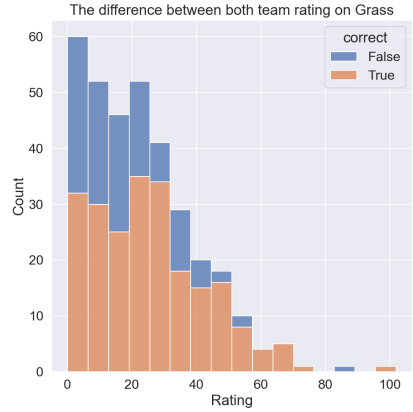
5.3.2 結合場地於雙打比賽

接著，我們利用本論文所提出的模型預測比賽結果，此模型與 3.2 節的模型差異在於有無加入場地資訊。並且評估每一場比賽的敗隊與勝隊的實力，並且將結果呈現於圖5.3.2。圖5.3.2由上至下分別是在比賽場地為紅土、草地、硬地下，每場比賽勝隊與敗隊實力分布情況，由左至右分別為在該場地下，勝隊與敗隊實力分布和兩隊實力之間的差距，橘色代表該比賽中預測勝隊的實力大於預測敗隊的實力，藍色則代表該比賽中預測勝隊的實力小於預測敗隊的實力。由圖5.3.2的上半部可知，不論比賽場地為何種，橘色的點皆比藍色的點還要多，大致而言，模型預測比賽結果是正確的。由圖5.3.2的右半部可知，不論比賽場地為何種，皆可發現隨著兩隊的實力差距越大，可以看出橘色的部分所佔的比例增加，代表模型愈能預測比賽結果，此外，我們發現當比賽場地為草地時，相較於其他場地，兩隊實力的差距分布較廣，其原因可從表5.1.1觀察出 σ^2_{grass} 明顯比 $\sigma^2_{clay}, \sigma^2_{hard}$ 大上不少。

DOI:10.6814/NCCU202101154







DOI:10.6814/NCCU202101154

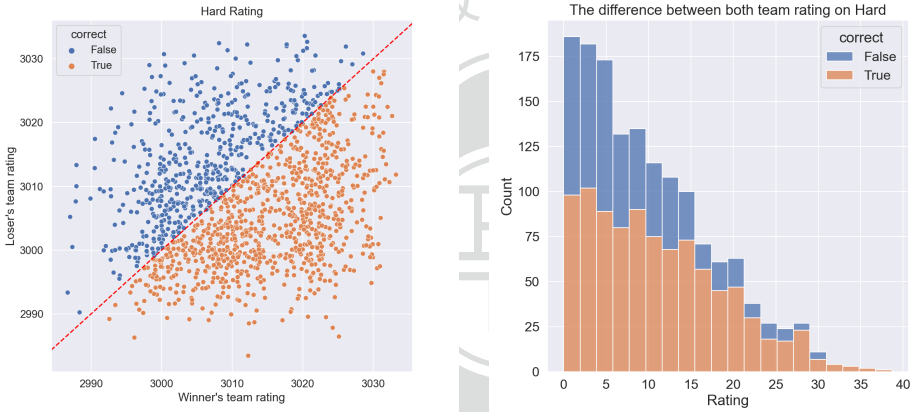


圖 5.3.2: 在測試集上，用本論文提出的模型分別在紅土、草地與硬地的比賽場地中，評估每場比賽中勝隊與敗隊的實力。

DOI:10.6814/NCCU202101154

5.3.3 整體

表5.3.1為模型在訓練集訓練完參數後，在測試集中，評估各模型的準確度與 mean log-likelihood，在準確度方面而言，未加入場地資訊且變異數更新的模型 (3.2 節) 的準確度 0.6359 與加入場地資訊且變異數未更新的模型 (3.3 節) 的準確度 0.6328，皆略高於未加入場地資訊且變異數未更新的模型的準確度 0.6317，在 mean log-likelihood 方面而言，未加入場地資訊且變異數更新的模型的 mean log-likelihood 達-0.6393 與加入場地資訊且變異數未更新的模型的 mean log-likelihood 達-0.6692，皆略低於未加入場地資訊且變異數未更新的模型的 mean log-likelihood 達-0.6392，整體而言，在雙打比賽中，加入場地資訊或變異數更新似乎可以增進模型的預測能力。

表 5.3.1: 在測試集中，評估模型預測的能力

Model	Accuracy	Mean log-likelihood
未加入場地資訊且變異數未更新的模型	0.6317	-0.6392
未加入場地資訊且變異數更新的模型	0.6359	-0.6393
加入場地資訊且變異數未更新的模型	0.6328	-0.6692

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202101154

第六章 結論與建議

本研究以男子職業網球比賽為例，藉由評分模型探討網球雙打選手的技能，並考慮比賽場地的因素，評估每隊在各種場地下所發揮的實力，從模型訓練的過程中，藉由訓練資料集估計超參數，接續在固定超參數的情況下，從測試集中估算比賽兩隊各別的實力，藉此預測獲勝的隊伍。

本研究將Weng and Lin (2011) 的多組別多隊員的評分方法運用於 ATP 雙打資料集，分別考慮不加入場地資訊與加入場地資訊兩種情況，預測準確率分別為 0.6359 與 0.6328，mean log-likelihood 分別為-0.6393 與-0.6692。雖然加入場地資訊的模型表現並未比較好，但該模型可透過共變異數矩陣 $\mathbf{\Lambda}$ 得知不同場地間的關聯性，在關聯程度上紅土與草地是相對低的，也驗證了紅土的特性在於球速慢，草地的特性在於球速快，兩者場地的特性極為不同。

雖然本研究是運用在網球的資料上，但一般而言，本研究可以同樣應運用在其他運動比賽上。在沙灘排球方面，Glickman et al. (2018) 曾利用 Elo、Glicko 等評分系統評估女子沙灘排球的排名，女子沙灘排球為雙打比賽項目，此外，比賽場地可以分為室外天然沙灘或人工建置的沙坑，本研究可以嘗試將結合場地於雙打之模型運用在女子排球沙灘比賽上，藉此評估女子沙灘排球的排名。

本文評估準則主要是參考Ingram (2021) 的作法，其中的 mean log-likelihood 並未考慮模型的複雜程度，未來可以使用 AIC 或 BIC 做為評估準則。此外，本研究發現考慮場地且變異數不更新的模型與不考慮場地且變異數更新的模型，準確度皆高於未考慮場地且變異數不更新的模型，建議未來可以往考慮場地且變異數更新的模型進行研究。

DOI:10.6814/NCCU202101154

參考文獻

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub., New York.

Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(3):377–394.

Glickman, M. E., Hennessy, J., and Bent, A. (2018). A comparison of rating systems for competitive women's beach volleyball. *Statistica Applicata-Italian Journal of Applied Statistics*, 30(2):233–254.

Ingram, M. (2021). How to extend elo: a bayesian perspective. *Journal of Quantitative Analysis in Sports*.

Kovalchik, S. (2020). Extension of the elo rating system to margin of victory. *International Journal of Forecasting*, 36(4):1329–1341.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286.

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202101154

Weng, R. C. and Lin, C.-J. (2011). A bayesian approximation method for online ranking. *Journal of Machine Learning Research*, 12(9):267–300.

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202101154