Evaluating an MMA Fight Markov Simulator with KMeans Fighting Style Matchups

Kodee Bonalewicz
Boston College
140 Commonwealth Avenue
Chestnut Hill, MA 02467
bonalewi@bc.edu

Benjamin Howell
howellbe@bc.edu

Daniel Howell
howelldb@bc.edu

Abstract

Our project used machine learning principles to predict outcomes of MMA matches between various details about fighters. It can be difficult to predict the outcome of a fight without large amounts of data and it may seem difficult to predict without computer models. Our goal is to use Machine Learning algorithms and practices alongside historical MMA data to predict the outcome of current MMA battles. These fight/match predictions can be an important outcome to measure the likelihood of a fighter winning a match to better help audiences understand the standings between opponents as well as help fighters know their weaknesses that they need to train to reduce. Our project will use K-means to determine style matchups for each fight, which can help determine if certain matchups are easier to predict than others. We will also be using Markov Chains to simulate the matches, where the transition probabilities are determined using fighter skill ratings, which are derived from bayesian expectation maximization procedures. Our data was compiled of results from previous matches, gathered from ufcstats.com.

1. Introduction

We would like to see if we can use machine learning algorithms and methods to model MMA matches and ultimately predict them. We decided to choose this topic because we were interested in seeing how MMA matches could be predicted since they appear to be much more unpredictable and random compared to other sports. We thought that a Markov chain would be a very good approach to this problem, since they capture complex system dynamics like what are present in MMA matches. We also thought that the k-means clustering method might help us understand stylistic matchups that lead to unpredictability and complexity in matches, in order to see what style matchups are more predictable than others.

2. Related Work

2.1. Modeling MMA with Markov Chains

This paper adopts and modifies the Markov chain model introduced by Holmes, McHale, and Żychaluk (2022) to forecast outcomes in mixed martial arts contests. Holmes et al. (2022) develop a Markov chain model to simulate mixed martial arts (MMA) contests by modeling the fight as a sequence of state transitions driven by fighter skill estimates. Rather than predicting a binary outcome, the model uses real fight statistics to estimate offensive and defensive skills—such as strike rate, takedown accuracy, submission success, and control duration—and then simulates the progression of a fight second by second. Transition probabilities are informed by Poisson, Binomial, and Gamma models fitted to historical UFC data. The chain includes absorbing states like knockouts and submissions, as well as a judging model to determine the winner if no finish occurs. The model demonstrates realistic behavior, outperforms benchmark models, and even supports profitable betting strategies when compared to bookmaker odds. We adopt their overall simulation framework and skill estimation methodology but diverge by generating skill estimates separately for each weight class, making the inclusion of weight class as a regression covariate unnecessary. In contrast, Holmes et al. pooled all fights across weight classes and included weight class as a parameter in their models.

2.2. Relevant k-means examples

David Wimser attempted to cluster fighter data through fighting styles using K-means clustering methods. He also used the public data set ufcstats.com which provided useful data to base fight style clustering on. After preprocessing and feature selection, Wismer applied K-means clustering to group fighters into distinct clusters based on their fighting styles.

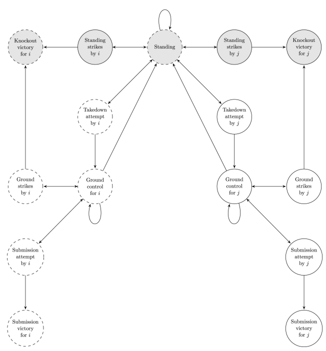

Figure A. General view of fight events. Sourced from Holmes et al. (2022).

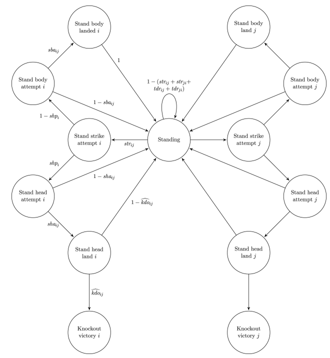

Figure B. Fight Events: Standing in depth. Sourced from Holmes et al. (2022).

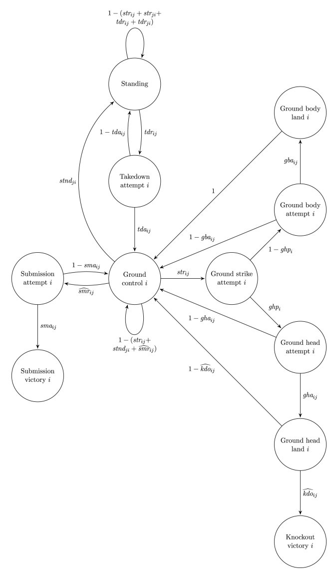

Figure C. Figure A. Ground in depth. Sourced from Holmes et al. (2022).

2.2.1. Strengths of previous k-means examples

Wismer's use of K-means clustering in this context demonstrates a practical and well-executed application of machine learning to sports data. By considering the nuances of fight dynamics, he effectively utilized clustering to uncover patterns that could inform betting strategies and deepen the understanding of fighter behaviors.

2.2.2. Weaknesses of previous k-means examples

A main issue of Wimser's findings were that he landed on using 10 clusters for fighting styles, but did not explain exact details for choosing this amount and left the cluster quantity choosing process ambiguous.

2.2.3. Our k-means interpretation

There are key points that Wismer makes in his paper that gives us a foundation for further exploration. These include defining how to cluster based of fighting styles, and some basic processes to decide on a quantity of clusters to use including

3. Methods

3.1. Overview

For our work, we determine fighting styles for different fighters, and we predict fight outcomes by running simulations for different fighter matchups. We determine the fighting styles through performing K-Means clustering on fight metrics. Fights are then simulated through Markov chains, based on generated skill estimates.

3.2. Markov Chains

Markov chains are used to simulate the outcome of a fight. A Markov chain models a stochastic process that undergoes transitions from one state to another within a finite state space, where the probability of transitioning to the next state depends only on the current state and not on the sequence of events that preceded it. Figure A briefly details the events and transitions that can occur in a fight, simplifying the standing actions and ground actions. Actions by each fighter here include standing, standing strikes, takedown attempts, ground control, ground strikes, submission attempts, submission victory, and knockout victory. Figure B describes standing actions in more detail, specifying where on the opponent a strike was attempted (head or body), and if it successfully landed or not. Figure C details possible ground actions by each fighter that can take place, such as ground strikes, where they landed and if they were successful or not. A knockout victory or submission victory are terminal states since those end the fight. Each state in the model represents a second in the actual fight. Since fights are typically 3 rounds and 5 minutes each round, we perform the chain for 900 steps to model the 900 seconds of a fight.

To find the probability of fighter i beating fighter j and

```
Optimization terminated successfully.
         Current function value: 0.368394
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 6440
Model:                          Logit   Df Residuals:                     6436
Method:                           MLE   Df Model:                            3
Date:                Tue, 13 May 2025   Pseudo R-squ.:                  0.4685
Time:                        14:09:59   Log-Likelihood:                -2372.5
converged:                       True   LL-Null:                       -4463.9
Covariance Type:            nonrobust   LLR p-value:                     0.000
=====================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
Delta_Strikes       0.0716      0.002     34.360      0.000       0.068       0.076
Delta_Takedowns     0.1334      0.018      7.275      0.000       0.097       0.169
Delta_Submissions   0.1861      0.033      5.616      0.000       0.121       0.251
Delta_Control       0.0035      0.000     16.902      0.000       0.003       0.004
=====================================================================================
```

Figure 1. Judging Model Regression Coefficients

vice versa, we simulate 2500 chains for a specific matchup, which represents 2500 possible fights that can occur. If the percentage of simulations where fighter i wins are greater than 50%, we predict a win for fighter i and a loss for fighter j. If the simulations predicting a win for i are less than 50%, we predict a loss for i, and a win for j.

A fight is determined a win or a loss either by a finish (a knockout or submission victory), or by a decision by the judges if a finish was not achieved by either fighter. To model non-finish outcomes, we created a logistic regression model that predicted a winner given previous match results that resulted in a win. For independent variables, we used the difference in the significant strikes landed, the difference in takedowns landed, the difference in submission attempts, and the difference in control time. All of these differences subtracted the winner's count from the loser's count, and the Y variable was the winner. Thus, to get the probability of a fighter winning a fight given the results for different actions, you subtract the opponent's fight action counts from theirs and plug them into the regression equation. To get the probability of the opponent winning, you subtract the probability you gained from the regression equation from 1.

We create the transition probabilities using a few probability distributions with skill estimates as inputs into them, which are based on UFC data we scraped from various UFC statistics websites mentioned in the data section. The transition probabilities are modeled using three types of probability distributions. Strike rate, takedown rate, and submission rate are modeled using Poisson distributions, as they represent count data over time. Accuracy and probability-based metrics—such as strike accuracies, takedown and submission accuracy, knockout probability, and strike location probabilities—are modeled using Binomial distributions. Ground control per takedown is modeled with a Gamma distribution, as it captures continuous, positive-valued durations, while stand-up probability is defined deterministically as the inverse of ground control time. How the skill estimates are determined will be described in the following section.

3.3. Skill Estimates

The skill estimates are estimated using Bayesian generalized linear models. They model the estimate using attack and defense parameters, which are estimated using the outcomes of previous actions, and an intercept. This is essentially a regression with two dummy variables for every single fighter where one variable is for the fighter's attacking strength, and the other is for their defending strength. If fighter I is the attacker and fighter J is the defender, then every single variable would be equal to 0 except for the variables corresponding to fighter I's attacking strength and fighter J's defending strength. These essentially would just add the coefficients together to get the skill estimate, which would then be plugged into either a binomial, poisson, or gamma distribution depending on the transition probability you are estimating. These skill estimates also add a cauchy prior to make fighters' skills converge to an average if data is lacking on their performance. The skill estimates are centered at 0 and can be positive or negative. They are derived from three different regressions: logistic regression, poisson regression, and gamma regression. The bayesglm() function from the arm package in R was used to estimate these. Fighters with no data for a certain action were removed from the skill estimate pool for that action, since the bayesglm() function requires that all the fighters be connected. Fighters who received no skill estimate for a certain action were given a default estimate of 0. Our data uses fights from 1999 until November 3rd, 2023, We split our data at the date May 5th, 2023 and used all fights before for training, and all fights after for testing. We also split each skill estimate up by weight class, as we thought that combining all of the weight classes together could potentially lead to weird results like predicting that a featherweight fighter would outstrike a heavyweight fighter, which would be impossible. We also removed female fights that were not in the strawweight class.

3.4. K-means clustering

K-Means clustering was used to organize our data from ufcstats.com into categorizable groupings. In order to make more logical predictions we needed to have our data organized into these groups based on the closeness of each of the data points relationship with each other. The average value of each cluster (our centroid values) were representatives in each data point category. For example, one of our clusters that will be discussed in the later Experiments section of the paper has a higher centroid value than other clusters in the head shots attempted/absorbed categories. Having distinguished centroid values between clustering helps us identify similarities and differences in data points which we can analyze and make further observations about in our experimentation and prediction method.

Cluster Type 1	Head Shots Landed	Head Shots Absorbed	Body Shots Landed
0	19.120412	19.002699	7.352031
1	41.142183	36.752857	14.697895
2	36.661355	12.080952	6.756227

Cluster Type 1	Body Shots Absorbed	Takedowns	Takedown Attempts \
0	7.157853	0.966645	2.512194
1	12.907257	0.764938	2.259959
2	2.756593	2.619597	5.230769

Cluster Type 1	Takedowns Absorbed	Takedowns Attempted Against \
0	0.926056	2.470272
1	1.124917	3.551214
2	0.164103	0.913553

Cluster Type 1	Ground Shots Landed	Ground Shots Attempted \
0	3.614541	5.057515
1	5.077771	7.216610
2	25.059158	35.379670

Cluster Type 1	Ground Shots Absorbed	Ground Shots Attempted Against
0	4.701668	6.625629
1	3.770766	5.330545
2	2.271429	3.283883

Figure 2. Total clustering data from our data sources labeled and divided across 3 clusters.

4. Experiments

4.1. Data

The data was scraped from ufcstats.com. The dataset includes over a dozen variables that record different events that happen in the match. The events include: shots landed and shots attempted, which are separated into head shots, body shots, leg shots as well as standing shots, ground shots and clinch shots, takedown attempts and takedown successes, reversals in ground control, total control time and knockdowns. It also includes the weight class of the fighter, the date the fight happened, and whether they are a male or female. We are using the data for k-means clustering as well as for skill estimates that will be used to determine transition probabilities on our Markov chain. We used fights from 1999 to April 2023, whereas Holmes et. al (2022) used fights from 2001 to 2017.

4.2. Clustering experiments

We were unsure about how many clusters to use for our data to have differentiated clusters that will aid our prediction model. After experimentation with different clustering amounts we landed on using 3 clusters (see Figure 2). When we used more than 3 clusters we found the centroid values to be inconclusive or harder to distinguish between other groups which defeated the purpose of using k-means clustering. For our 3 clusters, we labeled each one based on identifiable/unique centroid observations as explained below:

4.2.1. Cluster 0: Well-rounded fighters

We can see this cluster as more defensive in nature because of the overall fewer body head and body shots landed/absorbed meaning the fighters in cluster 0 are much more careful about when they attack. The fighters in this cluster have a more diverse mix of fighting style, including semi-aggressive and semi-defensive fight styles (see Figure 3).

Cluster Type 1	Head Shots Landed	Head Shots Absorbed	Body Shots Landed
0	19.120412	19.002699	7.352031
1	41.142183	36.752857	14.697895
2	36.661355	12.080952	6.756227

Cluster Type 1	Body Shots Absorbed	Takedowns	Takedown Attempts \
0	7.157853	0.966645	2.512194
1	12.907257	0.764938	2.259959
2	2.756593	2.619597	5.230769

Cluster Type 1	Takedowns Absorbed	Takedowns Attempted Against \
0	0.926056	2.470272
1	1.124917	3.551214
2	0.164103	0.913553

Figure 3. Cluster data representing cluster 0 differentiation.

Cluster Type 1	Head Shots Landed	Head Shots Absorbed
0	19.120412	19.002699
1	41.142183	36.752857
2	36.661355	12.080952

Figure 4. Cluster data representing cluster 1 differentiation.

Cluster Type 1	Ground Shots Landed	Ground Shots Attempted \
0	3.614541	5.057515
1	5.077771	7.216610
2	25.059158	35.379670

Figure 5. Cluster data representing cluster 2 differentiation.

4.2.2. Cluster 1: Aggressive fighters (strikers)

We can say that cluster 1 is a more aggressive fighter without much time on the ground because of the higher proportion of head and body shots landed/absorbed. This fighting strategy is a differentiator compared to the other clusters because of its uniquely high head shot attempted/absorption rate (see Figure 4).

4.2.3. Cluster 2: Wrestlers (ground fighters)

Cluster 2 has far greater ground shots landed and attempts indicating a fighting style that revolves more on the ground which we can cluster as wrestlers in our data set. With such large values for cluster 2 in ground values we assigned fighters in this cluster as wrestlers or ground fighters for a cluster title (see Figure 5).

4.3. Baseline and Markov Chain Results

We evaluate our Markov model and a baseline model by determining their accuracy–how many fights it correctly predicted the outcome of, out of all the fights it predicted. For a baseline, we have a logistic regression model that predicts a win or loss for fighter i against fighter j by their differences in cumulative significant strikes landed per cumulative minutes fought (CSLpCM). We use this as a baseline to see whether our k-means clustering, markov chain and skill estimates offer better predictive power than a simple logistic regression. Our Markov chain method had an accuracy of 56.83%. This is negligibly better than our baseline result, which is 56.68%. This poor improvement could be due to methodological modifications, such as only determining skill estimates within weight classes and not across them. Doing this limited the number of fights that are available to determine the skill estimates. We found that the k-means clusters are inconclusive when compared with the predictions from our markov model. We think that they best serve as a qualitative heuristic that someone wanting to predict an MMA fight can use to aid their thought process and predictions.

5. Conclusions

It is possible that 2500 chains per fight is not enough to determine a fight outcome. It could be explored whether more chains, say, 10,000, increase accuracy. We could also add a prior to the estimates that would make them time-varying. This would make skill estimates put more weight on more recent fights than older ones. Lastly, we could quantitatively compare our k-means with our markov chain results. When experimenting we found that the k-means style matchups do not currently tell us much about our predicted wins from the markov chain, however we think that if these were adjusted to have better and more interpretable clusters then they could possibly do so.

6. Contributions

The contributions of each member and Github Repo are below:
Kodee Bonalewicz - Creation and analysis of k-means clustering categories, Team management. Wrote about k-means clusters in section Related work, Methods, Experiments, Abstract and Bibliography References.
Benjamin Howell - Programmed Baseline, Markov Chains, and the Poisson, Binomial, and Gamma Probability Distributions. Discussed Holmes et al. (2022) in Related Work, Described Markov Chains in methods, and reported their results.
Daniel Howell - Created the skill estimates using R, retrieved the data from ufcstats.com and modified it, and created the judging model. Also wrote the introduction, conclusions, and part of the methods and experiments sections.
Zephan Keach - Discussed data and reviewed model accuracy.
Github Repo - https://github.com/howellbe02/Team7-ML-Project

References

[1] Benjamin Holmes, Ian G. McHale, and Kamila Żychaluk. A markov chain model for forecasting results of mixed martial arts contests. *International Journal of Forcasting*, 39(2):623–640, 2023.

[2] UFC. Ufc stats, 2025. http://ufcstats.com/statistics/events/completed.

[3] David Wismer. Clustering ufc fighters by fighting style. *A Project Borne of Meaningless Bets Among Friends*, 1(1):1–2, 2022.