University of Wolverhampton

Research Methods 7LN006, Thesis Proposal

Pregame Prognosis

A study in sentiment analysis and metric prediction using National Hockey League pregame interviews

Browning, Travis S.
Instructor
Dr. Emad Mohamed
2-18-2021

Abstract

1) Introduction and Related Work

Sentiment analysis is a widely distributed interest in the field of computational linguistics. It is currently utilized in many commercial applications such as social media monitoring and product analysis. Previous applications of sentiment analysis in team sports have already demonstrated that a player's use of language can be interpreted to amplify the results of traditional metric based prediction systems on an individual level.

Ice hockey statistics are less widely available than statistics kept for other sports. Scoring events are rare. Media coverage teams for the National Hockey League (NHL) are in the midst of adding new analytic metrics into the language of the everyday broadcast. However, due to the fast and fluid nature of ice hockey, individual events are often more difficult to define. The small datasets stand in contrast to information gathered from other sports such as successful passing percentage as utilized in basketball, or a rate of duels won to define the performance of defenders in football (soccer). By extension, the amount of data available for metric based prediction models for hockey is smaller than it is for other sports such as football or basketball.

Metric based predictions of sport are utilized widely in the commercial betting industry. Despite the small dataset available, computational predictions of success on a team level of NHL games have been explored by utilizing pregame summaries combined with statistical metrics. Although these methods are available for predictions of team performance, similar forecasting is mostly absent for individual hockey players.

This study will aim to evaluate a model to analyze the mood of a hockey player in a pregame interview by reviewing keywords within the player's answer responses for an associated polarity score to define the player's mood. This mood value will be used as a classifier input, both on its own, and again concatenated with performance metrics from the previous three games, to forecast statistical output for the player in the subsequent game. These values will be judged for accuracy and compared to predictions that had been received from only metrical data as a baseline. With this aim in mind, the research questions of this experiment are:

1. Can classification models utilize pregame interview text of NHL players to predict some level of variance in their own in-game performance at a game level?
2. Can text be combined with past performance metrics to produce better predictions?

This proposal is presented in three sections. In Section 1, background literature will be explored, NLP processes are explained, and a brief introduction into hockey metrics as well as statistical predictions will be given. Section 2 will discuss the process of gathering and processing data, as well as the details of data analysis. In Section 3, the potential for the application of this research is discussed, as well as suggestions for further research in this area.

1.1) NLP and Prediction of Human Behavior

Sentiment analysis [6] is well established in NLP. Widely utilized commercial applications include analyses of reviews, identifying patterns and trends, as well as automation of media and social media monitoring. The process involves imputing numerous texts on a similar subject for training. Analysis of these texts reveal not only the sentiment value of the text, but also how the author felt about the subject that they are writing about.

When applied to NHL coach interviews, sentiment analysis was found to have told the story of the series in the 2016 Stanley Cup Finals [2]. Coach interviews were analyzed using NRC Word-Emotion Association Lexicon [9], which categorized keywords into categories of sentiment. Dramatic swings into sentiments such as ‘joy’, ‘trust’, and ‘fear’ are reflected in events of the series. Highlighted in the included chart, are values of sentiments for the interviews. Following Game 6 there was a clear shift in sentiments to ‘positive’ and ‘joy’ from the winning coach. As well as a drop in frequency for similar emotions, the losing coach demonstrated a spike in ‘anger’ and ‘disgust’ (Figure 1).

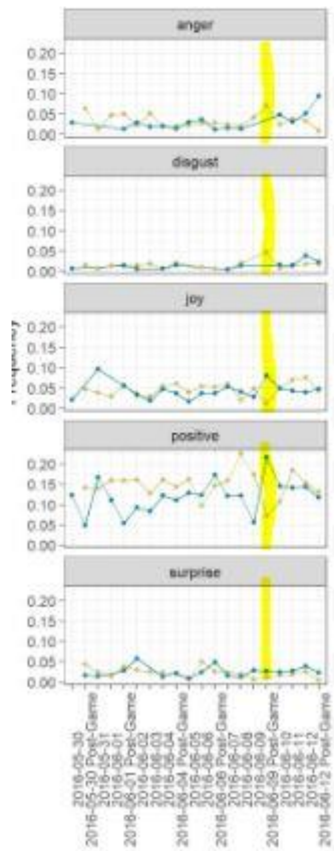

Figure 1

One research objective of Oved et al. was to identify an optimal model of language representation for predicting the in-game actions of basketball players based on mood analyses of pregame interviews. Three models were considered due to their performance in text classification: Convolutional Neural Networks (CNN) [13], Bidirectional Long Short-Term Memory (BiLSTM) [7], and Bidirectional Encoder Representations from Transformers (BERT) [11]. The CNN model excels in analyzing text on a word-by-word level. BiLSTM views a text on a sentence-level. For every sentence, a mean score is calculated based on the scores of every token within the sentence. BERT models excel at analyzing dependent statements simultaneously. The best performing text based model included primarily the architecture of BERT, as well as an attention mechanism adopted from BiLSTM. The output of this model (BERT-A-T) was then combined with a statistical summary of the previous three games (BERT-A-TM) to produce the highest scoring predictions.

Through tweets, rather than interview transcripts, Xu et al. also attempted to predict the on-court performance of NBA players through NLP. Tweets were selected as a medium due to the insight that they provided into players’ thoughts and opinions. Because players are entirely in control of the product of their tweets, the medium did provide the potential for self-selection bias when compared to the interview format. Similar to prediction models containing interview transcripts analyzed for sentiment, analyzed tweets were found to improve results when estimating a player’s statistical output.

1.2 Prediction and Analysis in Ice Hockey

Ice hockey data is gathered on a play-by-play level for events which are recorded in the box score. The following is a summary of metrics used to evaluate players found in a box score:

***Goal (G)**: A goal is scored when the puck crosses the goal line into the net. Credit to the goal is assigned to the last attacking player to have touched the puck. At the end of the game, the winning team has scored more goals. Goals can be categorized and tallied as **Even strength goals (EV), Powerplay Goals (PPG)**, **Shorthanded Goals (SHG)**. Every goal will fit into one of these categories. As well, goals can be recorded as **Game Winning Goals (GWG)** if they are the goal that decides the outcome of the game.*

Assist (A)**:** *Up to two assists are rewarded for teammates of a goal scorer, if they have passed the puck in sequence leading to the goal.*

***Point (P)**: The sum of goals and assists a player has been awarded.*

***Goal Differential (+/**-): *Goal differential is calculated for individual players in hockey, rather than being calculated for entire teams as it is in other sports. A player's +/- is defined as an integer, beginning at 0 for all players. If the player is on the ice for a goal scored their +/- will increase by 1, and if they are on the ice for a goal conceded it will decrease by the same amount.*

***Penalty minutes (PIM):** For most infractions a two-minute penalty will be received. While a player is penalized, they are banished to the penalty box, unable to reenter the ice until the penalty timer has expired. This leaves the team of the penalized player with a numerical disadvantage. More serious infractions such as fighting receive longer, major penalties.*

***Shot on Goal (SOG):** Any event in which the puck is heading to the net, destined to either result in a goal or a save from the goalie. Missed shots and blocked shots do not count as shots on goal. **Shooting percentage (S%)** is calculated as a ratio of goals over shots on goal. Goalie statistics will include a number of shots faced.*

***Saves**: SOG blocked by the goalie.*

***Hit:** Defensive action in which one player attempts to bump an opposing player off of possession of the puck. Hits do happen on open ice, but most often they occur along the boards on the boundary of the ice rink.*

***Block:** Defensive action in which a player obscures a goal bound shot (shots at a professional level are often recorded traveling over 100 mph (160 km/h)) with their stick or their body. This is typically done by sliding along the ice in between the shooter and the goal.*

***Faceoff win percentage (FO%):** In hockey, play is started by a faceoff. A referee drops the puck between two opposing players on one of nine locations on the ice. The players compete to win possession for their team.*

Time on Ice (TOI)**: Players perform rolling substitutions in ice hockey, they can enter and leave the ice from the bench during live play. As a result, players' involvement cannot be tracked with substitution events as it is in other sports. The time in between entering and leaving the ice is a **shift**.* Shifts normally last between 30 and 80 seconds at a professional level, separated by 3-to-5-minute periods of rest on the bench ***[17].

***Powerplay TOI**: Time on ice during powerplay situations. During a penalty situation, the team with the numerical advantage is on the 'powerplay'.*

***Shorthanded TOI**: Time on ice during penalty kill situations. During a penalty situation, the team with the numerical disadvantage is on the 'penalty kill' or playing 'shorthanded'.*

***Wins** and **Losses**: Are recorded for goalies*

***Save percentage:** Is a ratio of saves over SOG.*

While compiling data to predict performance on a team level, Weissbock [3] included advanced statistical features such as the Fenwick Close Percentage, a ratio of time a team is in possession of the puck compared to its opposition, and a representation of luck calculated by the sum of save percentage and shooting percentage. The Fenwick percentage was chosen due to having the highest r^2 correlation to points in the standings, not to be confused with player points, compared to G, +/-, Hits, and others (Charron, 2013, as cited by Weissbock, 2014). Luck is defined as the results of gameplay that fall outside of the boundaries and variance in a player's performance. Luck regresses to a value of 100% over the course of a season. A value over 100% is good luck, and a value below is bad luck. Team level statistics were fed into a tool called WEKA [8] for classification tasks. WEKA contains Neural Networks, Naïve Bayes, J48 and other algorithms. These algorithms were combined with traditional, advanced, and combined statistics were evaluated for effectiveness when predicting a result. Weissbock found Neural Networks combined with mixed data to be the most effective in predicting an outcome on the macro scale, and Neural Networks combined with traditional data to be most effective at predicting on a micro scale.

Rather than defining a norm, Oved et al. decided to use a player's statistical performance over the last three games to model metric based inputs into his classifier. This representation over three games could vary greatly from 3/82 of a player's statistics over an entire regular season, or even 3/7 of a playoff series.

2) Methodology

2.1) Research Hypothesis

A classification model can utilize player pre-game interview text to predict some variance of that player's performance on a game-by-game level. This text could then be combined with past performance metrics to produce better predictions.

2.2) Data Collection and Processing

A game of hockey is played on a rink between two teams. Under normal circumstances each team has five skaters and one goalie, controlling a rubber puck with sticks. Skaters are players who share the primary objective of scoring goals on the opposing team. The primary objective of the goalie is to block the opposing team from scoring goals. A game consists of three 20-minute periods. Teams play 82 games in an average regular season for a chance to compete in the Stanley Cup playoffs, leading to 1,230 games played in the entire league for the Regular Season. To win a round of the Stanley Cup playoffs you must beat your opponent four times in seven games. The Stanley Cup winner will have won four rounds of playoffs, or 16 games.

As a contractual obligation with their teams and the NHL, players are occasionally asked to attend interviews before, during, or after a game. A major validity concern going into this experiment will be the quantity of available pregame interviews. Interviews are a series of open-ended question and answer pairs, between a single member or multiple members of press media and another party, normally playing or coaching staff.

Transcribed pregame interviews of NHL players are available through ASAP Sports [5] from 1997 onwards. These interviews are presented in a Q/A format, already compatible with the BERT model. Within this archive there was, for most years, at least one playoff series containing a series of player interviews. For many seasons three series of interviews are available, in the form of two conference finals and one Cup final. Although most player interviews happen on practice days, in between periods, or post game, a vast majority of player pregame interviews are not captured in these transcripts. For modern players, most relevant interviews exist only in video format. Going into the experiment, it's anticipated volume of interviews may present a challenge. Through the transcripts at ASAP sports, 159 pairs of in game statistics and pregame interview text are available. In future research it is worth exploring the impact that speech to text engines could serve in enhancing the data set with audio and video interviews.

2.3) Model design and methods of data analysis

The goal, in analyzing each interview process will be to produce a sentiment score for the interviewee that can be interpreted as the player's mood. The AFINN sentiment lexicon [12] is used below in figures 1 and 2 demonstrate player sentiment analysis. AFINN, previously used by Tamming (2019) to define player sentiment in NHL interviews, is a dictionary of 3,000 keywords. Keywords are assigned a polarity score between -5 and +5. Texts are first filtered through stopword lists. Next, they are summarized by adding the polarity of the keywords in text. To address bias of statement length, the sum of these keywords is divided by the number of words in the statement. Pictured below is the end of an output of an interview with Mike Modano, before game 3 of the 1999 Stanley Cup finals, passed through AFINN for sentiment analysis in Figure 2. In Figure 3 the values of individual content words within questions and answers are displayed.

```
Q/A pair 9
Q) During the Playoffs Satan and Hasek were publicly criticized because they took themselves out of lineups and they could
n't help their team some fans and media didn't like that. Do you worry if you take yourself out of the lineup there will be
a stigma on you saying he couldn't go?
A) Well, you know, that feeling is between me and my teammates, I think; whether the fans like it or not, I don't really car
e. It is my decision and the players will understand that 100%. If I am able to go and contribute and play the way I can, as
long as I am skating well and my legs are fine, then I think everything else will kind of fall in place and the chances may
come, but, you know, what the consequences are of not playing doesn't really bother me.
The question has a mood score of: -0.0196078431372549
The answer has a mood score of: 0.0989010989010989

----------------------

Q/A pair 10
Q) Mike, did you feel any discomfort when you were out there this morning and, if so, where did it affect you the most?
A) Well, actually just -- I stretched a lot of ligaments in my wrist and my fingers at the same time while hurting my wrist
so there is a lot of discomfort in the whole area. But like I said, I could do a lot more than I thought I could do this mor
ning, but we will see tonight with the doctor and see what the magic of the needles can do.
The question has a mood score of: -0.08695652173913043
The answer has a mood score of: -0.028169014084507043

----------------------

Average mood score of questions: -0.02421305530741525
Average mood score of answers: 0.08491169556073383
Average mood score of interview: 0.03034932012665929
```

Figure 2

```
Q/A pair 9
Q:
[('During', 0.0), ('Playoffs', 0.0), ('Satan', 0.0), ('Hasek', 0.0), ('publicly', 0.0), ('criticized', -2.0), ('took', 0.
0), ('lineups', 0.0), ('could', 0.0), ("n't", 0.0), ('help', 2.0), ('team', 0.0), ('fans', 0.0), ('media', 0.0), ("n't",
0.0), ('like', 2.0), ('.', 0.0), ('Do', 0.0), ('worry', -3.0), ('take', 0.0), ('lineup', 0.0), ('stigma', 0.0), ('sayin
g', 0.0), ('could', 0.0), ("n't", 0.0), ('go', 0.0), ('?', 0.0)]
A:
[('Well', 0.0), (',', 0.0), ('know', 0.0), (',', 0.0), ('feeling', 1.0), ('teammates', 0.0), (',', 0.0), ('I', 0.0), ('th
ink', 0.0), (';', 0.0), ('whether', 0.0), ('fans', 0.0), ('like', 2.0), (',', 0.0), ('I', 0.0), ("n't", 0.0), ('really',
0.0), ('care', 2.0), ('.', 0.0), ('It', 0.0), ('decision', 0.0), ('players', 0.0), ('understand', 0.0), ('100', 0.0),
('%', 0.0), ('-', 0.0), ('If', 0.0), ('I', 0.0), ('able', 0.0), ('go', 0.0), ('contribute', 0.0), ('play', 0.0), ('way',
0.0), ('I', 0.0), (',', 0.0), ('long', 0.0), ('I', 0.0), ('skating', 0.0), ('well', 0.0), ('legs', 0.0), ('fine', 2.0),
(',', 0.0), ('I', 0.0), ('think', 0.0), ('everything', 0.0), ('else', 0.0), ('kind', 2.0), ('fall', 0.0), ('place', 0.0),
('chances', 2.0), ('may', 0.0), ('come', 0.0), (',', 0.0), (',', 0.0), ('know', 0.0), (',', 0.0), ('consequences', 0.0),
('playing', 0.0), ("n't", 0.0), ('really', 0.0), ('bother', -2.0), ('.', 0.0)]
The question has a mood score of: -0.0196078431372549
The answer has a mood score of: 0.0989010989010989
```

Figure 3

Sentiment analysis in the experiment will be generated by BERT. Oved et al. found BERT the most effective method at addressing the bias of the player only controlling the answer in the Q/A pair. Question/Answer pairs are represented as a vector that can be averaged to produce a numerical output for every pair. These vectors are averaged over the course of an interview transcript to generate a mood value for both participants, resulting in similar annotated values in word embeddings to those produced above by AFINN.

Question-Answer pair tokens are fed into BERT. The tag [CLS] will be used to mark the beginning of a pair. [SEP] will be used to mark the end of a statement in the pair. BERT will output an *h* score, denoting mood, for every segment of the pair. These *h* scores will be averaged to produce another *h* score for the entire pair. For every pair in the interview, an *h* score is inputted into the attention mechanism layer to produce an *h* score for the entire interview. Here, our two testing models will differ. In the BERT-A-TM model, this interview *h* score will be concatenated with the players metric output from the past three games and fed into a Deep Neural Network linear classifier to produce a prediction on the occurrence of an event. In the BERT-A-T model we will feed the interview *h* score into the classifier independently.

Ultimately, our goal is to recognize patterns in a player's game-by-game statistical output, correlated to the player's mood as defined by BERT. This understanding will lead to higher performing classifiers.

2.4) Results Validation

This research is expected to find similar findings to Oved et al. The findings are anticipated to demonstrate a text-based analysis combined with a metric analysis, which when fed into a classifier, will produce the strongest prediction scores for in-game actions.

A potential for bias exists, as a small group of media professionals are the only interviewers for our examples. The goal of these interviewers is to generate attention for their publication, and one of the skills nurtured by interviewers is bringing about an emotional reaction from interviewees with their questions. The objective of the interviewees is to promote themselves and their teams as a brand, before answering the interviewers' questions. Team provided media training, as well as access to other interviews throughout the league, can lead to interview answers feeling like a product of rote memorization, of scripted cliché, rather than a genuine conversation. This view is widely held in the realm of hockey analytics. In official NHL interviews, players have been recorded reflecting on the inauthentic and route nature of interview responses [14]. Despite the participant's familiarity with the process, transcripts of these interview events contain considerable amounts of linguistic data.

Factors outside of the player's mood will influence the statistical metrics. A player's skill level and athleticism will play as large an influence on their metrics as their cognitive function over the course of a season. These factors tend to remain constant for players over the course of a season, outside of events such as injury. On a game level, factors such as the ability of varying opposition, familiarity of ice conditions and the presence of an audience will have an influence on a player's output and can lead to the creation of a sampling bias.

As the models in our experiment are taken from Oved et al., these models will be expected to perform similarly with interviews and metrics of hockey players as it has previously with basketball players. To validate, 20% of our interviews will be used in tandem with corresponding performance metrics as a test set. A 5-fold cross validation procedure for each metric label is implemented, within each fold randomly sampling from a training set containing 64% of the interviews and a development set containing 16% of interviews. If our model performs similarly, the ratio of positive and negative examples in each subset is expected to be identical to the ratio in the data set.

3) Expected Contributions and Further Research

This experiment seeks to reproduce and validate a subsection of outcomes of Oved 2020 for a new subject matter. Findings of this experiment can be useful in team level personnel decisions, as well as in the sport betting industry. The value of basing decisions off of statistical data, rather than traditional insight, has been hailed in sports outside of hockey [10].

Large quantities of transcribed interviews with NHL coaching staff are available, as well as interviews with players recorded in between periods. It is worth exploring the effect that a coach's mood has on player performance on a game level. As well, this experiment could be repeated to evaluate the ability of predictive models which analyze players mood during in game interviews that occur between periods.

Any study of hockey culture in North America is of course incomplete without the inclusion of data from the hockey hotbed of Québec. It is common for Francophone players to be interviewed in French in multiple Canadian markets. To address this gap in data, first the transcripts of French language interviews would need to be obtained and then a parallel French language mood dictionary such as CamemBERT[15] would need to be utilized to replace BERT. As well, these models are worth extending into different team sports to test effectiveness.

Bibliography

[1] Oved, N., Feder, A. and Reichart, R., 2020. Predicting In-Game Actions from Interviews of NBA Players. [online] MIT Press Journals. Available at: <https://www.mitpressjournals.org/doi/full/10.1162/coli_a_00383> [Accessed 10 January 2021].

[2] Bray, M., 2017. Text Analysis of NHL Hockey Coach Interviews · Mathieu Bray. [online] Mathieubray.com. Available at: <https://www.mathieubray.com/2017/02/11/text-analysis-hockey-coaches/> [Accessed 4 February 2021].

[3] Weissbock, J., 2014. Forecasting Success in the National Hockey League using In-Game Statistics and Textual Data. [online] Citeseerx.ist.psu.edu. Available at: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.841.8005&rep=rep1&type=pdf> [Accessed 4 February 2021].

[4] Chenyan Xu, C., 2015. Hidden In-Game Intelligence in NBA Players' Tweets. [online] Cacm.acm.org. Available at: <https://cacm.acm.org/magazines/2015/11/193325-hidden-in-game-intelligence-in-nba-players-tweets/fulltext?mobile=false> [Accessed 3 February 2021].

[5] Asapsports.com. n.d. ASAP Sports Transcripts - Hockey. [online] Available at: <http://www.asapsports.com/showcat.php?id=5&event=yes> [Accessed 15 February 2021].

[6] Pang, B., Lee, L. and Vaithyanathan, S., 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. [online] Aclweb.org. Available at: <https://www.aclweb.org/anthology/W02-1011.pdf> [Accessed 5 February 2021].

[7] Liu, Y., Sun, C., Lin, L. and Wang, X., 2016. Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention. [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1605.09090.pdf> [Accessed 5 February 2021].

[8] Witten IH, Frank E (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edition. Morgan Kaufmann, San Francisco

[9] Mohammad, S. and Turney, P., 2013. Crowdsourcing a Word–Emotion Association Lexicon. [online] Saifmohammad.com. Available at: <http://www.saifmohammad.com/WebDocs/Crowdsourcing-MohammadTurney-CI.pdf> [Accessed 6 February 2021].

[10] Schoenfeld, B., 2019. How Data (and Some Breathtaking Soccer) Brought Liverpool to the Cusp of Glory (Published 2019). [online] Nytimes.com. Available at: <https://www.nytimes.com/2019/05/22/magazine/soccer-data-liverpool.html> [Accessed 6 February 2021].

[11] Delvin, J., Chang, M., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. [ebook] Google AI Language. Available at: <https://arxiv.org/pdf/1810.04805.pdf> [Accessed 2 February 2021].

[12] Snuif, F., 2017. fnielsen/afinn. [online] GitHub. Available at: <https://github.com/fnielsen/afinn/tree/master/afinn/data> [Accessed 2 February 2021].

[13] Kim, Y., 2014. Convolutional Neural Networks for Sentence Classification. [online] New York: New York University. Available at: <https://arxiv.org/pdf/1408.5882.pdf> [Accessed 3 February 2021].

[14] YouTube. 2019. NHL players admit to giving cliché answers to the media. [online] Available at: <https://www.youtube.com/watch?v=WBnxzgZvmUE&ab_channel=NHL> [Accessed 4 February 2021].

[15] CamemBERT. 2020. *CamemBERT*. [online] Available at: <https://camembert-model.fr/> [Accessed 6 February 2021].

[16] Tamming, D., 2019. A Quantitative Study of NHL Interviews. [online] Medium. Available at: <https://medium.com/analytics-vidhya/a-quantitative-study-of-nhl-interviews-25b28821364b> [Accessed 2 February 2021].