

# Text Mining in Trainingswissenschaft und Leistungssport

Otto Kolbinger

Technische Universität München

## Einleitung

In Sportwissenschaft und Sportpraxis müssen vielfach Informationen verarbeitet und analysiert werden, die ausschließlich in Textform vorliegen. Neben traditionellen Methoden der Inhaltsanalyse, wie der manuellen qualitativen Inhaltsanalyse, werden dafür immer häufiger sogenannte Text Mining Verfahren eingesetzt. Ziel ist es, ähnlich wie beim Data Mining, automatisch Wissen aus großen Textdatensätzen abzuleiten (Cai & Sun, 2009). Dies kann zu beschreibenden Zwecken erfolgen, z. B. um die Themen oder Stimmung (Sentiment) von Texten und Textabschnitten zu identifizieren, aber auch um Verhalten oder Ereignisse vorherzusagen. Der methodische Werkzeugkasten reicht dabei von lexikonbasierten Ansätzen, bei denen Textkörper auf bestimmte Wörter oder Wortgruppen durchsucht werden, bis hin zu überwachten und unüberwachten Verfahren maschinellen Lernens. In den Bereichen der Sportwissenschaft, in denen Forscher typischerweise mit Textdaten arbeiten, werden Text-Mining Verfahren bereits angewendet, zum Beispiel in der Sportsoziologie (u. a. zu Reaktionen auf politische Statements von Athleten: Frederick et al., 2022; Schmidt et al., 2019). In traditionell von numerischen Daten dominierten Bereichen, wie in der Trainingswissenschaft, wird Text Mining bisher kaum – wenn überhaupt – angewendet. In dieser Post-Doc-Vorlesung soll deshalb diskutiert werden, wie Text Mining in der Trainingswissenschaft und im Leistungssport im Allgemeinen eingesetzt werden kann. Konkret wird die Anwendung von Text Mining-Techniken in zwei Bereichen vertieft vorgestellt: Erstens wird am Beispiel des Videoassistenten im Fußball (VAR) demonstriert wie Social-Media-Posts genutzt werden können, um die Auswirkungen von Regeländerungen auf das Meinungsbild von Interessengruppen zu untersuchen. Zweitens wird anhand von zwei Beiträgen aufgezeigt, welchen Mehrwert Text Mining Verfahren im Talentscouting bieten können. Aufbauend auf der Vorstellung dieser beiden Bereiche werden weitere Einsatzmöglichkeiten aufgezeigt und abschließend diskutiert, wie Text Mining in weiteren Feldern der Sportwissenschaft genutzt werden kann.

## Anwendungsbeispiele

### *Evaluation von Regeländerungen*

Um Leistungssteigerungen zu ermöglichen und anzuregen, können Sportverbände die Rahmenbedingungen anpassen, unter denen sportliche Leistung erbracht werden muss. Dies kann erreicht werden, indem neue Ausrüstung zugelassen wird, Regeln geändert oder Schiedsrichter mit technischen Hilfsmitteln unterstützt werden. Solche Interventionen haben jedoch oft unbeabsichtigte Nebeneffekte und müssen umfassend evaluiert werden, auch bezüglich des Einflusses auf das Meinungsbild von Interessensgruppen. Kolbinger und Knopp (2020) werteten über 600.000 Tweets von Fans der englischen Premier League aus, um deren Reaktionen auf VAR-Einsätze zu untersuchen. Dafür trainierten die Autoren zwei automatische Text-Klassifikatoren, um Posts zu identifizieren, die sich auf den VAR beziehen und um zu bewerten, ob ein Tweet positive, neutrale oder negative Emotionen ausdrückt. So konnten sie zeigen, dass Tweets über den VAR mit 76,2% vornehmlich negativ

waren, was sich nicht mit dem Stimmungsbild für andere Posts während Fußballspielen deckte (positiv: 39,4%; negativ: 31,3%). Weiter konnten die Autoren zeigen, dass die durchschnittliche Stimmung aller Tweets nach einem VAR-Vorfall signifikant und dauerhaft abfiel. Während negative Reaktionen auf den VAR aufgrund anekdotischer Erfahrungen vielleicht bereits erwartet werden konnten, wurden mit dieser Studie – durch die Anwendung von Text Mining-Verfahren – die Auswirkungen erstmals empirisch belegt.

## **Talentscouting**

„Kobe weiß wie man Basketball spielt – und was es braucht um zu gewinnen“. Dieser Auszug aus einem Scoutingbericht über den viel zu früh verstorbenen Kobe Bryant erwies sich als sehr zutreffend (Sumsky, 2020). Interessanterweise wird sich dabei auf Leistungsmerkmale bezogen, die (noch) nicht über numerische Daten abbildbar sind. Derartige Scoutingberichte liegen in unzähliger Anzahl in unzähligen Vereinen und Verbänden vor. Anhand von zwei Studien soll gezeigt werden, wie diese mit Hilfe von Text Mining Verfahren ökonomisch genutzt werden können. Seppa et al. (2016) verknüpften für die Auswertung von Scoutingberichten im Eishockey Sentiment-Analysen mit einer lexikonbasierten Kategorisierung. Jeder Abschnitt eines Scoutingberichts wurde auf Basis von Schlagwörtern oder Wortketten einer Kategorie zugeordnet, z. B. „Einsatz“, wenn Wörter wie eben „Einsatz“ oder aber auch „faul“ oder Wortketten wie „Kampf annehmen“ auftraten. In Verbindung mit Sentiment-Analysen entsprechender Abschnitte, wurden die Spieler so jeweils für diese Kategorien eingeordnet, beispielsweise als Spieler mit „mangelndem Einsatz“ oder „gutem Puck-Gefühl“. Mit diesen automatisiert generierten Einordnungen als Prädiktoren konnten besser die Assist- und Tor-Raten der Spieler im Profibereich vorhergesagt werden, als auf Basis der erzielten Assists und Tore im Jugendbereich. Noch besser waren die Vorhersagen, wenn beides kombiniert wurde. Auch wenn einschränkend gesagt werden muss, dass jeweils nur eine niedrige Vorhersagequalität erreicht wurde, ist das Muster, dass die Berichte die Vorhersagequalität verbessern konnten, sehr interessant. In einer weiteren Studie zur Vorhersage der Leistung in der NBA entwickelte Maymin (2021) ein Modell, in welchem unter anderem Matchstatistiken und Scoutingberichte aus dem Collegebereich als Prädiktoren dienten. Da ich Maymin vor allem auf den Vergleich seines Modells mit der Draftperformance von NBA-Teams fokussierte, geht er nicht im Detail auf die Beiträge der einzelnen Prädiktoren ein. Allerdings zeigt seine Übersichtsgrafik, dass Inhalt, Umfang und Sentiment von einzelnen Textkategorien einen ähnlich hohen Vorhersagebeitrag leisteten wie Matchstatistiken. Der Forschungsstand zur Nutzbarkeit von Textdaten für die Talentediagnostik ist noch keinesfalls zufriedenstellend, zumal sich die wenigen Studien bisher auf kommerzielle Scoutingberichte und ausschließlich auf das nordamerikanische Sportsystem beziehen. Allerdings liefern diese Studien durchaus vielversprechende Anhaltspunkte und deuten darauf hin, dass die Textdaten Informationen abbilden, die in anderen Daten (noch) nicht vorhanden sind.

## **Literatur**

- Cai, Y., & Sun, J. (2009). Text mining. In Liu, L., & Özsu, M. (Eds.), *Encyclopedia of Database Systems* (p. 131). Springer.
- Frederick, E., Pegoraro, A., & Schmidt, S. (2022). „I’m not going to the f\*\*\*ing White House“: Twitter Users React to Donald Trump and Megan Rapinoe. *Communication & Sport*, 10(6), 1210–1228.

- Kolbinger, O., & Knopp, M. (2020). Video Kills the Sentiment—Exploring Fans' Reception of the Video Assistant Referee in the English Premier League Using Twitter Data. *PLoS ONE*, *15*(12), e0242728. doi.org/10.1371/journal.pone.0242728
- Maymin, P. (2021). Using scouting reports text to predict NCAA → NBA performance. *Journal of Business Analytics*, *4*(1), 40-54. <https://doi.org/10.1080/2573234X.2021.1873077>
- Schmidt, S., Frederick, E., Pegoraro, A., & Spencer, T. (2019). An Analysis of Colin Kaepernick, Megan Rapinoe, and the National Anthem Protests. *Communication & Sport*, *7*(5), 653–677.
- Seppa, T., Schuckers, M., & Rovito, M. (2017). *Text Mining of Scouting Reports as a Novel Data Source for Improving NHL Draft Analytics*. In Ottawa Hockey Analytics Conference (pp. 1–11).
- Sumsky, A. (2. Juli 2020). *Kobe Bryant's Scouting Report is Worth the Read*. *Basketball forever*. <https://basketballfor-ever.com/2020/07/02/kobe-bryants-scouting-report-worth-read>.