Predicting the Performance of Basketball Players Using Automated Personality Mining

Emergent Research Forum

Dominik Siemon
Technische Universität Braunschweig
d.siemon@tu-bs.de

Rangina Ahmad
Technische Universität Braunschweig
rangina.ahmad@tu-bs.de

Jan-Paul Huttner
Technische Universität Braunschweig
j-p.huttner@tu-bs.de

Susanne Robra-Bissantz
Technische Universität Braunschweig
s.robra-bissantz@tu-bs.de

Abstract

The Big Five personality traits, which provide a guide to the comprehensive assessment of individuals, have been found to be a valid predictor of behavior performance in the athletic context. However, traits have commonly been inferred via self-report questionnaires, which are criticized for reflecting only limited aspects of personality. Numerous studies have shown that human language reflects personality based on the frequency with which certain categories of words are used. Thus, we used an automated language-based measurement of personality and identified important traits of Basketball players - a combination of high conscientiousness and high agreeableness - that may be of significance for our prediction model of individual performance. Our findings are supported by prior research conducted on the prototypical "athletic personality profile". Predicting players' performance would not only further our understanding of competitive behavior in an athletic context but would also facilitate athlete selection and the design of training programs.

Keywords

Big Five Personality Traits, Automated Personality Mining, Basketball, Twitter

Introduction and Motivation

A plethora of studies have indicated that an individual’s personality influences many aspects of their behavior and sports performance (Allen et al. 2014; Piedmont et al. 1999; Whelan et al. 1991). Accordingly, increasing attention has been paid to the significant effects of personality on sports. The Big Five model, a generally accepted psychological taxonomy aimed to measure personality by formalizing five fundamental traits, is most commonly assessed through self-report measures. Such traditional questionnaires however, are criticized for reflecting only limited aspects of personality. Since a well-accepted theory of psychology is that human language reflects the emotional state and personality, based on the frequency with which certain categories of words are used (Boyd and Pennebaker 2017), we assume that an automated language-based measurement of personality would be a more reliable and valid way to identify traits. In order to see if there are any relations between an athlete’s personality traits and their sports performance, we chose to use the player data of the National Basketball Association (NBA). The NBA collects detailed player and game statistics and has additionally invested in an extensive new infrastructure (player tracking) to capture data. Moreover, the NBA players' personality traits were inferred by an automated personality mining tool via pre-existing digital data found on each player's Twitter account. In a next step, we want to build a statistical model that can provide a conditional expectation of basketball player performance based on automatically inferred personality traits. Predicting players’ performance would not only further our understanding of competitive behavior in an athletic context but would also facilitate athlete selection and the design of training programs.
Theoretical Background

Scouting in the NBA

As a multimillion-dollar sport business on a professional level, the NBA includes 30 Teams. Currently, each team in the NBA is worth at least $1 billion (Badenhausen 2018). The players' salaries also increased due to the higher revenues in the league. Hence, the pressure on the players to perform, increased as well (Abrams et al. 2008). Taking a closer look to the current scouting process in the NBA, there are several fundamental elements. Starting from traditional forms like on-site scouting or game film analysis, up to machine learning algorithms and statistical analysis to reveal insights about the players and correlations within their statistics (Atlas and Zhang 2008). Furthermore, Abrams et al. (2008) and Sandoval (2003) mentioned how much effort owners, general managers and the front-office put into selecting the right players. They focus on getting the good draft picks, hence, avoiding overpaid picks and an escalation of the players’ salaries. Therefore, the evaluation of the players’ potential and final selection of the team roster is a crucial process to all of the NBA teams. General managers tend to look out for the most talented players (Staw and Hoang 1995), but there are certainly more factors that influence a successful career of a rookie. Atlas and Zhang (2008) described four relevant domains that are considered by the scouts in their daily business: background (e.g. name, school, class, position played), characteristics (e.g. leadership, game knowledge, physical and mental ability), game skills (e.g. overall shooting, shot selection, decision making, free throw, three-point shooting, shot blocking, alert to the ball, aggressiveness) and personal skills (e.g. maturity, attitude, coach ability, feel for the game, agility, strength, endurance, poise). In this approach, we will focus on personal skills and characteristics and how they can be operationalized. Many of the factors (such as leadership or poise) that are comprised in these two domains, can be measured by the player’s personality traits.

Personality in Athletic Performance

Numerous studies have documented that personality has an impact on the behavior performance of athletes. Loosely defined as the construct that makes a human being’s behavior, thoughts and feelings (relatively) consistent but at the same time differentiates individuals from another, personality has been found to have predictive power (Allport 1961). According to Morgan (1986), the psychological construct has been shown to steadily explain 20% to 45% of the variance in sports performance (Piedmont et al. 1999). In this context, further studies have reported that personality traits are associated with interpersonal relationships as well as with long-term athletic success (Allen and Laborde 2014). The latter, being one of the factors influencing team performance, is considered to be an achievement-related outcome that can be predicted by the Big Five model of personality. This multifactorial model is the most widely used classification of personality that has formalized traits in order to measure a human being’s personality. More precisely, five fundamental traits or dimensions have been defined for a comprehensive assessment of individuals: Conscientiousness (a person’s tendency to display self-discipline and to act in an organized or thoughtful way), Openness (refers to the extent to which a person is open to experiencing a variety of activities and prefers novelty over convention), Agreeableness (a person’s tendency to be cooperative and compassionate towards others), Neuroticism (or also referred to as Emotional Range: the extent to which a person’s emotions are sensitive to the individual’s environment and the tendency to experience negative emotions) and Extraversion (refers to the extent to which people enjoy company and seek excitement and stimulation) (Costa and McCrae 2008; IBM Cloud Docs 2017). As an extension, the Revised NEO Personality Inventory (NEO-PI-R) assesses personality with six specific sub-categories or facets in each of the five dimensions (Costa Jr and McCrae 1995).

Earlier findings are consistent with recent studies that measures of personality can differentiate athletes from non-athletes or elite sportspeople from sport participants of different skill levels, respectively (Piedmont et al. 1999; Whelan et al. 1991). In order to identify the “athletic personality profile”, the Big Five model has been leveraged especially in the past two decades. Reviewed studies regarding correlations of the Big Five to athletic performance so far have provided the following perspectives: Athletes who score low on neuroticism or hold high levels of conscientiousness have more successful performance statistics throughout the course of a competitive season (Allen and Laborde 2014; Piedmont et al. 1999). It has further been found that sportspeople with high levels of agreeableness tend to have a more beneficial relationship with their teammates and coaches (Allen and Laborde 2014; Jackson et al. 2011). When it comes particularly to team athletes, extraversion has been found to be relevant as well, as sport
participants with higher levels of this trait respond to unsuccessful outcomes in a more positive manner (Allen et al. 2014; Allen and Laborde 2014; Taylor and Doria 1981). No salient assumptions have been made concerning the relationship between the dimension Openness (to experience) and athletic performance.

Based on these findings that (elite) athletes possess a unique combination of personality traits that are yet definable, personality testing would facilitate not only athlete selection by scouts and coaches but also the design of training programs (Whelan et al. 1991). With self-report questionnaires being a standard for measuring personality for several decades, this method however harbors a considerable problem: the degree to which people’s self-reported traits reflect who they truly are, has been put in doubt by many experts. An option to approach this problem is by utilizing language-based measurement of personality, as language constitutes another fundamental dimension of personality. In fact, language use has been scientifically proven to be unique, relatively reliable over time and internally consistent (Boyd and Pennebaker 2017). By scouring automatically vast amounts of User Generated Content (which is accessible on the internet), valuable personality data in the form of language can be used without the participants to complete traditional questionnaires (Boyd and Pennebaker 2017).

**Automated Personality Mining**

The worldwide accessibility to the internet and specifically Social Media platforms have been giving people around the world the possibility to connect to one another online. Such platforms come in many forms, including microblogs, social networks or photo-sharing platforms. Especially Twitter – a real-time microblogging online service where users can publish limited short messages that are visible to the public, called *tweets* – has been used by many public figures, such as NBA players. Since each person uses words in a different way and everyday words reflect a human being’s psychological state, the entries made on Twitter are, on a certain level, considered to be no different from everyday communication (Pennebaker 2011). Accordingly, tweets provide insights into an individual’s thinking and personality - thus, making Twitter a veritable goldmine of information. Aspects of personality can be predicted by the frequency with which certain categories of words are used as well as the variations in word usage (Golbeck et al. 2011; Yarkoni 2010). Personality mining systems are able to automatically infer personality traits from an individual’s text by applying linguistic analytics and personality models (Golbeck et al. 2011; Staiano et al. 2012; Yarkoni 2010). For the purposes of the present study, we used the IBM Watson Personality Insights (PI) service which utilizes personality theories and linguistic analysis to deduce traits from an individual’s written text.

**Methodology and Pre-analysis**

The data we used from the NBA, in order to predict individual basketball player performance, is considered to be highly accurate and precise (Eisenberg 2003) and is freely available online. With the PI service, the Big Five traits (i.e. the NEO-PI-R dimensions) can be assessed by the players’ Twitter timeline. The output of this analysis is a metric of all NEO-PI-R dimensions in a JSON file, where each dimension’s calculation results in a value between 0 and 1. For an accurate analysis of the personality traits, the Twitter user should have at least written 1600 words overall in his tweets. In order to predict a player’s basketball performance, we will perform a correlation and regression analysis with the data of 561 NBA players and their personality traits inferred via their Twitter feeds.

The basketball data was gathered using the website [www.basketball-reference.com](http://www.basketball-reference.com) where the statistics of all NBA basketball players are provided. We limited our data collection to native English-speaking players to have a consistent analysis for the personality analysis via Twitter. In a next step, we gathered the Twitter account-names of the collected NBA players and filtered the ones that do not have enough tweets for an accurate NEO-PI-R analysis. Additionally, we filtered out all the players’ retweets to avoid a bias in the results. Overall the data set contains 561 NEO-PI-R analyses of NBA players based on their analyzed tweets, and their individual game statistics consist of different 25 measures. In order to construct our regression model, we conducted a pre-analysis of the NEO-PI-R statistics of the NBA All-Star players of the last five years (2013-2017). Every year, the NBA hosts a basketball exhibition game with the best

---

1. [https://personality-insights-demo.ng.bluemix.net/](https://personality-insights-demo.ng.bluemix.net/)
players of the league, the NBA All-Star players, who are selected by a combination of fan, player, and media voting. In our pre-analysis we aim to find out, which personality traits are of the highest interest for our model. Therefore, we calculated the means of every personality dimension of the NBA players and calculated the difference of these dimensions compared to the rest of the NBA players. In addition, a Mann-Whitney U test was performed (due to non-normal distributed data) to calculate the significance of the differences. Overall 37 NBA All-star players were included in the calculation, whereas nine players that do not use Twitter or do have insufficient tweets and five who are not native English-speaker were excluded. Players can also play in consecutive All-Star games. The following table shows the three highest differences (+/-) in the personality traits of the All-Star players and all other players.

<table>
<thead>
<tr>
<th></th>
<th>Conscientiousness</th>
<th>Trust</th>
<th>Agreeableness</th>
<th>Intellect</th>
<th>Excitement-seeking</th>
<th>Self-consciousness</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-Stars</td>
<td>0.786</td>
<td>0.652</td>
<td>0.805</td>
<td>0.2133</td>
<td>0.5166</td>
<td>0.194</td>
</tr>
<tr>
<td>Other</td>
<td>0.677</td>
<td>0.557</td>
<td>0.686</td>
<td>0.2846</td>
<td>0.5599</td>
<td>0.1611</td>
</tr>
<tr>
<td>Difference</td>
<td>+0.109</td>
<td>+0.095</td>
<td>+0.119</td>
<td>-0.071</td>
<td>-0.043</td>
<td>-0.042</td>
</tr>
<tr>
<td>p-value</td>
<td>.002*</td>
<td>.034*</td>
<td>.002*</td>
<td>.069</td>
<td>.111</td>
<td>.211</td>
</tr>
</tbody>
</table>

*95% CI

Table 1. Big Five Difference of Basketball Players

The pre-analysis shows that the traits conscientiousness and agreeableness have the biggest significant positive difference, meaning that NBA All-star players have a higher level of conscientiousness and agreeableness than all other players, which aligns with prior studies. Trust, a facet of the dimension agreeableness, has the third highest significant positive difference between the All-stars and the other players. Extraversion has also a significant high positive difference (+0.070, p-value = .00253*), which is consistent with prior studies. Other positive differences are the sub-dimensions achievement-striving and activity level. The highest negative difference has intellect (not significant), a facet of openness. Followed by excitement-seeking (a facet of extraversion), where the difference is not significant as well, and self-consciousness (a sub-category of neuroticism). In a next step, we will compute a set of correlation analysis to discover and measure the strength and direction of the relationship between any NEO-PI-R dimension and any basketball performance of the players. With these findings and the findings from prior studies and theory, we will construct our model, which will be further evaluated with a qualitative analysis, where scouts and coaches of different basketball teams will report on their long experience and interaction with many professional basketball players. Ratings on their assessments of important personality traits for specific basketball positions will be taken into account, when refining our model. Subsequently our model will be tested with a regression analysis.

Discussion and Outlook

In our paper, we collected data from the social microblogging service Twitter by inferring the personality traits of NBA players via their Twitter accounts. The values of the personality traits will be further used as the independent variables of our model. In doing so, we however assume that each player’s Twitter account is verified, meaning that they use the service independently and write their own tweets. In a first step, we identified important personality traits that might be of high interest for our model, by calculating the differences between NBA All-star players and regular NBA players. Our identified traits are additionally supported by prior studies on the specific personality traits of successful athletes, that is a combination of high conscientiousness and high agreeableness. In a next step, we will conduct expert interviews, with scouts, coaches and team managers on their perception of important personality traits of particularly good players. A correlation analysis between the player’s personality traits and their basketball performance will further reveal valuable variables for our model. Subsequently our constructed model will be tested with a regression analysis. With the results of our prediction model, scouts and coaches will have another valuable tool for the selection of relevant basketball players without the need of an intensive self-report personality testing. Even without the use of Twitter, other sources of text can be also used for automated personality mining, and in the absence of researcher intervention. Furthering the understanding of competitive behavior in an athletic context facilitates athlete selection. Our model will not only provide valuable insights for research, but can be also adapted to other sport disciplines.
REFERENCES


