METHODS FOR UNDERSTANDING AND PROCESSING EXPERT OPINION

Gordon Irving
gordon.e.irving@gmail.com

June 4, 2021

Keywords:

NLP, BERT, TF-IDF, LDA, Expert Opinion

Abstract

Although methods in Natural Language Processing (NLP) have seen significant growth during the last ten years, surprisingly little has been written about how differences in textual features affect the viability of specific methods. Written expert opinion represents one form of corpora possessing distinct features which must be recognized and dealt with. We demonstrate how existing methods for text classification, sentiment analysis and topic modeling may be adapted to fit the unique features of written expert opinion. Specifically, we leverage a method known as Term Frequency Inverse Document Frequency (TF-IDF) for the purpose of text classification, create an ensemble model for sentiment analysis using Bidirectional Encoder Representations from Transformers (BERT), and make a first pass attempt at sentence-level topic modeling using a Latent Dirichlet Allocation (LDA) method on a small corpus of roughly 1,300 scouting reports of NHL draft eligible players. Throughout this work, we draw special attention to key differences in how these methods apply to our highly specialized text versus how they tend to apply with texts in which a wider freedom of expression (and less focussed occasion for writing) is common.

Introduction & Problem Statement

The presence of technical jargon and the relatively limited ranges of expression and vocabulary common in written expert opinion represent distinct features which differentiate these documents from other types of corpora (for example, collections of tweets or movie reviews, etc.). Where tone is suppressed in favor of impartiality and documents exist for the purpose of objective comparison between two or more things, attempts to capture topic or sentiment at the document level may be dubious at best. Attention to key details is necessary to understand the meaning of texts such as these, which requires that we reach the sentence level or lower when applying methods for sentiment analysis and topic modeling.

This paper examines how a combination of Natural Language Processing (NLP) techniques can be applied to a highly specialized corpus of written expert opinion. We pay close attention to the differences between how these methods apply in this instance, and how they tend to apply in use cases in which a wider freedom of expression, less focussed occasion for writing, and broader vocabulary is more common. Here we make the argument that the highly specialized nature of written expert opinion provides us with both unique challenges and unique opportunities to glean insight and represent the meaning of text.

Our corpus of documents here consists of over 1,300 scouting reports of NHL draft eligible players spanning the years 2008 to 2021, with each representing a relatively uniform description of player strengths and weaknesses. A Common barrier to understanding these documents is the presence of implied meaning, which is not explicit, but nevertheless serves to underly and organize meaning at higher levels. For the current corpus, the ontological structures underlying all other meaning come in the form of player position, which is a vital aspect along which player evaluation takes place. When working with reports at scale, these positional labels are sometimes missing (as was the case for the data used here). To extract this important underlying feature for the reports in our corpus, we construct a neural network which leverages Term Frequency Inverse Document Frequency (TF-IDF) text representation in order to identify player position even when positional labels are missing.

Next, we seek to understand sentiment. This is another aspect in which the current corpus of documents differentiates itself from more traditional corpora, such as movie reviews, user comments online, collections of tweets, etc. Because these reports represent a professional attempt at impartial evaluation, sentiment at the document level is not relevant. Where movie reviews, product reviews and tweets can all reasonably be said to contain document-level sentiment pointing either toward positive or negative, the current set of documents cannot. Sentiment here exists at the sentence level. To extract sentence-level sentiment from our documents, we make use of the Hugging Face Transformers package for Tensorflow 2 and create an ensemble of three fine-tuned Bidirectional Encoder Representations from Transformers (BERT) models, each trained through a slightly different process using a collection of 50,000 movie reviews labeled with positive and negative sentiment.

Finally, we apply a popular topic modeling algorithm known as Latent Dirichlet Allocation (LDA) to our data for the purpose of understanding the topics contained in each sentence and adapting our understanding of sentiment into an Aspect Based Sentiment Analysis (ABSA). It should be noted that the use of this algorithm at the sentence level is not a traditional approach. In fact, it has been previously noted that due to the way this algorithm models topics based on probability distribution, it is best applied to larger documents such as news articles (Ozyurt, and Ackayol, 2021). The position of this paper however, is that the unique nature of expert opinion documents such as those used here makes topic modeling at the sentence-level with LDA more reasonable to attempt than it would be for other corpora. Specifically, the highly specialized language, limited vocabulary, and adherence to relatively uniform criteria for player evaluation, allows the LDA model to achieve much higher sentence-level topic coherence than would be possible on a collection of documents drawn from more diverse sources.

Literature Review

The last decade of research provided several key innovations in the field of natural language processing. Beginning around that time, experiments in clustering movies by type using reviews 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).

Classification techniques have proven useful in a wide range of applications, from classifying tweets and comments online (Bilgin, 2020), to identifying and blocking pornographic websites (Chen et al. 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 (Nguyen, 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).

Perhaps the most important development of recent memory is that of the Biderectional Encoder Representations from Transformers (BERT) model, developed by researchers at Google. By implementing bidirectional training on transformers (as opposed to a sequential first to last training), the researchers behind BERT found that they could generate a model with greater contextual awareness for the meaning of words (Devlin, Chang, Lee, and Toutanova, 2018). Since the publication of this research and the release of several pretrained BERT models, many disparate fields have incorporated BERT models into their research such as the monitoring and analysis of investor sentiment (Li, Li, Wang, Jia, and Rui, 2021) and the use of text generation engines to tell stories based on sequential images (Su, Dai, Guerin, and Zhou, 2021). The public availability of several pretrained BERT models and BERT variants make the model both accessible and popular among professionals in the field (Ravichandiran, 2021).

Also within the last ten years, topic understanding has been combined with sentiment analysis for the purpose of creating Aspect-Based Sentiment Analysis (ABSA). The goal of this method is ultimately to provide insight into what aspects or qualities are being discussed when positive or negative sentiment is expressed toward a given entity. One common challenge to ABSA is that of reaching the level of granularity required for smaller documents which, due to data sparsity and a lack of co-occurrence patterns, resist the probabilistic Bag-of-Words assumption (that order does not matter) present in the LDA algorithm (Ozyurt and Akcayol, 2021). In other words, because the LDA algorithm views documents as random combinations of word samples drawn from a distribution determined by underlying topic, certainty regarding the true topics present in a document is easier to attain when the sample of words is larger. Adaptations of this algorithm for use in ABSA, have included the ADM-LDA approach, which eliminates the Bag-of-Words assumption (Bagheri, Saraee, and de Jong, 2014). In this algorithm a new assumption is made; that the topics of words contained in documents can be represented as a Markov chain in which prior topics of words influence the probability of topics for subsequent words. A similar approach was used in a more recent work in which the proposed method was called SS-LDA (Ozyurt and Akcayol, 2021). Other approaches to adapting LDA to work on short texts have included the aggregation of short documents into long documents for training (Nimala, Magesh, and Thamizh, 2018).

Data Description & Preparation

Our corpus consists of roughly 1,300 player scouting reports for NHL draft eligible players from the years 2008-2021. Report word counts range from as few as 60 to as many as 300. Reports have been collected from the work of NHL scouts Corey Pronman and Scott Wheeler, both of whom publish their work at *The Athletic*, as well as from a professional scouting organization who publishes their work online at nhlentrydraft.com. For most reports, our dataset also includes the name and position of each player, although positional labels were not always present in some of the older reports. To demonstrate the relative uniformity of reports (which is an important feature here) three examples of reports are shown below. Special attention should be paid to the repetition of themes across reports, which consistently discuss such player attributes as physicality, skill level, skating ability, decision making, overall offensive and defensive ability, as well as player pedigree (i.e. how they’ve performed at lower levels). Additionally, it should be noted that even for human readers there can be difficulty in discerning the positions played by individual players.

Karabacek has played ahead of his age group for the Czechs in international play, and was in Austria during 2012-13 ahead of his encouraging North American debut in 2013-14. He controls the puck with some flashy skills, and good overall instincts and decision making. His shot has developed well, as he gets the puck off his stick quickly and with accuracy. His size (5-11, 185 pounds) is a limitation, but not to the same degree as some of the other skilled forwards on this list.
(Pronman, 2014)

When you have the length that he does and you can still skate as he can, you've got a lot to work with. The raw talent is exciting because it shows up in flashes in games when he joins the rush and carries the puck through the neutral zone. But I still have concerns about his decision-making as well as his puck handling. He also struggles to make high-end plays with the puck once his speed has pushed him over the offensive zone blueline. In the right program, there's no question that there's a lot to tap into. But there's definitely a lot of risk associated with taking him in the first half of the first round.
(Wheeler, 2019)

Cernak has been a very impressive prospect for some time, being one of Slovakia's top under-18 products of the past 10-15 years. He has a ton of high-level experience, including a great season in Slovakia's top league and playing for the national squad. His tool kit is extremely appealing, and he has a raw upside that's sky high: He projects to skate and handle the puck at above-average NHL levels. Cernak is also a physical defender who can make defensive stops as well as play a part on a team's power play. His

> main issue is his hockey IQ, as he shows a moderate frequency to make bad decisions on hits, pinches and puck decisions. At the top of his game, he can look like a dominant two-way defenseman, but that is not always the guy who shows up.
>
> (Pronman, 2015)

Data cleaning procedures varied depending on the model purpose. For instance, one model used here, a simple two-layer Neural Network which takes in TF-IDF vectorized representations for the purpose of determining player position, required removal of punctuation and non-alphabetic characters using the Python RegEx (regular expression) module, as well as the removal of player names. Text preparation for this model also involved the changing of all characters to lower case and the filtering out of stop words as identified in the Natural Language Toolkit (for example, ‘which,’ and ‘these’). TF-IDF n-gram range was set to one, which here means we permit one word per term, no two or three-word combinations. Finally, we constructed a single matrix to represent reports in which each word present in the corpus was represented as a single column, with reports represented as rows. The row-wise values of this matrix were TF-IDF vectors representing each player report and indicating the TF-IDF value for every word across all reports.

As mentioned previously, the features specific to our corpus make document-level sentiment analysis inappropriate and require us to perform sentiment analysis at the sentence level. Thus, the first stage of data preparation for the BERT based sentiment models required that each report be divided at the sentence level. Next, the BERT tokenizer available in Tensorflow 2 was used to prepare our texts for the specific operations necessary for a BERT based sentiment model. These BERT specific preprocessing steps involve the creation of special tokens to indicate the beginning and ending of text sequences, as well as to create padding tokens, which ensure all processed documents (or, in our case sentences) are a fixed length.

Text preparation for topic modeling took on a unique form as well. Just as was done in preparation for sentiment analysis, we broke each document down to the level of the individual sentence. Next, all punctuation and numeric characters were removed and all text was made to be lower case, leaving only the words for each sentence. Finally, using the parts of speech tagger from Natural Language Toolkit, extraneous parts of speech were removed. This tagger recognizes over 30 categories of word based on the eight English language parts of speech and was leveraged in this case to remove all words except nouns and adjectives. Certain verbs, which the tagger mis-identified as nouns (for example, the present tense verb “skates,” which was misidentified by our tagger as the plural noun “skates”) remained even after preprocessing. Out of data handling necessity, a dataframe was constructed in which each row represented one report, with each column representing a sentence within that report. The variable length of reports here required that filler sentences be included for most reports in order to obtain a uniform report length. Words not found elsewhere in the corpus were chosen for the filler sentence and the same exact sentence was used as filler in every instance.

Methods & Modeling

Training a neural network classifier to predict player position, is the first and most important key to interpreting the meaning of each report. Here the raw text data was cleaned and tokenized using the TF-IDF vectorization process described above. Then, holding 20% of the reports out for test purposes, we ran the remaining 80% through a simple two-layer neural network with the number of input nodes matching the total length of the vocabulary contained in the corpus. In other words, the total length of our TF-IDF vectors (described above) determined the number of input nodes. Using a 20% dropout from one layer to the next helped avoid overfitting. Finally, a binary output determined whether the positional label should be classified as forward or defender. This model trained for only 10 epochs.

Due to a lack of labeled sentiment data in our corpus of interest, we made use of the Stanford IMDB dataset (available at https://ai.stanford.edu/~amaas/data/sentiment/) for the purpose of fine-tuning sentiment classifiers. Three separate BERT models were then fine-tuned on this dataset, each time using a different training and test set as well as a variable number of training epochs (two, three and four epochs respectively). This process yielded an ensemble of three slightly different BERT based sentiment classifiers which were then used for sentiment analysis on the corpus of interest. Sentiment classification was then performed at the sentence level. The resulting set of sentiment scores ranged from negative three to positive three. Instances where there was disagreement among the three fine-tuned models yielded scores of either negative one or positive one, indicating slight lean toward either the positive or negative class.

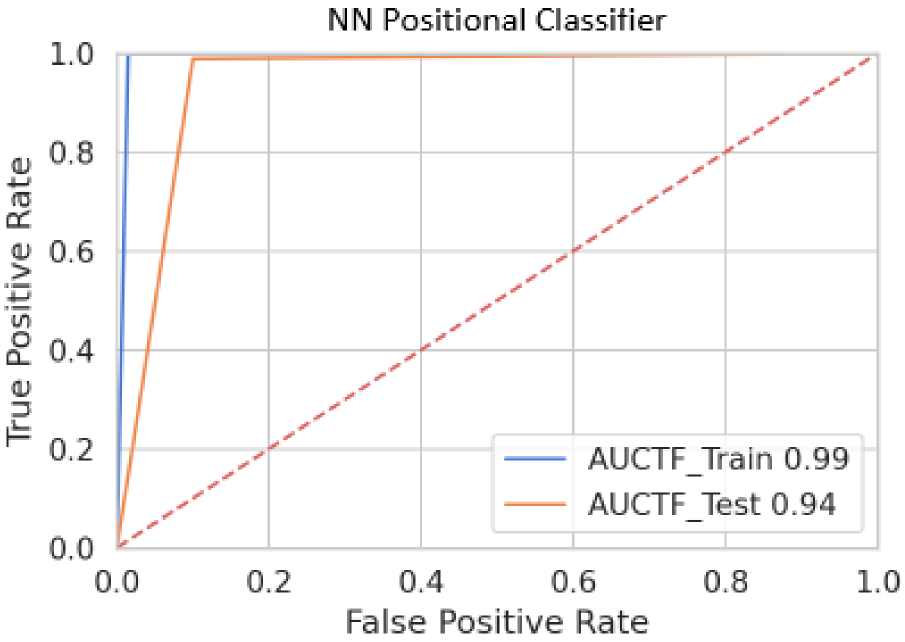
To leverage this sentence-level sentiment analysis into an aspect-based sentiment analysis, the final necessary step was to train a topic model for use at the sentence level. Here we selected the LDA algorithm for topic modeling. Training passes were set at 100 with the number of topics set at 13. It is worth noting that a traditional method for determining the proper number of topics in an LDA topic model is to calculate perplexity and coherence scores. However, with our extremely small documents (in some cases only a few words per sentence) traditional perplexity and coherence measures did not make sense. Instead, through experimentation and subjective human evaluation, 13 topics was determined to be the best number.

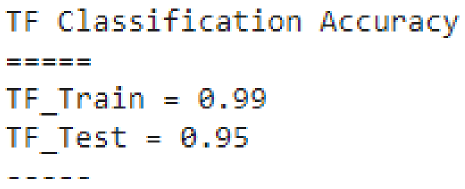
Finally, determining the most likely topic associated with each sentence was accomplished by creating a function capable of iterating through all sentences, extracting the probability of each individual word belonging to a particular topic, and then labeling the sentence as belonging to the topic cluster with the highest sum probability according to all its individual words. Additionally, because preprocessing steps produced approximately 80 sentences which were either completely blank or otherwise unsuitable for topic classification, a fourteenth topic label was created manually so that these sentences could be held out and not affect the integrity of any other topic clusters.

Results (TF-IDF Classification)

Using our Neural Network and TF-IDF vectorized representations of reports for the purpose of classifying player position proved highly successful. Depending on training and test splits, test accuracy varied from roughly 94% at the low end to 97% at the high end. Figures 1.1 and 1.2 below show area under the curve and precision/recall statistics respectively for one typical trial run of this experiment.

Figure 1.1





```
TF Classification Accuracy
=====
TF_Train = 0.99
TF_Test = 0.95
-----
```

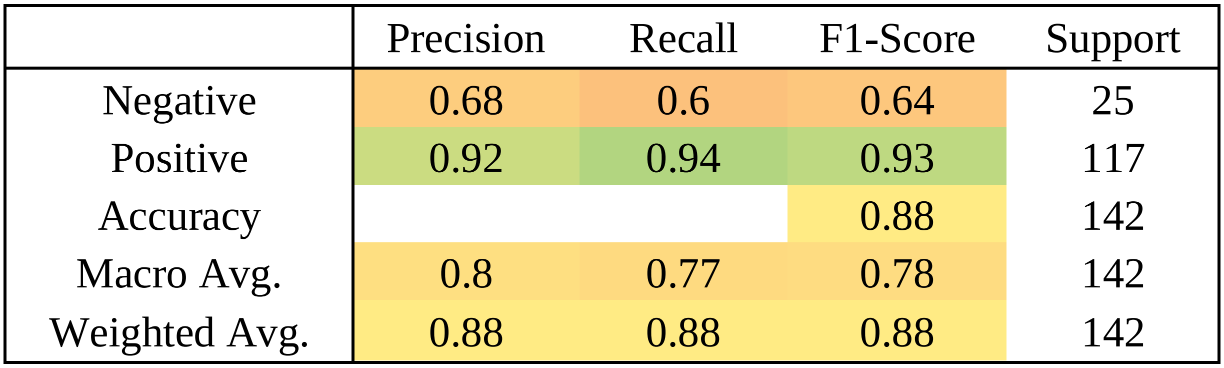
Figure 1.2

	Precision	Recall	F1-Score	Support
Defense	0.98	0.9	0.94	90
Forward	0.95	0.99	0.97	178
Accuracy			0.96	268
Macro Avg.	0.96	0.94	0.95	268
Weighted Avg.	0.96	0.96	0.96	268

Results (Sentiment Analysis)

Because our corpus did not come with strict positive and negative sentiment labels, results for the ensemble of BERT based sentiment classifiers are slightly harder to quantify. To establish some baseline level of results for analysis, we took a random sample of 142 sentences and classified them as positive or negative based on human judgement. In instances where single sentences contained both positive and negative sentiment, assignment to the negative class was always chosen. This was done both for the sake of convenience and to bolster the number of negative sentiments found in our corpus, which is heavily skewed to the positive class. The results from this sample are shown in figure 1.3 below.

Figure 1.3



	Precision	Recall	F1-Score	Support
Negative	0.68	0.6	0.64	25
Positive	0.92	0.94	0.93	117
Accuracy			0.88	142
Macro Avg.	0.8	0.77	0.78	142
Weighted Avg.	0.88	0.88	0.88	142

Results (LDA Topic Model)

Running the LDA algorithm at the sentence level produced the following topic clusters. Figure 1.4 below shows the breakdown of each topic cluster, as well as the top words for each. Although there is some overlap between certain topics, a clear hierarchy emerges. Please note that the filler sentences described above under data preprocessing all clustered together under the same topic and have been excluded from this example.

Figure 1.4

Pedigree	**Cluster Number**	**Top Words**
	5	draft, best, better, teams, class, pick, nhl, awareness, situations
	9	nhl, teammates, junior, level, consistent, team, opportunity, team, opportunity
	10	players, teammates, games, points, scorer, line, needs, possession, past, forward
Physicality	**Cluster Number**	**Top Words**
	11	rush, guy, game, defenders, able, shots, aggressive, nice, able
	1	hard, true, battle, drive, muscle, element, contact, crease, loads, added
	0	game, strength, physical, defensive, size, solid, body, pass, feet, level
Puck Skills & Skating	**Cluster Number**	**Top Words**
	8	puck, speed, good, hockey, strong, stride, skater, ice
	7	dynamic, steep, abilities, effort, explosive, flashy, edges
	2	ice, puck, times, smart, areas, play, passes, real, point, boards
Offense	**Cluster Number**	**Top Words**
	3	shot, play, puck, good, vision, quick, hands, stick, power
	6	zone, offense, year, high, skill, level, good, open, impressive, position
	4	offensive, game, ability, puck, mobility, upside, opponents, excellent, space, skill

Note that where obvious overlap exists, figure 1.4 organizes the overlapping topics by block. Overlap is not ideal, but in this instance it is not a fatal flaw either, as what we end up with is a hierarchy of four main topics with three related sub-genres each. The figure above shows this hierarchy. The broader themes here appear to be about player pedigree (i.e., performance at the junior levels, where they played, and where they stack up with the competition), physicality, special skills (skating and puck skills mostly), and offensive ability.

Drilling down further into the validity of these individual topics, we can examine the individual sentences. We do this by first determining the topic cluster each sentence most likely belongs to (remember, LDA gives us a probability distribution, not a strict label) and then examining whether there is any coherence among sentences belonging to the same topic. The most internally consistent topic appears to be topic cluster zero. Figure 1.5 below shows some examples of sentences which our methods determine belong to this cluster as well as how each was classified according to the sentiment model discussed in earlier sections.

Figure 1.5

Topic Cluster	**Player**	**Sentiment**	**Sentence**
0	David Pasternak	Leans Negative	His main weakness is in his **physical game**, as he loses a fair amount of battles.
0	Nick Ritchie	Positive	Ritchie's **physical game** is fantastic.
0	Nick Ritchie	Leans Negative	It will obviously be harder on him in the pro game, but it was near-impossible to get the puck off of him at the junior level.
0	Nick Ritchie	Positive	Ritchie's **strength level** is great, and I've seen him **bulldoze through people** on the way to the net, and shoulder off multiple checks without much effort.
0	Sam Reinhart	Leans Positive	He has **average size** (6-1, 183 pounds), but will work hard in battles and can play an average **physical game**, although I have heard some scouts criticize him in that area.
0	Jared McCann	Negative	He certainly needs to get **stronger**, and despite a good **phyical** work ethic he projects as a below-average **physical** player at the top level.

1.5 (Continued)

Topic Cluster	**Player**	**Sentiment**	**Sentence**
0	Aaron Ekblad	Positive	Ekblad is the complete package, a **strong** two-way defenseman with better than NHL-desirable **size.**
0	Aaron Ekblad	Positive	He gets to where he needs to just fine with a long, **powerful** stride, and consistently uses his natural gifts such as **strength** and reach to make the right plays.
0	Aaron Ekblad	Leans Positive	He is **physical** when he wants to be and can make life difficult on opponents.
0	Brady Tkachuk	Leans Positive	In terms of his **physical tools**, he is [ahead of others].

What makes this set of sentences so impressive is that not only are we able to clearly identify the coherent topic (physicality), we are also able to accurately classify directional sentiment within the topic identified. This is not always the case, however. As seen in figure 1.6 below, not all topic clusters are as internally consistent or identifiable.

Figure 1.6

Topic Cluster	Player	Sentiment	Sentence
9	Philip Tomasino	Leans Negative	Between teammates Ben Jones, Akil Thomas, Kirill Maximov, Jack Studnicka, Jason Robertson, Ivan Lodnia and draft-eligible teammate Kyen Sopa, there are only so many goals to go around right now in Niagara.
9	Rasmus Andersson	Leans Positive	If you told me two years ago Andersson would not be in my top 15, I'd have called you crazy.
9	Aleksi Haatanen	Positive	Haatanen had a solid campaign between the Finnish junior and second-tier pro level.

1.6 (Continued)

9	Anntii Saarela	Positive	Saarela came into the season as a top-30 prospect after impressing in his 15- and 16-year-old seasons.
9	Leevi Aaltonen	Positive	Aaltonen has been a top player in his age group for years, often playing ahead of his age level.
9	Rasmus Kupari	Positive	Kupari is all of the things the NHL is trending towards: Agile, up-tempo, crafty.

In each of the above instances, we observe that the theme is discussing how the player stacks up to their draft class or how they performed in previous seasons. This matches the overall label of pedigree, which is the higher-level category to which this one topic cluster belongs. However, the internal consistency of these sentences, where some comment on previous performance and others comment on specific player attributes, is debatable. The usefulness of clusters like these is questionable because it does not help us understand very much about the content of what is being discussed.

Analysis and Interpretation (TF-IDF Classification)

Reaching overall classification accuracy of between 94-97% (depending on training and test splits) demonstrates the power of TF-IDF as a means of representing the contents of documents within a corpus. Here we find a computationally simple way of making inferences about player position, which is an important underlying structure along which player evaluation is carried out. Unsurprisingly, this model performs slightly worse when attempting to identify the minority class, but even here a 90% recall statistic is respectable.

Analysis and Interpretation (Sentiment Analysis)

Although our sentiment analysis ensemble model seems to classify positive sentiment fairly well, it struggles with the negative class. Figure 1.4 below takes a report we are already familiar with (see Data Description above) and breaks down the sentiment classification sentence by sentence, with each voting component of our ensemble model shown.

Figure 1.4

Sentence	Model 1	Model 2	Model 3	Total
Cernak has been a very impressive prospect for some time, being one of Slovakia's top under-18 products of the past 10-15 years.	Positive	Positive	Positive	3
He has a ton of high-level experience, including a great season in Slovakia's top league and playing for the national squad.	Positive	Positive	Positive	3
His tool kit is extremely appealing, and he has a raw upside that's sky high: He projects to skate and handle the puck at above-average NHL levels.	Positive	Positive	Positive	3
Cernak is also a physical defender who can make defensive stops as well as play a part on a team's power play.	Positive	Positive	Positive	3
His main issue is his hockey IQ, as he shows a moderate frequency to make bad decisions on hits, pinches and puck decisions.	Negative	Negative	Negative	-3
At the top of his game, he can look like a dominant two-way defenseman, but that is not always the guy who shows up.	Positive	Positive	Negative	1

Here we can see that the first five sentences result in total agreement across all three sentiment classifiers, with the sixth sentence being the only one in which there is disagreement.

Looking at that sentence more closely, it is easy to understand why that would be the case. Here there is a combination of positive and negative sentiment present. Although our scoring criteria, which was developed prior to analyzing these results, would have classified this as a negative statement, there is a reasonable argument to be made either way. This one sentence may be more accurately understood to contain two opposing sentiments, which leads us to a key insight; that the ability of our ensemble classifier to distinguish positive, negative, *and* ambiguous statements is a strength of the model. Of course, there are still many sentences which this model legitimately struggles to accurately classify, but our results here are highly encouraging overall.

Analysis and Interpretation (LDA Topic Model)

The application of Latent Dirichlet Allocation to the sentence level is unconventional. The uniqueness of our current set of documents made our experiment worthwhile, and in a few key topic clusters we found an encouraging level of topic integrity and coherence, but on the whole we were unable to establish a working topic model for use at the sentence level due to the same data sparsity issues commonly known to make sentence level LDA untenable in other settings. While seeing some signs of promise, we cannot deny that further work is needed to establish a topic model which accurately captures the meaning of text at the sentence level. In order to reach any level of coherence with Aspect Based Sentiment Analysis, improvements to our topic model must be made first.

Conclusions

The features specific to written expert opinion differentiate this form of text data from more common corpora such as movie reviews, tweets, and user comments. Specifically, the finite number of aspects along which player evaluation is carried out provides a consistency of topics which is uncommon in less focused corpora. The opportunity exists for organizations who generate written expert opinion to glean insights about their own process and manage reports at a high level. Where implicit underlying features are critical to interpreting meaning (as was the case here with player position) the opportunity exists to extract these features and organize reports accordingly using only basic TF-IDF methods and Neural Network architectures. Although document-level sentiment is more difficult to establish in reports where the occasion for writing is objective assessment, sentence level sentiment analysis shows potential for tremendous value in identifying positive, negative, and ambiguous statements. Most crucially, sentence-level topic modeling methods which are ineffective in more diverse corpora appear better suited to the relative rigidity with which field experts express their analysis. Although imperfect in this one example, the potential for establishing highly effective sentence-level topic models for expert opinion documents is clear.

Directions for Future Work

Two key areas of focus for advancing this research are recommended. First, the development of metrics by which diversity of expression and size of vocabulary *relative to size of corpora* can be more exactly measured. If indeed the features of limited expression, repetition, and uniformity are the reason that sentence level LDA topic modeling showed more promise here than it has in previous work, then these features ought to be more closely defined. The exact effect of these features on topic modeling algorithms must be quantifiable in some way if it really exists. Secondly, it is recommended that our sentence level topic modeling be adjusted in order to more accurately portray the content of each sentence. One method of accomplishing this lies in adjusting the training process of the LDA algorithm by inserting hand-crafted documents with even greater uniformity of terms and co-occurrence patterns. Here again, the relative uniformity of documents provides us with a greater opportunity to accomplish this task because there are a finite number of possible topics and much more uniform expressions.

Reference List

Bagheri, Ayoub, Mohamad Saraee, and Franciska de Jong. 2014. “ADM-LDA: An Aspect Detection Model Based on Topic Modelling Using the Structure of Review Sentences.” *Journal of Information Science* 40 (5): 621–36. doi:10.1177/0165551514538744.

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.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding.” *arXiv preprint arXiv: 1810.04805.*

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

Li, Menggang, Wenrui Li, Fang Wang, Xiaojun Jia, and Guangwei Rui. 2021. "Applying BERT to Analyze Investor Sentiment in Stock Market." *Neural Computing & Applications* 33 (10): 4663–76. doi:10.1007/s00521-020-05411-7.

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, 2011.19

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.

Nimala, K., S. Magesh, and R. Thamizh. 2018. "Hash Tag Based Topic Modelling Techniques for Twitter by Tweet Aggregation Strategy*." Journal of Advanced Research in Dynamical and Control Systems*. 3 (1): 571-578.

Ozyurt, Baris, and M. Ali Akcayol. 2021. “A New Topic Modeling Based Approach for Aspect Extraction in Aspect Based Sentiment Analysis: SS-LDA.” *Expert Systems with Applications* 168 (April): N.PAG. doi:10.1016/j.eswa.2020.114231.

Pronman, Corey. 2014. “Corey Pronman’s Top 100 Draft Prospects Index – NHL Draft 2014.” *ESPN+*. May 12, 2014

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

Ravichandiran, Sudharsan. 2021. *Getting Started with Google BERT.* Birmingham: Packt Publishing Ltd.

Su, Jing, Qingyun Dai, Frank Guerin, and Mian Zhou. 2021. “BERT-HLSTMs: BERT and Hierarchical LSTMs for Visual Storytelling.” *Computer Speech & Language* 67 (May): N.PAG. doi:10.1016/j.csl.2020.101169.

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.

Valkov, Venelin. 2020. “Intent Recognition with BERT and TensorFlow 2 in Python | Text Classification Tutorial.” YouTube Video, 1:19:50. February 8, 2020. https://www.youtube.com/watch?v=gE-95nFF4Cc&t=851s&ab_channel=VenelinValkov

Wheeler, Scott. 2019. “Wheeler: Final Ranking for the 2019 NHL Draft’s Top 100 Prospects.” *The Athletic*. May 6, 2019.

Zhao, Alice. 2019. “Natural Language Processing (Part 5): Topic Modeling with Latent Dirichlet Allocation in Python” YouTube Video, 24:13. January 5, 2019. https://www.youtube.com/watch?v=NYkbqzTlW3w

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.