Self-Scouting the Scout: Using Text Analytics to Improve Scout Grades in American Football

Scouting in American football is an inexact science with countless first round draft busts and Hall of Famers selected in the later rounds or not at all. Yet due to the importance of identifying the best prospects, player personnel departments continue to exhaust substantial resources through their scouting departments, searching for the slightest competitive edge. Meanwhile, researchers independently develop predictive models that use the data available to them, also with varying degrees of success. These isolated approaches will inevitably lead to disagreement between player rankings, leaving the general manager to make a judgment call on when to adhere to one or the other.

We propose an alternative option that integrates the wealth of knowledge that scouts' possess into the statistical analysis. Our approach uses natural language processing to extract useful predictors from a written scouting report. When used as features in gradient boosting decision trees, a scout's written opinion can be shown to improve predicted success by over 14% in most position groups when compared to just the scout's overall grade. Since we use text as features, we are also able to provide insight into characteristics that are important at each position and inappropriately valued in an individual scout's assessment. Also, unlike most work in this area that focuses on game level statistics to measure success, we use the average yearly value of the second signed contract, properly scaled, which is a more representative measure of success.