REFINING EXPERT OPINION: AN EXAMINATION OF MACHINE LEARNING & NLP TECHNIQUES APPLIED TO SCOUTING DATA

Gordon Irving
MSDS 453, Section 55
November 22, 2020

Keywords:

NLP, Machine Learning TF-IDF, Word2Vec, Expert Opinion, Hockey

Abstract

The areas of human life in which expert opinion is critical to decision making have long remained insulated due to the opacity of domain specific lingo and the general difficulty of translating natural language into terms understandable by machine learning methods. However, advances in natural language processing (NLP) have opened an analytical seam into these subtle and flexible spaces of domain specific expert opinion. This paper explores that seam along one front – by applying TF-IDF Word2Vec and Doc2Vec representation methods to a small corpus of more than 700 scouting reports of NHL draft eligible players. We begin by applying various classification methods (Random Forests, Gradient Boosting, Neural Networks) to vectorized representations of these documents with the goal of classifying documents according to the human ontological structures along which they are organized. We then explore the data further by attempting to map from vectorized representations of the words themselves to observed outcomes in player performance.

Introduction & Problem Statement

With the proliferation of easily accessible natural language processing tools in recent years, organizations which rely upon expert opinion for decision making now face the prospect of a tremendous opportunity – that of applying more rigorous methods of interpreting the recommendations of experts and potentially even improving upon them. Vital to the success of this endeavor lie two core assumptions. First, that the recommendations of experts hold quantifiable value, and second, that the language used to convey and justify these recommendations provide additional information beyond the surface level assessment. If these two conditions hold, then the possibility exists for rigorous analytical methods to improve upon the surface level recommendations and assessments of experts by identifying clues that recommendations are unwittingly biased, too modest, or too extreme. The possible area of effect here encompasses all forms of written reporting ranging from damage and insurance estimates to police reports, judgements and beyond.

This paper explores one specific domain ripe analysis: the scouting and evaluation of NHL draft eligible players. First, we establish that simple NLP techniques and machine learning methods can reach a high level of correspondence with the human ontologies which structurally underlie and organize this domain specific topic. This is accomplished via the classification of player reports by position, which is a critical aspect along which scouts evaluate overall player ability. Our results demonstrate that basic NLP and machine learning methods can reach a high level of accuracy (97%) in player classification.

Next, we make a first pass attempt at leveraging these same techniques for the purpose of improving scout recommendations. This is accomplished by first adopting a one-size-fits-all metric for evaluating player performance (GSVA, or Game Score Value Added is the metric of choice) and then training various algorithms to predict player performance based on the language data contained in each report. To compare the performance of algorithms trained on the language data, we create a benchmark measure based on the original ranking of players provided by the scout. This benchmark comes in the form of a simple linear regression model which relies on the numerical ranking of each player according to the expert who has written these reports. The underlying assumption here is that the root mean squared error of this linear regression model will provide a reasonable benchmark against which our more complex models can compete with the goal of predicting player performance according to GSVA based on the natural language data contained in each report. This paper concludes with recommendations for future work.

Literature Review

Tremendous advancements have been made in natural language processing methods since the start of the last decade. Beginning around that time, experiments in clustering movies by type using vectorized representations of descriptions from IMDB began to show serious promise for the future of these methods (Maas et al. 2011). Soon after, researchers at Google developed Word2Vec and Doc2Vec methods for use on extraordinarily large corpora (Mikolov et al. 2013). More recently, the refinement of document clustering methods has led to improvement in document searches (Buatoom, Kongprawechnon, and Theeramunkong, 2020), and document clustering methods have merged with topic modeling techniques to yield more powerful and scalable topic modeling frameworks (Brena and Ramirez 2019).

The ubiquity of text data in the modern world means that the promise of these techniques does not end with the refinement of search technology. Today, researchers are leveraging both traditional TF-IDF methods as well as the more cutting edge Word2Vec and Doc2Vec methods for the purposes of automating classification tasks and improving human-expert recommendations in a variety of fields. Classification tasks range from the large scale social oriented tasks, such as classifying tweets and comments (Bilgin 2020) as well as identifying and blocking pornographic websites (Chen et al. 2020), to the small scale business oriented tasks, such as automating the translation of software requirements into the classes of functional and non-functional requirements (Canedo and Mendes, 2020). Research oriented toward improving upon expert opinion has recently touched such areas as written examination questions at universities (Mohammed and Omar 2020), expert assessments of risk for catastrophic failure aboard open ocean tankers (Hoang, 2017), and detection of hazardous conditions in heavy construction (Tixier et al. 2017).

Concurrent with this research, others have continued to advance the neural network architectures which underly and empower NLP classification, clustering, and sentiment analysis. Recent work on the hierarchical encoder attention-based models (HEA) for example, has led to valuable advancement in the ability of NLP techniques to discern high and low-quality information online (Kinkead, Ahmed, and Krauthammer 2020). Elsewhere, emerging architectures such as the Interactive Dual Attention Network (IDAN) has led to advancements in sentiment analysis and classification (Zhu, Zheng, and Tang 2020).

Data Description & Preparation

Our corpus consists of 713 ranked scouting reports of NHL draft eligible players between the ages of 17-20 years. All reports used in this analysis were written by public scout and writer Corey Pronman between 2013 and 2020. Word count for each report can range from as few as 60 words to as many as 300. In addition to the text report data, our dataset also includes the name, position, draft year, and ranking for each player. Where applicable, statistics regarding player performance at the NHL level (games played, points, and GSVA) is also included. Due to a low quantity of observations (less than 10), the position of goaltender was excluded from the reports list, leaving only offensive and defensive position players. Because many players listed in these reports failed to make the NHL following their draft seasons, imputation for GSVA is necessary. For players who failed to make the NHL, we have imputed GSVA scores between -1 and 0, which indicates replacement level or lower. Our GSVA statistic very loosely approximates a normal distribution.

Other data cleaning procedures included the following; removal of punctuation and non-alphabetic characters using the RegEx module, the removal of player names from reports, standardizing the case of all characters to lower case, filtering out words identified in NLTK as stop words (such as ‘which’ and ‘these’). Word stemming was not performed on this corpus due to experimental findings which will be discussed later suggesting that it may be detrimental to the integrity of the corpus. After initial text data cleaning, two separate lists were created for use in TFIDF and Doc2Vec analysis, each containing the remaining words contained for every individual document.

Methods and Modeling

In the interest of establishing a baseline level of correspondence between human level and machine understanding of the documents, our initial experiment tests the extent to which TF-IDF and Doc2Vec text representations can be leveraged by various machine learning methods to classify players according to position. Here we seek answers only in the simplest terms, classifying along two dimensions: offensive position players and defensive position players.

To accomplish this, the ‘Center’ and ‘Wing’ positions were grouped into a ‘Forward’ category, while ‘left defense’ and ‘right defense’ were grouped into a ‘Defense’ category. In each of these classification experiments, our models trained on 570 examples and tested on the remaining 143. In the TF-IDF version of this experiment we use an n-gram range of one, which yields a total of 3,670 words, making our data frame 713 rows by 3,670 columns wide. In the Doc2Vec version of this experiment, we used a vector size of 500, a window of 10, and a minimum word count of five.

Having established correspondence between our classification methods and the human ontology which underlies all player evaluation, we make a first pass attempt to predict player performance at the NHL level based on the language data provided in each report with the objective of improving upon the raw player rankings provided by the author of the reports. For this experiment we pare down on the size of the corpus, using only reports from the period of 2013 – 2017. This is done to mitigate the effect of year over year differences which tend to penalize the youngest players in this sample who simply have not yet reached their potential.

Improvement here is measured against a benchmark which is admittedly imperfect. We use a linear regression model as a surrogate for the 'goodness' of the rankings provided by the scout. This regression model considers only two variables to establish the baseline value for the recommendations of the scout: the raw player rankings associated with each report, and the year in which the report was written. The linear regression model then predicts player GSVA and we measure error using root mean squared error (RMSE). Finally, the performance of this baseline model is compared with more sophisticated models which make use of the language data to make predictions.

Classification Results

The initial experiment, which focused on classifying the position of each player as offensive or defensive based solely on the natural language data contained in each scouting report, showed promising results when applied to TF-IDF scores. Table 1 below summarizes the results of four models trained on a random sample of 570 reports and tested on 143.

Table 1. TF-IDF Classification Results

Logistic Regression					Decision Tree				
	Precision	Recall	F1-Score	Support		Precision	Recall	F1-Score	Support
Defensive Position	1	0.46	0.63	48	**Defensive Position**	0.93	0.85	0.89	48
Offense Position	0.79	1	0.88	95	**Offense Position**	0.93	0.97	0.95	95
Macro Average	0.89	0.73	0.75	143	**Macro Average**	0.93	0.91	0.92	143
Weighted Average	0.86	0.82	0.8	143	**Weighted Average**	0.93	0.93	0.93	143
Accuracy			0.82	143	**Accuracy**			0.93	143
Random Forests					**Neural Network**				
	Precision	Recall	F1-Score	Support		Precision	Recall	F1-Score	Support
Defensive Position	0.95	0.75	0.84	48	**Defensive Position**	0.96	0.94	0.95	48
Offense Position	0.89	0.98	0.93	95	**Offense Position**	0.97	0.98	0.97	95
Macro Average	0.92	0.86	0.88	143	**Macro Average**	0.96	0.96	0.96	143
Weighted Average	0.91	0.9	0.9	143	**Weighted Average**	0.96	0.97	0.96	143
Accuracy			0.9	143	**Accuracy**			0.97	143

Although tree based models and logistic regression models show comparable (or even superior) recall in identifying the offensive position players, reaching as high as 100% positive identification for this class, they do so at the expense of precision and in some cases do very poorly at identifying defensive position players. One contributing factor which explains this is the general imbalance of classes, with offensive position players outnumbering defensive position players by a two to one ratio. The ability of the neural network to overcome this problem is evident.

As can be seen in figure 2 below, classification based on Doc2Vec text representation did not produce particularly strong results.

Table 2. Doc2Vec Classification Results

Logistic Regression					Decision Tree				
	Precision	Recall	F1-Score	Support		Precision	Recall	F1-Score	Support
Defensive Position	0	0	0	48	Defensive Position	0.43	0.19	0.26	48
Offense Position	0.66	1	0.8	95	Offense Position	0.68	0.87	0.76	95
Macro Average	0.33	0.5	0.4	143	Macro Average	0.55	0.53	0.51	143
Weighted Average	0.44	0.66	0.53	143	Weighted Average	0.6	0.64	0.6	143
Accuracy			0.66	143	Accuracy			0.64	143
Gradient Boosting					**Neural Network**				
	Precision	Recall	F1-Score	Support		Precision	Recall	F1-Score	Support
Defensive Position	0.62	0.17	0.26	48	Defensive Position	0.33	0.6	0.42	48
Offense Position	0.69	0.95	0.8	95	Offense Position	0.65	0.37	0.47	95
Macro Average	0.65	0.56	0.53	143	Macro Average	0.49	0.49	0.45	143
Weighted Average	0.67	0.69	0.62	143	Weighted Average	0.54	0.45	0.45	143
Accuracy			0.69	143	Accuracy			0.45	143

In this instance we have included a gradient boosting model in place of a random forests model since the gradient boosting approach appears to be better, achieving respectable overall recall and the highest overall classification accuracy of any model. As was the case with classification using TF-IDF data, most models tended to lean toward classification of reports as belonging to offensive position players due to the imbalance of classes. However, unlike with the experiment in TF-IDF data, the neural network proved to be the worst model for this data as it severely over-fit the noisy Doc2Vec training data and was the only model to achieve less than 50% classification success as a result.

With only 713 documents available, both Doc2Vec and Word2Vec techniques should be at a similar disadvantage with this dataset compared with the more simplistic TF-IDF. To further explore this, we establish a Word2Vec model and explore the terms this model believes to be similar. This time using a minimum word count of five, a vector size of 100, and a window of 10, we examine a few of the ways in which our more complex vectorization models are at a disadvantage by looking more closely at what this Word2Vec model knows about our corpus. This exploration shows only the beginnings of a reasonable lexical command with several glaring issues present. For example, the word "two-way" which in this corpus of documents is used to characterize players who are well rounded both offensively and defensively, is in this model considered to be most similar to the following words shown in table 3 below.

Table 3. Word Similarity to "two-way"

Word/Token	**Similarity Score**
6-foot-4	0.99
mobile	0.99
big	0.99
highly	0.99
strong	0.99
fantastic	0.99
solid	0.99
pretty	0.99
tenacious	0.99

Some of these words are reasonable matches, such as ‘solid,’ which often appears as a descriptor of players who fit the ‘two-way’ type. However, ‘6-foot-4’ and ‘pretty’ are poor matches. Another well-known feature of the Word2Vec model is its ability to model the relationships between words in such a way as to infer new word matches based on the adding or subtracting of word vectors from one another. All experiments with this in our corpus have shown poor results. Example: ‘passer’ plus ‘creative’ should produce something like ‘playmaker.’ In this instance however it produces likely matches such as ‘rush,’ ‘handler,’ ‘coordinated,’ ‘zone,’ and ‘offensive.’ Although some of these likely matches are reasonable, none are optimal. Apparent in all the above analysis is the fact that more complex vectorized representations are not appropriate for this small corpus.

Regression Results

The baseline regression model we develop here considers only the numerical ranking of the player and the year of the draft. The idea behind this construction is that the numerical ranking itself represents the scouting recommendation we seek to match or improve upon using text data. Additionally, by including the year in which the draft took place, we offer the model an opportunity to adjust for developmental differences (e.g. differences in whether a player has had enough time to reach his true potential). This regression model is trained only on that simple data and is asked then to predict GSVA (our best estimate for overall performance). The resulting error (RMSE) for all predictions is our benchmark against which we may measure the relative success of other models.

In contrast to this benchmark, we test out models which start out with *only* the language data, and then create blended models which have access to information from both the benchmark and the language-based models. Results from these regression experiments can be seen in table 4 below.

Table 4. Player Performance Prediction Models

Model	Data	RMSE
Linear Regression (Baseline)	Rank Only	0.83
Random Forests	Language Only	1.07
Gradient Boosting	Language Only	1.1
Neural Network	Language Only	1.06
Random Forests	Rank & Language	0.77
Gradient Boosting	Rank & Language	0.81
Neural Network	Rank & Language	0.92

The linear regression model representing our baseline, or how uniformly the numerical rankings translate into performance, shows a root mean squared error of 0.83. As may be expected, predictions which only use TF-IDF scores and do not account for numerical rank perform somewhat worse. However, they still manage to provide predictions which are sensible given the information they have to work with. The blended, tree based models, which use all available data (raw rankings as well as TF-IDF scores), apparently reduce error and improve upon the baseline in this experiment, but there are several reasons to be skeptical that these results reflect true, sustainable improvement, which we will discuss in detail in the next section.

Analysis & Interpretation - Classification Experiment

The failure of our Doc2Vec player position classifier to meet its full potential on our relatively small corpus is evident and requires very little explanation. Here the Doc2Vec method simply overpowers our corpus by seeking out more complexity than this narrow set of documents can offer. There likely are not enough connections here to establish the level of term contextuality that this method seeks out. On the other hand, the extent to which our model interpretations of the TF-IDF data aligns with the human ontological classification of players by position is impressive, especially considering that many of the scouting reports contained in this corpus do not make the position of the player explicit and may be ambiguous even to human readers. Take the following report from a 2017 ranking, for example:

> There are a whole host of reasons to be skeptical about Smirnov. He's small and not overly physical, he's a mediocre skater at best and Penn State's schedule was tame before the divisional games began. However, at the end of the day, Smirnov was one of the top scorers and playmakers in college hockey as a teenage freshman. His skill level and work ethic are high-end, and he has shown the ability to be a game-breaker at the college level. He can make some risky mistakes, but in general his great creativity and vision allow him to make tough plays seem routine. NHL scouts are tentative on him in part because of how slow he is for a smaller guy. But when considering all of his attributes, there is value here. (Pronman 2017)

A closer examination of the terms determined to be predictive of position by our tree-based models reveals that success is dependent on them finding subtle and domain specific codewords embedded in the text; terms which are used by *this particular scout* in describing player attributes at each position. For instance, the following words shown in table 5 were critical in guiding our model to correctly classifying the above report as belonging to a forward position player.

Table 5. Key Terms for Classification

Word/Token	**Regression Coefficient**
scorer	0.9
skill	0.73
ethic	0.69
high-end	0.43
vision	0.41

Although ‘scorer’ is unsurprisingly a good indicator that this player is a forward, the inclusion of terms like ‘skill’ and ‘ethic’ (which in this corpus is always used in the word pair ‘work ethic’), show a keener insight into the subtle usages of language. For reasons that are not entirely obvious, this author does not typically write about work-ethic when assessing defensive players, nor does he often use the word ‘skill’ in describing their attributes, even though there is surely a high degree of skill required for them to succeed in their position. When discussing this aspect of performance among defensive position players, he will tend to remark about them being ‘skillful’ in regard to a *particular* aspect of performance rather than make a blanket comment about their overall ‘skill’ as a flat attribute. In experimentation, subtle differences such as these are what made word stemming less desirable as a pre-processing step. Insights like this one betray elusive yet profound insights into the way this author thinks about player evaluation.

Perhaps even more fruitful for analysis is to look at the types of reports which our models mis-categorize. For the neural network, there are very few in that category. However, one such is quoted in full below:

> With Carlsson, it's more about the scouting projection as opposed to the statistical production, because there wasn't much of the latter. To put it simply: He's great defensively. He is always coming over the boards for any tough defensive zone draw, and

to start most penalty kills. He's very smart in his own end, closing on his checks well, getting in lanes and disrupting offensive setups. His puck skills are not the greatest, but he can jump up into the play here and there. (Pronman 2015)

Table 6 below sums up some of the key terms for this report along with coefficients borrowed from the logistic regression model.

Table 6. Key Terms for Classification

Word/Token	Regression Coefficient
Skills	0.62
Production	0.36
Penalty	0.36
Boards	0.19
Defensive	-0.73
Defensively	-1.25

Although there are tells here which guide the model toward classifying this player as a defensive player, such as the words “defensive” and “defensively,” these words are not exclusively used to describe players of the defense class (unlike the very similar term “defenseman” which, unsurprisingly, indicates a player of the defensive class more strongly than any other term). Additionally, there are a few terms, such as “skills” and “production,” which typically appear in descriptions of offensive position players. In this case, our models all misclassified this report as belonging to an offensive position player rather than a defensive one. For the neural network model, this was one of only three defensive player reports to be mis-categorized and one of only five misclassifications (out of 143) total. Worth noting is the fact that the very words which most mislead our classifiers were also used here in the negative case.

The author intended to express a lack of skill and production, but due to the simplicity of our single n-gram TF-IDF model, this nuance was missed.

Analysis & Interpretation – Regression Experiment

As has already been noted, there are several reasons to be cautious regarding the apparent improvement of our Random Forests and Gradient Boosting regression models over the raw rankings provided by the scout. In this section we will present a few of the reasons why this apparent improvement should be viewed skeptically.

First, the difference is marginal, and it should be noted that these are the results from only a single experiment. Running this same experiment with different random assignments of reports to training and test sets alters performance of the models slightly from trial to trial. Additionally, when looking closely at the words identified by these models as predictive of player performance, many of them are clearly poor predictors which just happen to align with player performance due to random variation within our small dataset. Table 7 below summarizes various words selected as predictive by the individual trees within the larger Random Forests model as well as the frequency of these words across documents.

Although some of these words seem to hold potential as being truly predictive, such as “forechecker,” and “highend,” or “average,” many are clearly not suitable for this purpose. The use of the word “amount” for example cannot be linked to any meaningful player attribute, nor can the use of the word “league.”

Figure 7. Words Predictive of Performance

Word	Document Frequency	Model Popularity
forechecker	4	4
league	108	2
amount	28	2
multiple	12	2
defense	237	2
concerned	7	2
prospect	185	1
highend	168	1
average	262	1
fantastic	63	1

Finally, there is a key limitation in measuring the ‘goodness’ of the raw rankings provided by the scout using a linear regression method as we did here. The linear nature of such a model prevents it from accounting for random shifts in year over year player ability. Despite the fact that we provided this model with the draft year as a predictor, the linear nature of this model will provide it with a general ability to adjust for later drafts producing players who have not yet reached their full potential, but it cannot adjust for the possibility that some drafts (2015 for example), produced players who were superior to those who came both before and after. Although the Random Forests model appears to consistently outperform the linear benchmark of player ranking when trying to predict performance, we cannot ignore the fact that it is doing so in a manner which is as yet inexplicable and therefore may not be truly viable going forward.

Conclusions

In exploring this corpus of scouting reports, we have shown that a high degree of accuracy in classification of player position (97%) is possible using only the TF-IDF data from each report. Additionally, we’ve shown that although *some* correspondence exists between

Word2Vec and Doc2Vec representations of this relatively small corpus and the human ontological categories which organize and underly them, the size of the corpus is ultimately insufficient for this method to reach its true potential.

More importantly, we have taken a first pass attempt at leveraging TF-IDF data and various machine learning techniques for the purpose of improving upon the rankings which accompany these reports. In so doing, we have laid the groundwork for improving upon expert opinion using NLP and machine learning techniques in a particular domain in which expert opinion currently constitutes the state of the art (and indeed only) basis for decision making. Although the results shown in this experiment are not to be taken as evidence of absolute success in this objective, they constitute a promising beginning for future work.

Future Work

The work outlined in this paper represents a start on a problem for which there are many possible paths toward further refinement. Our TF-IDF language data was examined only at the level of a single n-gram, which left a lot to be desired regarding context. Particularly where reports used keywords in the negative sense, this resulted in unresolvable difficulty for our very basic model. Possibilities for overcoming this difficulty include increasing the n-gram range and seeking out the assistance of pre-trained Word2Vec and Doc2Vec models to leverage their more context-aware abilities. Both options will likely require extensive feature engineering work and might successfully combine into an effective single model if dimensionality reduction can be performed to a sufficient degree to make the merging of these models practical.

Reference List

Bilgin, Metin. 2020. “Classification of Turkish Tweets by Document Vectors and Investigation of the Effects of Parameter Changes on Classification Success.” *Sigma: Journal of Engineering & Natural Sciences / Mühendislik ve Fen Bilimleri Dergisi* 38 (3): 1581–92.

Brena, Ramon, Eduardo Ramirez, David Pinto, and Vivek Singh. 2019. “Scalable Text Semantic Clustering around Topics.” *Journal of Intelligent & Fuzzy Systems* 36 (5): 4645-57. Doi:10.3233/JIFS-179015.

Buatoom, Uraiwan, Waree Kongprawechnon, and Thanaruk Theeramunkong. 2020. “Document Clustering Using K-Means with Term Weighting as Similarity-Based Constraints.” *Symmetry (20738994)* 12 (6): 967. doi:10.3390/sym12060967.

Chen, Yang, Rongfeng Zheng, Anmin Zhou, Shan Liao, and Liang Liu. 2020. “Automatic Detection of Pornographic and Gambling Websites Based on Visual and Textual Content Using a Decision Mechanism.” *Sensors (14248220)* 20 (14): 3989. doi:10.3390/s20143989.

Dias Canedo, Edna, and Bruno Cordeiro Mendes. 2020. “Software Requirements Classification Using Machine Learning Algorithms.” *Entropy* 22 (9): 1057. doi:10.3390/e22091057.

Maas, Andrew L., Daly, Raymond E., Pham, Peter T., Huang, Dan, Ng, Andrew Y. and Potts, Christopher. "Learning Word Vectors for Sentiment Analysis." Paper presented at the meeting of the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, Stroudsburg, PA, USA, 2011.

Mikolov, T., I. Sutskever, K. Chen, GS Corrado, and J. Dean. “Distributed Representations of Words and Phrases and Their Compositionality.” In *Advances in Neural Information Processing Systems* 26 (NIPS 2013), 2, edited by C.J.C. Burges, L. Bottou, M. Welling, Z.Ghahramani, and K.Q. Weinberger, 3111-3119. Red Hook, NY: Curran Associates, 2013.

Mohammed, Manal, and Nazlia Omar. 2020. “Question Classification Based on Bloom’s Taxonomy Cognitive Domain Using Modified TF-IDF and Word2vec.” *PLoS ONE* 15 (3): 1–21. doi:10.1371/journal.pone.0230442.

Nguyen, Hoang. 2017. “Fuzzy Methods in Risk Estimation of the Ship System Failures Based on the Expert Judgments.” *Journal of Konbin* 43 (1): 393–403. doi:10.1515/jok-2017-0058.

Pronman, Corey. “Pronman: Final 2015 NHL Mock Draft” *ESPN+*. June 25, 2015.

Pronman, Corey. “Pronman’s 2017 NHL Draft Rankings: Top 100 Prospects” *The Athletic*. June 7, 2017.

Tixier, Antoine J.-P., Matthew R. Hallowell, Balaji Rajagopalan, and Dean Bowman. 2017. “Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes Using Data Mining.” *Automation in Construction* 74 (February): 39–54. doi:10.1016/j.autcon.2016.11.001.

Zhu, Yinglin, Wenbin Zheng, and Hong Tang. 2020. “Interactive Dual Attention Network for Text Sentiment Classification.” *Computational Intelligence & Neuroscience*, November, 1–11. doi:10.1155/2020/8858717.