Birds of a Feather: Using Word Embeddings to Evaluate Athlete Brand Similarity

Bradley Baker, University of Massachusetts Amherst

Marketing - Branding (Professional Sport)

20-minute oral presentation (including questions) Mode: Asynchronous
Session: Live Q&A for Marketing 3 (Asynchronous) Saturday June 5, 2021, 3:50 PM - 4:50 PM

According to the precepts of Human Brand Theory (Thomson, 2006), consumers form attachments to well-known personas that are attractive and those attachments form the basis for positive marketing results. Strong personal brands are among the most valuable assets of professional athletes (Su et al., 2020) as endorsement income can dwarf financial returns directly from sport. For example, in 2019, Roger Federer made $100 million from marketing deals with Uniqlo, Credit Suisse, Rolex, and other partners, compared to $6.5 million in winnings as a top-ranked tennis player (Badenhausen, 2020). Thus, athletes face strong incentives to establish monetizable personal brands (Napoli et al., 2016). Athlete brand equity derives from consumers’ brand image and associations (Arai et al., 2013, 2014; Ross, 2006).

Understanding how consumers think about athlete brands is essential for partnerships between corporate brands and athlete endorsers. While established scales (e.g., Arai et al., 2014; Gladden & Funk, 2002; Ross et al., 2006) can be used to evaluate brand associations, recent advances in natural language processing (NLP) promise an alternative approach to capturing the latent representation of athlete brands in consumers’ minds based on consumers’ own words. Word embeddings are a statistical approach to language modeling that captures underlying semantic meaning based on the context in which words appear in large text corpora (Mikolov et al., 2013a, 2013b; Pennington et al., 2014). The context in which words are used provides unique insight into mental representations of those words.

Data for the current study come from User Generated Content (UGC) collected via Twitter’s Application Programming Interface (API) and comprise tweets mentioning professional basketball players (NBA; WNBA), including athletes just entering the leagues via 2020 drafts. Word embeddings will be used to project the sparse, high-dimensional set of word features in this text corpus into a dense, 100-dimensional vector representation. These embeddings will be used both to derive key associations with athlete brands based on point-wise mutual information and singular value decomposition (Hvitfeldt & Silge, 2020) and to calculate similarity scores between pairs of athletes in semantic space. Latent semantic similarity captures how consumers talk about athletes and the mental associations of athlete brands in consumers’ minds. Thus, it is possible to identify athletes who share similar associations and may be comparable brands. Given exclusivity arrangements that limit firms’ ability to partner with established athletes, insight into athlete brand comparability supports identifying substitutable endorsers. Further, algebraic transformations of vectorized word representations can identify complex semantic relationships such as analogies (e.g., A’ja Wilson:WNBA::???:NBA). Results will provide knowledge regarding consumer cognitions about athlete brands and relationships between athletes based on latent associations in consumers’ minds. This contributes to theoretical understanding of branding and brand associations. Limitations of the current approach will also be discussed. For example, models based on human-generated text corpora are prone to social biases such as gender-based stereotypes (Bolukbasi et al., 2016; Caliskan et al., 2017) and such bias can negatively impact downstream uses (Zhao et al., 2018).