NHL Goalie Starter Model

Sasank Vishnubhatla

November 1, 2018

Currently, success in the NHL is largely due to the success and consistency of a starting goalie. However, with the influx of new, young talent in the league, determining the starting goalie can be difficult at times. So, I have decided to work on creating a model to determine if a goalie is of starter caliber.

First, let's import the `neuralnet` package.

```
library(neuralnet)
```

Preliminary Neural Network Model

Let's import the 2015-2016 season data.

```
data20152016 = read.csv("data/goalie_stats_20152016.csv")
```

Now, let's determine which goalies during the 2015-2016 season were starters. If a goalie starts more than 40 games, then the goalie shall be considered a starter. So, in 2015-2016, notable non-starters are Johnathan Bernier and Jimmy Howard.

```
starters = list(rep(0, length(data20152016$Player)))[[1]]
starters[which(data20152016$GP >= 40)] = 1
```

Now, we can create the dataframe that we will use in our neural network.

```
dataframe20152016 = data.frame(
    s = starters,
    gsaa = data20152016$GSAA / max(data20152016$GSAA),
    dsvp = data20152016$dSv. / 100,
    ldsvp = data20152016$LDSv. / 100,
    mdsvp = data20152016$MDSv. / 100,
    hdsvp = data20152016$HDSv / 100.
)

dataframe20152016 = dataframe20152016[complete.cases(dataframe20152016),]
```

Now, with the dataframe made, we can create our first neural network.

```
features = names(dataframe20152016[2:6])
formula = paste(features, collapse = " + ")
formula = paste("dataframe20152016$s ~ ", formula)
formula = as.formula(formula)

net20152016 = neuralnet(formula,
                        dataframe20152016[2:6],
                        linear.output = FALSE)
```

Now, let's just see if it thinks Marc-Andre Fleury is a starter.

```
flowerRaw = subset(data20152016, Player == "MARC-ANDRE.FLEURY")
flowerFrame = data.frame(
    gsaa = flowerRaw$GSAA / max(data20152016$GSAA),
```

```
    dsvp = flowerRaw$dSv. / 100,
    ldsvp = flowerRaw$LDSv. / 100,
    mdsvp = flowerRaw$MDSv. / 100,
    hdsvp = flowerRaw$HDSv. / 100
)

flowerPrediction20152016 = compute(net20152016, flowerFrame)
flowerPrediction20152016$rounded = round(
    flowerPrediction20152016$net.result[1])
```

Neural Network Based on Multi-Year Data

Now, let's build a model using more recent data.

```
data20162017 = read.csv("data/goalie_stats_20162017.csv")
data20172018 = read.csv("data/goalie_stats_20172018.csv")
```

With this multi-year data, we can build a model based on 3 years of data. So, let's do that!

```
starters20152016 = list(rep(0, length(data20152016$Player)))[[1]]
starters20152016[which(data20152016$GP >= 40)] = 1

dataframe20152016 = data.frame(
    s = starters20152016,
    gsaa = data20152016$GSAA / max(data20152016$GSAA),
    dsvp = data20152016$dSv. / 100,
    ldsvp = data20152016$LDSv. / 100,
    mdsvp = data20152016$MDSv. / 100,
    hdsvp = data20152016$HDSv / 100.
)

dataframe20152016 = dataframe20152016[complete.cases(dataframe20152016),]

starters20162017 = list(rep(0, length(data20162017$Player)))[[1]]
starters20162017[which(data20162017$GP >= 40)] = 1

dataframe20162017 = data.frame(
    s = starters20162017,
    gsaa = data20162017$GSAA / max(data20162017$GSAA),
    dsvp = data20162017$dSv. / 100,
    ldsvp = data20162017$LDSv. / 100,
    mdsvp = data20162017$MDSv. / 100,
    hdsvp = data20162017$HDSv / 100.
)

dataframe20162017 = dataframe20162017[complete.cases(dataframe20162017),]

starters20172018 = list(rep(0, length(data20172018$Player)))[[1]]
starters20172018[which(data20172018$GP >= 40)] = 1

dataframe20172018 = data.frame(
    s = starters20172018,
    gsaa = data20172018$GSAA / max(data20172018$GSAA),
```

```r
    dsvp = data20172018$dSv. / 100,
    ldsvp = data20172018$LDSv. / 100,
    mdsvp = data20172018$MDSv. / 100,
    hdsvp = data20172018$HDSv / 100.
)

dataframe20172018 = dataframe20172018[complete.cases(dataframe20172018),]

dataframe20152018 = rbind(dataframe20152016,
                         dataframe20162017,
                         dataframe20172018)
```

Now, we can create the neural net.

```r
features = names(dataframe20152018[2:6])
formula = paste(features, collapse = " + ")
formula = paste("dataframe20152018$s ~ ", formula)
formula = as.formula(formula)

net20152018 = neuralnet(formula,
                        dataframe20152018[2:6],
                        linear.output = FALSE)
```

With our multi-year neural network, let's test it out on Marc-Andre Fleury again.

```r
flowerRaw = subset(data20152016, Player == "MARC-ANDRE.FLEURY")
flowerFrame = data.frame(
    gsaa = flowerRaw$GSAA / max(data20152016$GSAA),
    dsvp = flowerRaw$dSv. / 100,
    ldsvp = flowerRaw$LDSv. / 100,
    mdsvp = flowerRaw$MDSv. / 100,
    hdsvp = flowerRaw$HDSv. / 100
)

flowerPrediction20152018 = compute(net20152018, flowerFrame)
flowerPrediction20152018$rounded = round(
    flowerPrediction20152018$net.result[1])
```

So, now we can see if he is a starting goalie from both neural networks.

```r
flowerPrediction20152016$rounded
```

```
## [1] 0
```

```r
flowerPrediction20152018$rounded
```

```
## [1] 0
```

Now, if we use the 2017-2018 Marc-Andre Fleury data, we get different results.

```r
flowerRaw = subset(data20172018, Player == "MARC-ANDRE.FLEURY")
flowerFrame = data.frame(
    gsaa = flowerRaw$GSAA / max(data20172018$GSAA),
    dsvp = flowerRaw$dSv. / 100,
    ldsvp = flowerRaw$LDSv. / 100,
    mdsvp = flowerRaw$MDSv. / 100,
    hdsvp = flowerRaw$HDSv. / 100
)
```

```
flowerPrediction20152016 = compute(net20152016, flowerFrame)
flowerPrediction20152018 = compute(net20152018, flowerFrame)

flowerPrediction20152016$rounded = round(
    flowerPrediction20152016$net.result[1])
flowerPrediction20152018$rounded = round(
    flowerPrediction20152018$net.result[1])

flowerPrediction20152016$rounded
```

```
## [1] 1
```

```
flowerPrediction20152018$rounded
```

```
## [1] 1
```

Matt Murray

Preface

During the playoffs, Matt Murray has been known to be spectacular when healthy. Leading the Pittsburgh Penguins to back to back Stanley Cups in 2016 and 2017, Murray established himself as a premier goalie during his rookie and sophomore season. However, Murray has yet to prove himself as a consistent goalie during the regular season. This can be seen from his stats last year: 45 games started, 27-15 record with 3 overtime losses, 2.92 goals against average, 90.7% save percentage, and -8.11 goals saved above average. If we look at his first full season in the league, 2016-2017, Matt Murray looked like an elite goalie: 47 starts, 32-10 record with 4 overtime losses, 2.41 goals against average, 92.3% save percentage, and +14.48 goals save above average.

Whenever Matt Murray is on his game, he is dominant. However, a goalie can only do so much. The Penguins defense during the 2017-2018 season was worse than expected. This translated to Murray being put in more situations where he just could not make the save.

Analysts believe that Murray's "down" year in 2017-2018 was merely a product of his team's defensive effort in front of him. However, with a shaky start this season, it's hard to tell if that is actually true. In Murray's first two starts this year, he gave up 6 goals on 36 shots (in a winning effort) against Washington and 5 goals on 29 shots (in a losing effort) against Montreal. His save percentage after his first two starts was an abysmal 83.1%. Even worse was his goals against average: approximately 5.50. This is not starter caliber by any means.

However, Murray immediately rebounded with a convincing 38 save shutout performance in Toronto before being victorious in a shootout in Edmonton (5 goals on 46 shots). Then, he was nearly perfect in Calgary where he gave up a fluke goal on 39 shots in a dominant performance. However, in his most recent start against the New York Islanders, Murray was yanked after giving up 4 goals on 9 shots. His past start should be taken with a grain of salt as he was left out to try by his defense, who played horrifically in front of him.

With his streakiness, Murray could be a starter who plays 45 to 50 games. However, for a large percent of those, he may not show up as an elite goalie, or he is a brick wall. His consistancy lacks, though he is only in his third year as a starter.

Using the models we made before, let's see if Matt Murray has the ability to be a starter in today's NHL.

Analysis

First, we need to get Murray's data.

```
currentData = read.csv("data/goalie_stats_20182019.csv")
murrayRaw = subset(currentData, Player == "MATT.MURRAY")
murrayFrame = data.frame(
    gsaa = murrayRaw$GSAA / max(currentData$GSAA),
    dsvp = murrayRaw$dSv. / 100,
    ldsvp = murrayRaw$LDSv. / 100,
    mdsvp = murrayRaw$MDSv. / 100,
    hdsvp = murrayRaw$HDSv. / 100
)


murrayPrediction20152016 = compute(net20152016, murrayFrame)
murrayPrediction20152018 = compute(net20152016, murrayFrame)

murrayPrediction20152016$rounded = round(
    murrayPrediction20152016$net.result[1])
murrayPrediction20152018$rounded = round(
    murrayPrediction20152018$net.result[1])

murrayPrediction20152016$rounded
```

```
## [1] 0
```

```
murrayPrediction20152018$rounded
```

```
## [1] 0
```

As we see here, Murray is not starter caliber through the beginning of this young season. His percent chances of being a starter, given the neural network, are:

```
murrayPrediction20152016$net.result[1]
```

```
## [1] 0.1948361297
```

```
murrayPrediction20152018$net.result[1]
```

```
## [1] 0.1948361297
```

Comparisons to a Baseline Goalie

I believe a good baseline to measure Murray to is Devan Dubnyk. Dubnyk is the go to starter for the Minnesota Wild. His results from the networks are:

```
compareRaw = subset(currentData, Player == "DEVAN.DUBNYK")
compareFrame = data.frame(
    gsaa = compareRaw$GSAA / max(currentData$GSAA),
    dsvp = compareRaw$dSv. / 100,
    ldsvp = compareRaw$LDSv. / 100,
    mdsvp = compareRaw$MDSv. / 100,
    hdsvp = compareRaw$HDSv. / 100
)


comparePrediction20152016 = compute(net20152016, compareFrame)
comparePrediction20152018 = compute(net20152016, compareFrame)
```

```
comparePrediction20152016$rounded = round(
    comparePrediction20152016$net.result[1])
comparePrediction20152018$rounded = round(
    comparePrediction20152018$net.result[1])

comparePrediction20152016$rounded
```

```
## [1] 1
```

```
comparePrediction20152018$rounded
```

```
## [1] 1
```

```
comparePrediction20152016$net.result[1]
```

```
## [1] 0.8253619695
```

```
comparePrediction20152018$net.result[1]
```

```
## [1] 0.8253619695
```

Comparison to Murray's Backup

The previous comparison shows that Dubnyk is by far the better goalie. However, Dubnyk is a veteran while Murray is still only in his third year. A better comparison would be Casey DeSmith, the backup to Murray.

```
compareRaw = subset(currentData, Player == "CASEY.DESMITH")
compareFrame = data.frame(
    gsaa = compareRaw$GSAA / max(currentData$GSAA),
    dsvp = compareRaw$dSv. / 100,
    ldsvp = compareRaw$LDSv. / 100,
    mdsvp = compareRaw$MDSv. / 100,
    hdsvp = compareRaw$HDSv. / 100
)

comparePrediction20152016 = compute(net20152016, compareFrame)
comparePrediction20152018 = compute(net20152016, compareFrame)

comparePrediction20152016$rounded = round(
    comparePrediction20152016$net.result[1])
comparePrediction20152018$rounded = round(
    comparePrediction20152018$net.result[1])

comparePrediction20152016$rounded
```

```
## [1] 0
```

```
comparePrediction20152018$rounded
```

```
## [1] 0
```

```
comparePrediction20152016$net.result[1]
```

```
## [1] 0.3099955369
```

```
comparePrediction20152018$net.result[1]
```

```
## [1] 0.3099955369
```

Compared to Murray, DeSmith has a better percent chance of being a starting goalie on any given day. Looking at DeSmith's season totals (4 games started, 2-0 record with 2 overtime losses, 2.23 goals against average, 93.2% save percentage, and 3.55 goals saved above average), it is easy to see that DeSmith has played better than Murray.

Conclusion

Even though Matt Murray has not been up to stuff, he still is showing signs up improvement. His game is becoming more well-rounded and hopefully his streakiness will turn into consistent elite goaltending.

Caveat

Currently, the goalie who has played the most games is Anders Nilsson and David Rittich, both who have played 12 games. In statistics, and especially in sports analytics, there must be a large enough sample size for an analysis to have any bearing. So, current goalies need to have at least (this is the smallest number of starts that I am willing to use in order for any result to have any major significant) 20 games played Currently, no goalie is close to that value, so the analysis on Matt Murray with comparisons to Devan Dubnyk and Casey DeSmith are not significant. Though the analysis does shed light, it does not have complete bearing given that the season is still young. Come the end of November, when more games have been played, a more significant analysis can be done.