Predict Rising Stars in Sports: An Example of NBA Players

Ying-Ho Liu, Yu-Xiang Hong*

Department of Information Management, National Dong Hwa University, Taiwan
daxliu@gms.ndhu.edu.tw, 611035102@gms.ndhu.edu.tw

Abstract

Predicting rising stars is of public interest. Professional personnel search for rising sports stars is based on experience, which is usually highly subjective. Typically, players' statistical data are used to make predictions, but rookie players lack abundant statistics, and some personal characteristics may not be reflected in the statistics. Alternatively, online discussions like tweets contain information about players, so the Text Sentiment Deep Prediction for Athletes Career (TSDPAC) algorithm was proposed to use tweets related to NBA players and sentiment analysis to predict whether a player is a rising star. The TSDPAC adopted deep learning techniques to construct the prediction model, which effectively predicted rising stars in the experiments and outperformed the compared algorithms.

Keywords: Rising star, Deep learning, Sentiment analysis, NBA player

1 Introduction

Exploring employees' career development is of public interest, particularly in predicting the "rising star" in the workplace, that is, an individual who shows the potential to become a star. The search for rising stars is a fascinating and potentially substantive activity, typically performed by professionals who use subjective judgment to identify such individuals based on relevant information. Given the limited human experience and possible biases, studies have been dedicated to finding rising stars using algorithms and referring to a large amount of data.

Among these studies, some focused on finding rising stars among scholars [1-9], and other research attempts to find rising stars in sports [10-19]. Sports stars attract a lot of attention as competitive sports arouse people's emotions and immerse people in them. Coupled with the rapid development of media technology, people have convenient access to various sports events, which have become a part of daily life.

The existence of sports stars is a widely discussed phenomenon. Rosen [20] pointed out that superstars are a relatively small group of people who earn a lot of money and dominate a field; this is particularly true in professional sports. Fans often focus on a star's every move on and off the field, and such stars bring economic benefits to those engaged in sports event-related industries, especially the operators of professional teams. Lucifora and Simmons [21] proposed that professional teams benefit from the superstar effect, reflected in their revenue. Consequently, it will be a very cost-effective investment for team operators to recruit potential stars as early as possible.

Some studies devoted to predicting rising stars have used the players' statistical data to make predictions. Still, unquantifiable characteristics, such as the ability to read the game, perseverance, concentration, etc., can only be presented partially from this data. Moreover, rookie players may have limited statistical data, so it isn't easy to make an adequate evaluation based solely on statistical data. Alternatively, there may be a large amount of text data. For example, sports commentators may write about particular players in sports columns, and there is a vast amount of data in the form of discussions and comments left by the general public on the Internet. Even experts post and interact with people through social media, so such text resources are worthy of development. In addition, due to the limited number of experts and market considerations (sponsor requirements, need to increase click-through rates, etc.), the experts often discuss some well-known players. However, potential rising stars who are less well-known may be unlikely to be mentioned in their articles. The general public does not have the burden of the experts. Everyone has different interests and concerns, so they have a greater chance of noticing, discussing, and evaluating less well-known players, providing more diverse content than expert commentary. The emotions hidden in the text can be used to determine the author's sentiment tendency toward the article's topic. Therefore, sentiment analysis of the player-related text is an important step in understanding the author's player evaluation of players. This study applied sentiment analysis to predict the career development of NBA players using tweets on X (https://twitter.com/) such as:

- Devin Booker is gonna be a great NBA player.
- For an 18-year-old kid to come out there and make the plays, it was pretty great.
- Myles Turner has stood out as much as anyone in USA Select team workouts this week.
- Mikal Bridges is the next Russell Westbrook.
- Mikal Bridges is just everything you'd want in a basketball player.
- Collin Sexton continues making that pull-up jumper he could not consistently hit a year ago. Great progress already.

*Corresponding Author: Ying-Ho Liu; Email: daxliu@gms.ndhu.edu.tw
DOI: https://doi.org/10.70003/160792642026012701002

- Jakob Poeltl is again making a difference at the rim - nice help to deny Jeff Green's lefty scoop.

The first three tweets commented on a player's overall performance, the fourth is related to a player's career growth, and the last three are related to a player's skills and career growth. Looking at these tweets, apart from emotional words, there is also a correlation between the tweets and the player's career development, suggesting that using social media posts to predict players' career development is feasible. Therefore, this study proposed to use tweets on social media X to predict whether an NBA player is a rising star. Due to the flourishing development of deep learning techniques in text analytics, deep learning techniques were combined with sentiment analysis to develop the Text Sentiment Deep Prediction for Athletes Career (TSDPAC) algorithm to predict players' career development using textual information.

The contributions of this study were four-fold. First, this study proposed the TSDPAC algorithm to predict whether a player is a rising star. Second, this study used deep learning techniques and sentiment analysis to make predictions. To the best of our knowledge, this methodology has never been developed to predict rising stars in the literature. Third, we conducted experiments using actual tweets to verify the performance of the TSDPAC algorithm. Fourth, we compared the TSDPAC with four machine learning algorithms. The remainder of this paper is organized as follows: Section 2 presents the literature review, Section 3 introduces the proposed method, Section 4 evaluates the model's performance, and Section 5 presents the conclusions and recommendations for future work.

2 Literature Review

This section reviews studies on predicting rising stars in sports and academia.

2.1 The Studies on the Sports Area

Chou *et al*. [10] predicted whether NBA players would be elected to the Hall of Fame using statistics such as steals, blocks, points, and field goal percentages as input to construct artificial neural network (ANN) and convolutional neural network (CNN) models, with the ANN model exhibiting better predictive power. Patton *et al*. [11] used players' videos released by the NCAA to predict the probability that college players would enter the NBA. Ahmad *et al*. [12] explored how to predict the rising stars in cricket using machine learning techniques, the players' data from the same match, and team status to rank the rookie players. Mahmood *et al*. [13] considered the relevance of NBA players in the same game for predicting rising stars using players' statistics and machine learning techniques. Graig and Winchester [14] used data on college soccer players and some ESPN scouting reports to predict whether a college quarterback would be drafted by an NFL team. Mulholland and Jensen [15] used college data, NFL composite scores, and physical metrics to predict NFL draft orders and NFL tight-end players' career development. Pitts and Evans [16] showed a positive correlation between the Wonderlic score and quarterback performance in the NFL. Rosen and Olbrecht [17] found that quarterbacks who demonstrated "functional mobility" in college performed better in their NFL careers than those who did not. In addition, some studies have shown that college statistics are not very useful in predicting NFL players' performance [18]. Berri and Simmons [19] evaluated the factors that NFL teams consider when selecting quarterbacks in the draft and the relationship between draft position and NFL performance, reporting that many college metrics that improve a quarterback's draft position were unrelated to their future performance in the NFL.

2.2 The Studies on the Academia

Li *et al*. [1] proposed the PubRank algorithm to identify rising academic stars using a PageRank-like algorithm to map the relationships between scholars' articles into a network. Wijegunawardana *et al*. [2] considered scholars' interactions on social media and extracted their co-authorship, citation, and publication venue information to obtain a list of recommendations for rising stars. Duad *et al*. [3] proposed using entropy to assess publications' quality dynamically and suggested that when calculating mutual influence among scholars, the extent of each scholar's contribution to academic articles should be considered. Daud *et al*. [4] regarded whether a researcher is a rising star as a classification application and extracted eleven characteristics from papers' authors, co-authors, and publication information using machine learning techniques to perform classification. Daud *et al*. [5] conducted another study on the rising stars of the co-author network (co-author network) in which they introduced the weighted mutual influence rank (WMIRank) technique. When scholars' research papers are published in multiple research areas, it is also valuable to refer to the information of popular research areas to identify the rising stars. Other studies used topic modeling [6], case studies [7], QRank [8], and cluster analysis [9].

Since identifying rising stars is an important and exciting topic and the existing studies seldom considered textual material like tweets and deep learning techniques to predict sports rising stars, This study developed the TSDPAC algorithm to address this research gap.

3 The Proposed Method

This section introduces the proposed method. Section 3.1 describes the collected data and defines star players and rising stars. Section 3.2 elaborates on the TSDPAC algorithm.

3.1 Data Collection and Definitions

In total, 216 players who joined the NBA during the 2010~2011 season through the 2017~2018 season and played at least 200 games in their first five seasons were identified. Tweets related to the first two seasons of each player were collected from social media X using X API (https://developer.x.com/en/docs/twitter-api). The number of collected tweets is 1,822,862. The average, maximum, and minimum number of tweets of a player are 8,439, 11,951, and 859. Common positive words that appeared in tweets of rising stars included *good*, *greatest*, *decent*, *excellent*, *fine*, *well*, and *nice*. In contrast, negative words like *bad*, *awful*, *dead*, *dumb*, *hilarious*, *horrible*, *nasty*, *poor*, and *terrible* were more likely to appear in non-rising stars' tweets. This suggested that using sentiment analysis was appropriate for finding rising stars.

The PER (Player Efficiency Rating) designed by John Hollinger [22] was used to evaluate the players' performance. Compared to the EFF (Efficiency) [23], the PER is a more detailed representation of a player's performance by adding weights to their various statistics and considering their playing time and the team's tempo. The NBA league average PER is set at 15 for each season, so comparing a player's PER across seasons is possible. Table 1 shows the PER reference range set by John Hollinger. Star players and rising stars were defined as follows:

Definition 1. A **star player** in a season is a player with a PER of 20 or more in this season.

Definition 2. If a player is a star player in his fifth season, he is a **rising star** in his first two seasons.

For example, Devin Booker's PER in his fifth season was 20.6, i.e., he was a star player this season and a rising star in his first two seasons.

Table 1. PER reference range [22]

Classification	PER range
All-time great season	35.0+
Runaway MVP candidate	30.0–35.0
Strong MVP candidate	27.5–30.0
Weak MVP candidate	25.0–27.5
Definite All-Star	22.5–25.0
Borderline All-Star	20.0–22.5
Second offensive option	18.0–20.0
Third offensive option	16.5–18.0
Slightly above-average player	15.0–16.5
Rotation player	13.0–15.0
Non-rotation player	11.0–13.0
Fringe roster player	9.0–11.0
Player who won't stick in the league	0–9.0

The number of rising stars in each season is listed in Table 2. The first two seasons' tweets were used to predict whether a player would be a star in his fifth season because a rookie-scale contract is guaranteed for two years, with an average NBA contract of 3.3 years [24]. The team operators must assess whether to sign or extend a player as early as their second season and typically, a three-year contract (on average) is considered. Therefore, they evaluate if a player will perform well during their fifth season.

Table 2. The number of rising stars in each season

Season	Total number of players	Number of rising stars
2010-2011	26	7
2011-2012	31	9
2012-2013	28	5
2013-2014	29	5
2014-2015	28	6
2015-2016	30	4
2016-2017	22	2
2017-2018	22	4

3.2 The TSDPAC Algorithm

The TSDPAC algorithm was designed as a deep learning multi-layered structure to predict whether an NBA player is a rising star (Figure 1), and each layer is detailed below.

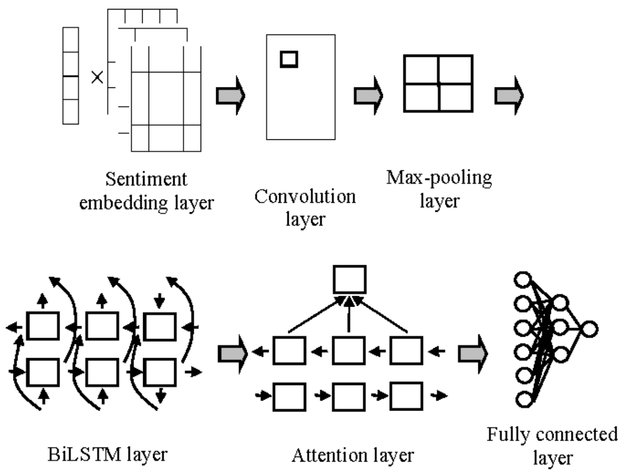

Figure 1. The TSDPAC algorithm

3.2.1 Sentiment Embedding Layer

The text data (player-related tweets on social media X) were converted into a *word vector matrix* by the BERT (Bidirectional Encoder Representations from Transformers) algorithm [25] and then weighted with the sentiment scores defined by SentiWordNet [26]. Player-related tweets were retrieved via the Twitter API, and special symbols, such as punctuation marks and emojis, as well as semantically irrelevant content, such as URL links, were removed. Then, each text was converted into word vectors with BERT representing a word as a vector to form a two-dimensional word vector matrix (if a text consists of w words and the length of the word vector is v, a two-dimensional matrix with dimensions $w \times v$ can be formed).

Since some words have sentiment meanings, the sentiment tendency of a tweet was crucial, so the sentiment words were weighted to reflect the sentiment tendency of the tweet using SentiWordNet. Each sentiment word collected in SentiWordNet has three sentiment scores representing the degrees of positivity, negativity, and objectivity. Each score was a value between 0 and 1, and the three scores for a word added up to 1.

After converting a tweet into a word vector matrix, the word vectors of the sentiment words were further weighted. For a sentiment word, subtract the negative sentiment score from the positive sentiment score of the word and multiply it by 10 to obtain its "Sentiment Tendency Weight." A Sentiment Tendency Weight greater than 0 means the word is more positive, less than 0 means the word is more negative, and close to 0 means the word is more neutral. The sentiment vector E_V was defined for each tweet, and the length of E_V is the number of words in the tweet. E_V served as the weight when weighting the word vector matrix. The i^{th} element e_i in E_V corresponds to the i^{th} word in the tweet, and e_i is defined as in Equation 1:

$$e_i = \begin{cases} s_i,\ if\ \ w_i \in SWN\ \ and\ \ s_i \neq 0 \\ 1,\ otherwise \end{cases} \quad \textbf{(1)}$$

Where w_i is the ith word in a tweet, s_i is the Sentiment Tendency Weight of w_i, and SWN is the SentiWordNet. e_i quantifies the impact of words according to their sentiment intensity. A word with strong positive or negative sentiment had a significant value. The weighted word vector t_i of w_i was derived using Equation 2:

$$t_i = e_i \cdot o_i \quad \textbf{(2)}$$

Where o_i is the word vector of w_i (i.e., i^{th} row in the word vector matrix), e_i is the value obtained from Equation 1. The *weighted word vector matrix* could be obtained by applying Equation 2 to each word vector in the word vector matrix.

3.2.2 Convolution Layer and Max-pooling Layer

These layers extracted the feature matrices with high-level concepts by applying convolution and pooling operations to the weighted word vector matrix generated in the sentiment embedding layer. The convolution layer performs a "convolution operation," as shown in Figure 2. The "Input data" is a two-dimensional matrix, and the "Kernel" is a 3×3 convolution kernel. The convolution kernel moves horizontally and vertically on the two-dimensional matrix and executes a Hadamard product with the content of the overlaying matrix every time it moves to a new position, thus obtaining the matrix presented by the "Convoluted feature" in Figure 2. Accordingly, the convolution operation retrieved the convoluted feature from a two-dimensional matrix.

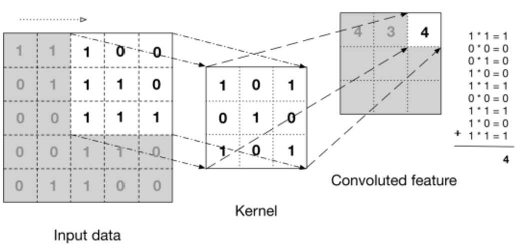

Figure 2. The convolution operation [27]

The convoluted feature was then processed by the max-pooling operation in the max-pooling layer. Figure 3 presents a max-pooling operation with a 2×2 kernel. The original matrix on the left was pooled by taking the maximum value to produce the matrix on the right, thereby reducing the matrix's size and avoiding overfitting for subsequent training.

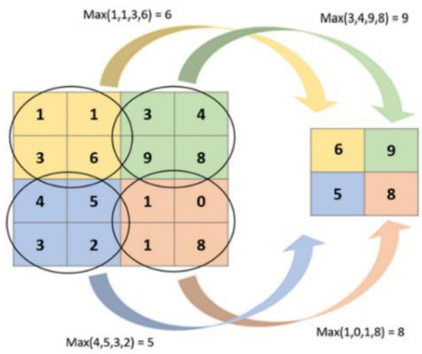

Figure 3. The max-pooling operation [28]

These layers can be iterated many times and use kernels of different dimensions on the weighted word vector matrix to retrieve higher-level feature matrices to explore the association between words.

3.2.3 BiLSTM Layer

Long short-term memory (LSTM) is a recurrent neural network (RNN) variant that can choose whether to retain or forget certain information through the forget gate, input gate, and output gate. Therefore, LSTM excels in dealing with long sentences. However, the text semantics need to be considered as a whole to be understood, and the meaning of a word is often related to the preceding and following text. The BiLSTM is a two-layer LSTM architecture, where one LSTM is used to retrieve the contextual semantics from start to end of text, and another LSTM is used to retrieve the contextual semantics in the reverse direction. The BiLSTM output is a combination of the outputs of the two LSTM layers to integrate the contextual relationships. A flattened operation converts the feature matrix generated by the convolution and max-pooling layers to a one-dimensional vector that serves as the BiLSTM layer's input.

3.2.4 Attention Layer

For an application only generating a decision (like determining whether a player is a rising star), the LSTM used the final time step's hidden state as output, but this ignores the information in the previous time steps' hidden states. Bahdanau *et al*. [29] proposed the Attention mechanism which integrates each time step's hidden state to generate the output. Figure 4 shows the BiLSTM combined with the Attention with the hidden states of each time step aggregated in the Attention layer to generate the output.

3.2.5 Fully Connected Layer

The output of the Attention layer is fed into the fully connected layer to generate the final decision. The fully connected layer is a vanilla neural network with one hidden layer used by the TSDPAC to infer whether a player is a rising star.

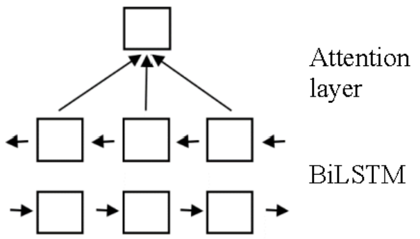

Figure 4. The BiLSTM layer and Attention layer

4 Experiment Results

Three experiments verified the performance of the TSDPAC. The first experiment considered the prediction accuracy within the same season, the second experiment examined all-time prediction, and the third experiment verified the prediction performance for the next season. In each experiment, the sentiment embedding layer first concatenated all player tweets and then processed the concatenated tweet to retrieve a weighted word vector matrix. The first 1000 words of the concatenated tweet were used in the sentiment embedding layer to give each player's weighted word vector matrix an equal dimension. All experiments were performed on a PC with an Intel(R) Core i7-11700, 32-GB main memory, running Microsoft Windows 10 (64-bit). The methods were implemented using Python packages (Keras, TensorFlow, and scikit-learn).

Two CNN layers and two max-pooling layers were used in the experiments. The first CNN layer had 16 kernels with a dimension (32, 32) and the second 16 kernels with a dimension (4, 4). The dimensions of the first max-pooling layer are (16, 16) and the second (2, 2). The number of units in BiLSTM was 128.

4.1 Rising Star Prediction within a Season

The first experiment not only assessed the ability of tweets to predict rising stars but also verified the consistency of tweets on different players in the same season using five-fold cross-validation. For the season with less than five rising stars, the number of folds was set as the number of rising stars.

Table 3 shows the results. The prediction accuracy of each season was above 90%, and while the number of rising stars in some seasons was low, missing a rising star resulted in a lower recall. However, the precision was excellent, which led to a good F1 score, and the AUC was good; therefore, the TSDPAC predicted rising stars well within a season, and tweets differed between rising stars and non-rising stars.

The TSDPAC was also compared to the algorithms using players' statistics to make predictions, including the support vector machine (SVM), artificial neural network (ANN), C4.5 decision tree, and naïve Bayesian classifier (NBC). Players' statistical data were downloaded from the ESPN website, including twenty-five statistics such as games played, minutes per game, average field goals made, 3-point field goal percentage, rebounds per game, assists per game, steals per game, blocks per game, points per game, etc. Each compared algorithm used the players' average statistics from the first two seasons to construct a model to predict whether the players are rising stars. The accuracy of each algorithm is presented in Table 4. The TSDPAC significantly outperformed the compared algorithms.

Table 3. The performance within a season

Season	R	P	F1	AUC	Acc
2010-2011	0.857	0.857	0.857	0.923	92.3%
2011-2012	1.0	0.9	0.947	0.975	96.7%
2012-2013	0.6	1.0	0.75	0.8	92.8%
2013-2014	0.6	1.0	0.75	0.8	93.1%
2014-2015	0.67	1.0	0.795	0.8	92.8%
2015-2016	0.5	1.0	0.67	0.75	93.3%
2016-2017	0.5	1.0	0.67	0.75	95.4%
2017-2018	0.5	1.0	0.67	0.75	90.9%

R: recall, P: precision, F1: F1 score, Acc: accuracy

Table 4. The accuracy within a season

Season	TSDPAC	SVM	ANN	C4.5	NBC
2010-2011	92.3%	70.3%	77.7%	66.6%	81.4%
2011-2012	96.7%	80.6%	80.6%	67.7%	77.4%
2012-2013	92.8%	82.1%	85.7%	89.2%	75.0%
2013-2014	93.1%	75.8%	68.9%	68.9%	72.4%
2014-2015	92.8%	92.8%	89.2%	78.5%	75.0%
2015-2016	93.3%	83.3%	83.3%	86.6%	90.0%
2016-2017	95.4%	86.3%	90.9%	90.9%	90.9%
2017-2018	90.9%	81.8%	68.1%	77.2%	81.8%

4.2 All-time Rising Star Prediction

In this experiment, the 216 players were pooled, and five-fold cross-validation was applied to test the TSDPAC performance with more players. The results are shown in Table 5 and Table 6. The TSDPAC's performance was excellent, with high recall, precision, and F1 score, and it significantly outperformed the other algorithms.

Table 5. The all-time performance

R	P	F1	AUC	Acc
0.881	1.0	0.937	0.93	97.2%

R: recall, P: precision, F1: F1 score, Acc: accuracy

Table 6. The all-time accuracy

TSDPAC	SVM	ANN	C4.5	NBC
97.2%	82.9%	80.1%	80.1%	77.8%

4.3 Rising Star Prediction for the Next Season

It is practical for professional teams to predict rising stars using previous data, so this experiment simulated this process. The players before the nth season were used to construct the prediction model to identify rising stars in the nth season. The results are listed in Table 7 and Table 8. Overall, the accuracy was high, and although two seasons achieved an accuracy under 90%, the recall rate was high. It may be more important to find a rising star than misclassify a non-rising star, considering all the positive effects brought by a rising star. The TSDPAC outperformed other algorithms.

Table 7. The prediction results for the next season

Season	R	P	F1	AUC	Acc
2011-2012	1.0	0.69	0.818	0.9	87.1%
2012-2013	0.8	1.0	0.89	0.9	96.4%
2013-2014	0.8	0.57	0.66	0.84	86.2%
2014-2015	0.66	1.0	0.8	0.833	92.8%
2015-2016	0.75	1.0	0.857	0.875	96.6%
2016-2017	1.0	1.0	1.0	1.0	100%
2017-2018	0.75	1.0	0.857	0.875	95.4%

R: recall, P: precision, F1: F1 score, Acc: accuracy

Table 8. The accuracy for the next season

Season	TSDPAC	SVM	ANN	C4.5	NBC
2011-2012	87.1%	74.1%	67.7%	61.2%	74.1%
2012-2013	96.4%	89.2%	85.7%	85.7%	78.5%
2013-2014	86.2%	86.2%	82.7%	75.8%	82.7%
2014-2015	92.8%	78.5%	78.5%	78.5%	75.0%
2015-2016	96.6%	90.0%	83.3%	83.3%	90.0%
2016-2017	100%	90.9%	95.4%	81.8%	72.7%
2017-2018	95.4%	72.7%	77.2%	77.2%	81.8%

5 Conclusion and Future Work

Predicting rising stars in the sports area is important and exciting to the public and team operators. Typically, player statistics have been used to make predictions but there is a colossal amount of text information on social media. Therefore, the TSDPAC algorithm was developed using CNN and BiLSTM to convert and weight text data from tweets to predict rising NBA stars. The TSDPAC extracted BERT embeddings from tweets and weighted the embeddings of sentiment words. Then, the CNN and BiLSTM were used to construct the prediction model. The TSDPAC effectively predicted rising stars, particularly in predicting the next season's rising stars with high recall and precision, and outperformed all other algorithms tested. Therefore, it provides team operators with reliable information. Future work will consider more text sources and weighting of other word types, like basketball jargon, as well as apply the TSDPAC to other sports. It is also possible to apply the TSDPAC to find rising stars in workplaces and academics.

References

[1] X. L. Li, C. S. Foo, K. L. Tew, S. K. Ng, Searching for rising stars in bibliography networks, *International Conference on Database Systems for Advanced Applications*, Brisbane Australia, 2009, pp. 288–292.

[2] P. Wijegunawardana, K. Mehrotra, C. Mohan, Finding rising stars in heterogeneous social networks, *International Conference on Tools with Artificial Intelligence*, San Jose, CA, USA, 2016, pp. 614–618.

[3] A. Daud, R. Abbasi, F. Muhammad, Finding rising stars in social networks, *International Conference on Database Systems for Advanced Applications*, Wuhan, China, 2013, pp. 13–24.

[4] A. Daud, M. Ahmad, M. S. I. Malik, D. Che, Using machine learning techniques for rising star prediction in co-author network, *Scientometrics*, Vol. 102, No. 2, pp. 1687–1711, February, 2015.

[5] A. Daud, N. R. Aljohani, R. A. Abbasi, Z. Rafique, T. Amjad, H. Dawood, K. H. Alyoubi, Finding rising stars in co-author networks via weighted mutual influence, *International Conference on World Wide Web Companion*, Perth, Australia, 2017, pp. 33–41.

[6] A. Dauda, F. Abbasc, T. Amjadc, A. A. Alshdadib, J. S. Alowibdi, Finding rising stars through hot topics detection, *Future Generation Computer Systems*, Vol. 115, pp. 798–813, February, 2021.

[7] J. Zhang, F. Xia, W. Wang, X. Bai, S. Yu, T. M. Bekele, Z. Peng, Cocarank: A collaboration caliber-based method for finding academic rising stars, *International Conference Companion on World Wide Web*, Montréal, Québec, Canada, 2016, pp. 395–400.

[8] L. Li, X. Wang, Q. Zhang, P. Lei, M. Ma, X. Chen, A quick and effective method for ranking authors in academic social network, in: J. Park, S. C. Chen, J. M. Gil, N. Yen (Eds.), *Lecture Notes in Electrical Engineering*, Vol. 308, Springer, Berlin, Heidelberg, 2014, pp. 179–185.

[9] G. Panagopoulos, G. Tsatsaronis, I. Varlamis, Detecting rising stars in dynamic collaborative networks, *Journal of Informetrics*, Vol. 11, No. 1, pp. 198–222, February, 2017.

[10] P. H. Chou, T. W. Chien, T. Y. Yang, Y. T. Yeh, W. Chou, C. H. Yeh, Predicting active NBA players most likely to be inducted into the basketball Hall of Famers using artificial neural networks in Microsoft Excel: Development and usability study, *International Journal of Environmental Research and Public Health*, Vol. 18, No. 8, Article No. 4256, April, 2021.

[11] A. N. Patton, M. Scott, N. Walker, A. Ottenwess, P. Power, A. Cherukumudi, P. Lucey, Predicting NBA Talent from Enormous Amounts of College Basketball Tracking Data, *Annual MIT Sloan Sport Analytics Conference*, Cambridge, MA, USA, 2021, pp. 1–14.

[12] H. Ahmad, A. Daud, L. Wang, H. Hong, H. Dawood, Y. Yang, Prediction of rising stars in the game of cricket, *IEEE Access*, Vol. 5, pp. 4104–4124, March, 2017.

[13] Z. Mahmood, A. Daud, R. A. Abbasi, Using machine learning techniques for rising star prediction in basketball, *Knowledge-Based Systems*, Vol. 211, Article No. 106506, January, 2021.

[14] J. D. Craig, N. Winchester, Predicting the National Football League potential of college quarterbacks, *European Journal of Operational Research*, Vol. 295, No. 2, pp. 733–743, December, 2021.

[15] J. Mulholland, S. T. Jensen, Predicting the draft and career success of tight ends in the National Football League, *Journal of Quantitative Analysis in Sports*, Vol. 10, No. 4, pp. 381–396, October, 2014.

[16] J. D. Pitts, B. Evans, Evidence on the importance of cognitive ability tests for NFL quarterbacks: what are the relationships among Wonderlic scores, draft positions and

NFL performance outcomes? *Applied Economics*, Vol. 50, No. 27, pp. 2957–2966, December, 2018.

[17] J. Rosen, A. Olbrecht, Data-driven drafting: Applying econometrics to employ quarterbacks, *Contemporary Economic Policy*, Vol. 38, No. 2, pp. 313–326, April, 2020.

[18] J. Wolfson, V. Addona, R. H. Schmicker, The quarterback prediction problem: Forecasting the performance of college quarterbacks selected in the NFL draft, *Journal of Quantitative Analysis in Sports*, Vol. 7, pp. 1–19, July, 2011.

[19] D. Berri, R. Simmons, Catching a draft: On the process of selecting quarterbacks in the National Football League amateur draft, *Journal of Productivity Analysis*, Vol. 35, No. 1, pp. 37–49, February, 2011.

[20] S. Rosen, The economics of superstars, *The American Economic Review*, Vol. 71, No. 5, pp. 845–858, December, 1981.

[21] C. Lucifora, R. Simmons, Superstar effects in sport: Evidence from Italian soccer, *Journal of Sports Economics*, Vol. 4, No. 1, pp. 35–55, February, 2003.

[22] Wikipedia, Player efficiency rating, https://en.wikipedia.org/wiki/Player_efficiency_rating, accessed on 2024/4/18.

[23] Wikipedia, Efficiency (basketball), https://en.wikipedia.org/wiki/Efficiency_(basketball), accessed on 2024/4/18.

[24] Spotrac, NBA contracts, https://www.spotrac.com/nba/contracts/limit-2000/, accessed on 2024/4/18.

[25] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint*, arXiv: 1810.04805v2, May, 2019.

[26] S. Baccianella, A. Esuli, F. Sebastiani, SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, *International Conference on Language Resources and Evaluation*, Valletta, Malta, 2010, pp. 17–23.

[27] Y. Li, A summary of neural network layers, https://medium.com/machine-learning-for-li/different-convolutional-layers-43dc146f4d0e, accessed on 2024/4/18.

[28] Convolution Neural Network, https://medium.com/%E9%9B%9E%E9%9B%9E%E8%88%87%E5%85%94%E5%85%94%E7%9A%84%E5%B7%A5%E7%A8%8B%E4%B8%96%E7%95%8C/%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-ml-note-convolution-neural-network-%E5%8D%B7%E7%A9%8D%E7%A5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF-bfa8566744e9, accessed on 2024/4/18.

[29] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, *arXiv preprint*, arXiv: 1409.0473, May, 2016.

Biographies

Ying-Ho Liu received a Ph.D. degree in Information Management from the National Taiwan University, Taiwan, in 2009. He is an associate professor in the Department of Information Management at National Dong Hwa University. His research interests include big data analysis, machine learning, data mining, and knowledge management.

Yu-Xiang Hong received a M.S. degree in Information Management from the National Dong Hwa University, Taiwan, in 2021. His research interests include big data analysis, machine learning, and data mining.