A Topic Model Approach to Represent and Classify American Football Plays

Jagannadan Varadarajan[1]
vjagan@adsc.com.sg
Indriyati Atmosukarto[1]
indria@adsc.com.sg
Shaunak Ahuja[1]
shaunak@adsc.com.sg
Bernard Ghanem[12]
bernard.ghanem@kaust.edu.sa
Narendra Ahuja[13]
n-ahuja@illinois.edu

[1] Advanced Digital Sciences Center of Illinois, Singapore
[2] King Abdullah University of Science and Technology, Saudi Arabia
[3] University of Illinois at Urbana-Champaign, IL USA

Abstract

We address the problem of modeling and classifying American Football offense teams' plays in video, a challenging example of group activity analysis. Automatic play classification will allow coaches to infer patterns and tendencies of opponents more efficiently, resulting in better strategy planning in a game. We define a football play as a unique combination of player trajectories. We develop a framework that uses player trajectories as inputs to MedLDA, a supervised topic model. The joint maximization of both likelihood and inter-class margins of MedLDA in learning the topics allows us to learn semantically meaningful play type templates, as well as, classify different play types with 70% average accuracy. Furthermore, this method is extended to analyze individual player roles in classifying each play type. We validate our method on a large dataset comprising 271 play clips from real-world football games, which will be made publicly available for future comparisons.

1 Introduction

Any group activity analysis solution requires modeling both temporal and spatial relationships amongst interacting members. Automated understanding of group activity would prove useful in many domains such as surveillance, retail, health care, and sports. Despite active research in automated activity analysis, the sports domain is extremely under-served. One particularly difficult but significant sports application is the automated labeling of plays in American Football ("football"), a sport in which multiple players interact on a playing field during structured time segments called plays. Football plays vary according to the initial formation and trajectories of players - making the task of play classification quite challenging.

In this paper, we address the combined problem of automatic representation and classification of football plays as summarized in Fig. 1. The motivations to develop an automated

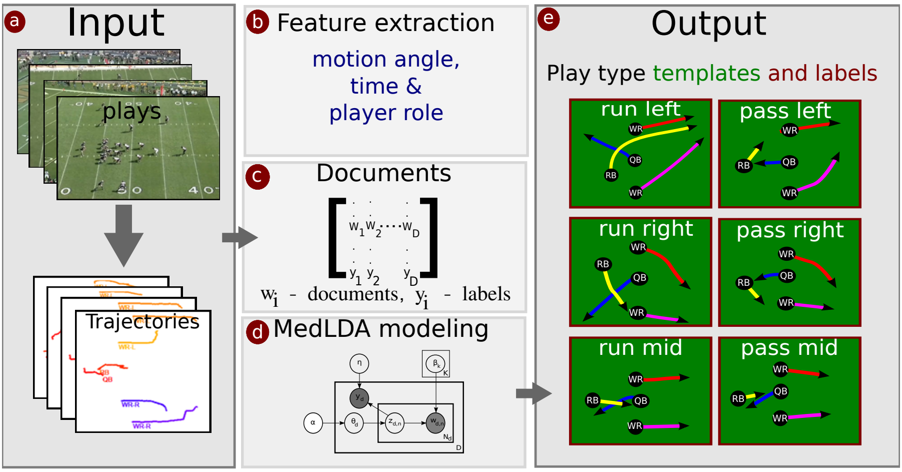
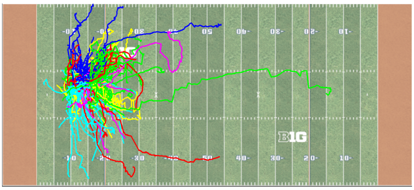


Figure 1: The goal of our work is to represent and classify play types in football videos. Given football plays, our framework consists of a) player tracking, b) feature extraction, c) document creation and d) modeling using MedLDA [22], resulting in e) two outputs: play type templates and labels. The templates shown in (e) are typical trajectories of four players (QB, RB, WR-L and WR-R).

Figure 2: Trajectories of quarterback (QB) mapped onto the field and colored corresponding to play type, illustrating significant trajectory variation within and across play types.



classification system are three-fold: **(i)** it would greatly help coaches and broadcasters efficiently analyze large collections of video and better understand team strengths and weaknesses; **(ii)** understand and evaluate the contributions of every player on the field; and finally, **(iii)** generate a collection of all plays executed by different teams and learn play tendencies in different situations. For instance, Team A can input Team B's videos and quickly understand commonly chosen play types and underlying strategies of Team B.

Developing a vision based play type classification system poses significant hurdles. It is difficult to track players due to high occlusion and irregular player motion in a dynamic scene. Even with accurate player trajectories, the system must classify plays with low inter-class and high intra-class trajectory variation. These plays are intentionally designed to appear similar across play types. For example, Fig. 2 shows all the trajectories of a single player, illustrating high variation in players' trajectories.

We define a play as a unique combination of offense players' trajectories. This dynamic group activity (play) is labeled by play type (run or pass) and direction (left, middle, or right). A *run* play is simply an attempt to advance the ball via running with the ball, while a *pass* play is defined as an attempt to advance the ball by throwing the ball. A *run* is considered left/middle/right if the player attempting to advance the ball runs to the left/middle/right of the offensive line. A *pass* play is considered left/middle/right if the ball is thrown to the left/middle/right third of the field. An offense consists of 11 players of which six play a critical role: quarterback (QB), running back (RB), wide receiver left (WR-L), wide receiver right (WR-R), tight end left (TE-L) and tight end right (TE-R). The trajectories of these six players are input to our system (Fig. 1(a)).

Given a play (short video clip), we consider two problems: 1) *representation*, *i.e.*, derive canonical templates (common routes taken by players) for each play type, and 2) *classification*, *i.e.*, predict a label from one of the six play types for a given play, as shown in Fig. 1(e). The above mentioned player trajectories allow analysis of players and their importance in a play type, both individually and collectively. In section 5.2, we present an experiment proving that trajectories contain vital clues for play type classification. [1]

Approach and Contributions. Our approach to modeling play types from trajectories builds on recent success of supervised topic models (STMs) such as SLDA [3] and MedLDA [22]. This approach overcomes significant trajectory variations in shape as demonstrated in Fig. 2. STMs capture dominant co-occurrent patterns, called “topics”, in data discriminatively. Intuitively, such patterns correspond to unique player interactions. For instance, run plays usually involve an intersection of QB and RB. STMs are effective in learning such salient co-occurring actions while achieving good classification performance.

The contributions of this paper are as follows: **(1)** we propose a framework to analyze football plays based on player trajectories that combines the objective of a) deriving semantically meaningful play type templates and b) achieving good classification performance. The framework uses the MedLDA [22] model on documents derived from play clips. We identify semantically meaningful templates and achieve high classification performance, thanks to the MedLDA optimization procedure that combines both maximum likelihood estimation and maximum margin classification. **(2)** We treat each trajectory as a document and analyze the importance of each player in each play type. **(3)** Finally, to our knowledge there is no standardized dataset to compare different methods developed for football play analysis. Thus, to encourage further work on this topic, we release a completely annotated football trajectory dataset we used, containing 271 plays, on [1].

2 Related Work

We compare our approach to previous work in three related domains: activity analysis, football play classification, and topic models.

Activity Analysis. Recently, several approaches were proposed to address group activity analysis which are broadly categorized as either *pixel centric* or *object centric*. Some examples of pixel centric methods include spatio-temporal interest points (STIP) used to capture human interactions in [13, 15], 3D histogram of gradients [10], quantized optical flow [17] and foreground pixels [12] used in discovering activity patterns from crowded surveillance videos. While pixel centric features can be readily extracted from videos, they do not preserve object identities and suffer from ambiguities due to presence of multiple objects in the scene. An alternative approach is to use object trajectories [9, 11, 21]. The extracted trajectories are then used to model long-term motion patterns. Although they can suffer from tracking errors and occlusions, trajectories help us isolate individuals from rest of the scene.

Play type classification. There is little previous work on football play classification in video. In [9] , trajectory motion of players are characterized using a probabilistic generative model to recognize offensive play strategies. Similarly, a manifold representation is derived from

[1] Tracking the ball carrier is a better indicator of play type, however, it is extremely difficult.

positions of the players in [11]. Both of these works rely on manually annotated trajectories. In [16], each play type is modeled using kernel density estimation on the player position with the temporal relations modeled using a non-stationary Hidden Markov Model. None of the above methods focus on semantic representation of play types, instead focus mainly on classification. Furthermore, these methods do not provide scope to analyze play types based on individual players and are not scalable to large datasets.

Topic Models. Topic models such as Latent Dirichlet Allocation (LDA) [4] are unsupervised text mining methods aimed at capturing latent themes from large collections of documents. Lately, they have been successfully adapted for vision tasks such as scene classification [19] and activity analysis [6, 17]. Using conventional topic models for classification typically involve learning the topics first, followed by a supervised learning step with the topic weights as features. Since the two steps are performed independently, this procedure results in poor classification performance. Recently developed supervised topic models (STMs), such as Supervised LDA [3] and Maximum entropy discriminative LDA (MedLDA) [22], incorporate class labels in the topic learning process to improve prediction performance as well as discover salient topics.

Our approach to play type classification is driven by three goals: **(i)** extract latent themes to create play type templates from player trajectories; **(ii)** classify play types with high accuracy, and **(iii)** understand typical routes of players, their roles and importance in a play type. We build on the success of STMs and use them to learn and classify play types. In section 5, we show that the learned topics indeed capture common routes taken by the players in a play.

3 Problem definition and Model overview

In this section, we formally state our problem and review the basic concepts of MedLDA model. We then detail how we apply MedLDA to player trajectories.

Problem definition. We are given D plays (or training clips) denoted by $\{v_d\}, 1 \leq d \leq D$, where a play v_d is represented as a set of trajectories from R players: $\{X_r^d(t) : 0 \leq t \leq T, 1 \leq r \leq R\}$, T is the maximum possible duration of a trajectory, R is the maximum number of tracked players, and each play, indexed by d, is also associated with a unique class label $y_d \in \mathcal{C} = \{1, 2, \cdots, C\}$ called a play type. Our objective is to learn a template τ_y corresponding to each class y, and a classifier function $F(v_d)$ that predicts a label $\hat{y}_d$ with high classification accuracy.

3.1 Supervised Topic Models

Topic models such as pLSA and LDA [4, 8] are mixed membership models that assume documents as a mixture of many topics, where each topic is represented as a multinomial distribution over the vocabulary. In case of LDA, the inputs are the document corpus represented as a document count matrix. The outputs are the discovered topics and their respective weights in the training documents, where topic weights serve as a lower dimensional representation of the original documents.

Supervised topic models like SLDA [3] and MedLDA [22] introduce a response variable y to each document as shown in the graphical model in Fig. 3. To motivate the use of the MedLDA model, we briefly review its predecessor, the SLDA model, and highlight the

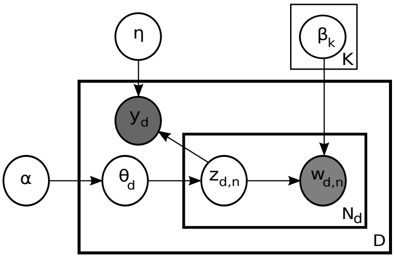


Figure 3: Supervised LDA model [3]. The MedLDA model follows the same generative process except for predicting y_d for classification.

improvements made in MedLDA. Let D be the number of documents[2] each containing N_d words from a vocabulary of V terms. Let K be the number of topics and β the topics matrix, where each topic β_k is a multinomial distribution over the V terms. The generative process to realize each document $\mathbf{W}_d$ in our dataset $\mathcal{D} = \{\mathbf{W}_1, \mathbf{W}_2, \cdots \mathbf{W}_D\}$ is as follows:

1. draw topic proportions $\theta_d|\alpha \sim \text{Dirichlet}(\alpha)$, where θ_d is a multinomial distribution parameter and α is the parameter for a Dirichlet distribution;
2. for each word $w_{d,n}$ in $\mathbf{W}_d = \{w_{d,n}\}_{n=1}^{N_d}$:
   - draw a topic assignment $z_{d,n}|\theta_d \sim \text{Multi}(\theta_d)$;
   - draw a word $w_{d,n}|z_{d,n} \sim \text{Multi}(\beta_{z_{d,n}})$;
3. draw a response variable: $y_d|z_{d,1:N_d}, \eta, \delta^2 \sim \mathcal{N}(\eta^{\mathrm{T}}\bar{z}, \delta^2)$; where $\bar{z} = \frac{1}{N_d}\sum_{n=1}^{N_d} z_{d,n}$.

The MedLDA model follows the above generative process to draw $(\theta, z, \mathcal{D})$. It deviates from SLDA in obtaining the labels y, that is, instead of drawing the labels from a normalized distribution as in SLDA, MedLDA uses a max margin learning framework to predict the labels. Given the latent topic assignments $z_{1:N_d}$, MedLDA learns a discriminant function in the form $F(y, z_{1:N_d}, \eta) = \eta_y^{\mathrm{T}}\bar{z}$ where η_y is a class specific K dimensional vector associated with class y and η is a vector obtained by stacking $\eta_y, y \in \mathcal{C}$. Note that $z_{1:N_d}$ act as our features here. The parameters of the model are obtained by optimizing Eq. 1, which combines both maximum likelihood and maximum margin estimation.

$$\min_{q,q(\eta),\alpha,\beta,\xi} -\mathcal{L}^u(q;\alpha,\beta) + ||\eta||^2 + \kappa\sum_d \xi_d \tag{1}$$
$$\forall d, y \in \mathcal{C}, \text{s.t.} \left\{\eta_{y_d}^{\mathrm{T}}\mathrm{E}(\bar{Z}_d) - \eta_y^{\mathrm{T}}\mathrm{E}(\bar{Z}_d) \geq \Delta(y, y_d), \xi_d \geq 0\right.$$

In Eq. 1, q refers to the variational posterior distribution $q(\theta, z)$, $\bar{Z}$ is the random variable corresponding to $\bar{z}$ and $E[.]$ is taken over variational distribution $q(\cdot)$. The first term $\mathcal{L}^u(q)$ in the objective function is the variational lower bound on $\log P(\mathcal{D}|\alpha,\beta)$, which is the log likelihood of the data. This is same as in unsupervised LDA (cf. [4] for details). κ is a positive regularization constant that is typically used in SVM learning framework. $\Delta(y, y_d)$ is the loss function indicating the cost of misclassifying y_d to be y. This is typically the $0/1$ loss function, *i.e.* $\Delta(y, y_d) = 1$ if $y \neq y_d$ and 0 otherwise. ξ_d are the slack variables corresponding to each document. For more details, we refer the reader to [4, 19, 22]. Looking at the objective function, we see that the first term maximizes the data likelihood while the other terms resemble a typical maximum margin classification objective such as in SVM. Intuitively, this can be considered as a regularized maximum margin learning where the regularization comes from the data likelihood term. This combined learning paradigm enables us to extract salient topics with good classification accuracy, as seen in Section 5.

[2] In our case a document represents a play and hence we use the same notation here.

3.2 Analyzing football plays using MedLDA

The process of applying MedLDA to football play analysis consists of three parts: *player tracking*, *vocabulary creation* and, *document creation* using topic models.

Player Tracking: Initializing trajectories automatically is challenging as it needs recognizing the initial formation and player positions robustly [2]. In our work, we obtain trajectories by initializing a Multi Object Tracker (MOT) [20] with bounding boxes around six offense players (QB, RB, WR-L, WR-R, TE-L, TE-R (cf. Section 1)) at the start of the play clip. The six players are tracked till the end of the play clip or until they disappear from the field of view. The trajectory observations from player tracking is mapped back to the field image using registration parameters obtained from the method in [7]. The trajectories of the six tracked offense players are the input to our framework.

Vocabulary Creation: In our case, a word w in the vocabulary is defined by three different aspects of trajectories namely: motion direction, time and player role. **(1)** Play types in football are determined by distinct motion directions and routes taken by different players. In order to capture this information, we measure motion directions from observations in the trajectory. The motion directions ranging between $0^\circ - 359^\circ$ are quantized into 18 equal sized directional bins of 20 degrees each. **(2)** Play types differ largely based on how trajectories evolve over time. To capture temporal information, we group observations from every consecutive 10 frames into a single time step. Thus, each observation is also associated with a time-stamp based on its time of occurrence from the beginning of the play. The maximum duration of a play in our dataset is 250 frames, thus we have 25 different temporal bins. Note that in this temporal quantization, longer trajectories result in more words than shorter trajectories[3]. **(3)** To understand the participation of each player in a play type, the observations from a player's trajectory are associated with the player role, *e.g.*, QB, RB, etc. In our experiments, we use trajectories of six different players in creating our vocabulary (cf. Section 1).

Therefore, using the aforementioned quantization steps, each observation from the trajectory results in a word w represented as a triplet $w = (v_m, v_t, v_r)$, where v_m is the index of the motion bin, v_t is the temporal bin index, and v_r the player role. Thus each trajectory can be summarized using a table of $18 \times 25 = 450$ bins, and the vocabulary is given by $450 \times 6 = 2700$ words.

Document Creation: Each document in MedLDA is a play clip spanning a single offense play. These documents are given by word counts, $\mathbf{W}_d = \{n(w_i)\}_{i=1}^{V}$, *i.e.*, the number of times each word w_i appears in the document. Additionally, the document also has a label y_d corresponding to one of the six play types (cf. Fig. 4(a)). Note that our document creation process assumes that the individual players are independent given a topic. Nevertheless, word co-occurrences within a document implicitly captures interactions among the players.

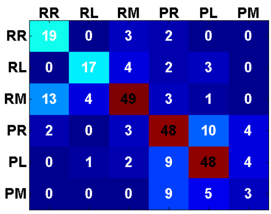
4 Dataset

There is currently no publicly available dataset containing football video play clips and corresponding player trajectories. To test the efficacy of our method in analyzing football plays, we compiled a trajectory dataset from 3 football games containing 271 play clips. Trajectories of the six players on offense were extracted using a MOT tracker. The play types and trajectories were labeled by a sports expert. The plays were categorized on two levels. At the top level, the plays were labeled as either *run* or *pass* plays. At the second level, the labels

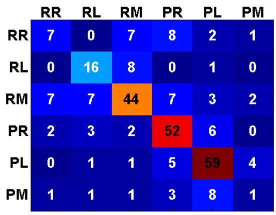
[3]Since football plays are of short duration, we do not explicitly model variations in speed.

Run			Pass		
Right (RR)	Left (RL)	Middle (RM)	Right (PR)	Left (PL)	Middle (PM)
24	27	70	67	64	19

(a)



(b)



(c)

Figure 4: a) Class hierarchy and distribution of samples in our dataset; note its highly unbalanced nature; b) confusions from sports expert, c) confusions from our proposed model.

were determined by the direction of the play: *left, middle, right*. In total, our dataset contains 121 run plays and 150 pass plays. The class distribution of the dataset is unbalanced, Figure 4(a) shows the number of instances in each class. This dataset of annotated football trajectories will be made publicly available for future comparison [■].

5 Experimental Results

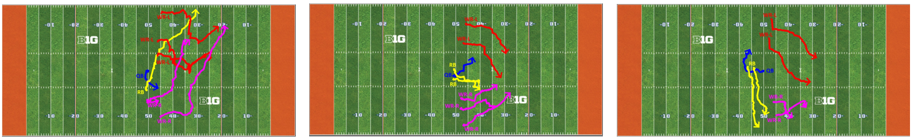
In this section, we evaluate our approach for both play type representation and classification.

5.1 Qualitative analysis - play type representation

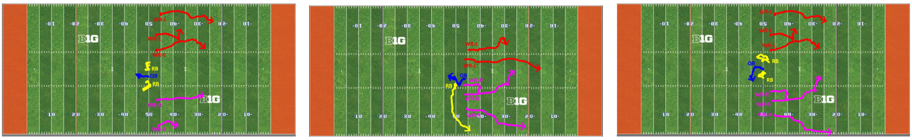
Our objective here is to extract templates corresponding to each play type by applying the MedLDA model on our document collection consisting player trajectories. Our method to select these templates is based on the topic weights obtained for documents of each class. The template τ_y for class y is selected as the most representative topic z out of the K topics obtained from MedLDA. Mathematically this is given by, $\tau_y = \arg\max_z \sum_{\{d|y_d=y\}} P(\theta_d|\mathbf{W}_d, \alpha)$, where $P(\theta_d|\mathbf{W}_d, \alpha)$ is the posterior Dirichlet estimate obtained for each document d after applying MedLDA. Given that a play type template is a topic, *i.e.* multinomial distribution over words $P(w|z)$, we can analyze the results qualitatively by generating trajectories from $P(w|z)$ or by selecting the best matching trajectory from the training set for each player.

We adopt the latter approach and select the closest trajectory for each player in a template by the following method. Given the set $\mathcal{T}_{r'}$, of all trajectories for player r' in the training set, we first generate a discrete representation $\hat{X}_{r'}(t)$ for each trajectory $X_{r'}(t) \in \mathcal{T}_{r'}$ using the quantization steps described in section 3.2. The closest matching trajectory for the player in the template is the one with highest score for the normalized cross-correlation between $\hat{X}_{r'}(t)$ and $P_{r'}$, where $P_{r'} = P(\cdot,\cdot,v_r = r'|\tau_y)$.

The templates for all (six) play types, generated from topics learned by applying MedLDA, are shown in Figs. 5(a–f). These templates highlight several salient aspects of the play types. Firstly, from the trajectories of QB and RB, we note that run plays in Fig. 5(a,b,c) are fundamentally different from pass plays in Fig. 5(d,e,f). In run plays, the QB typically moves back to hand over the ball to the RB who then takes it forward, resulting in an intersection of the blue and yellow trajectories. Whereas in pass plays, there are no such intersections as the ball is thrown to one of the WRs (in red or magenta). The play types are determined by the direction (left, middle, right) in which the ball is taken forward. In run left/mid/right (cf. Fig. 5(a,b,c)), the RB's trajectory in yellow indicates this difference. For instance, in



(a) Run left (b) Run mid (c) Run right



(d) Pass left (e) Pass mid (f) Pass right

Figure 5: Templates generated for all play types. Each trajectory color represents a player role: QB-blue, RB-yellow, WR Right-magenta and WR Left-red. Tight ends are not shown here as they have very low weights in the topic.

Fig. 5(b), we see the RB going straight between the hash-line marks in the field whereas in Fig. 5(c) the RB moves to the right. For pass plays, this distinction is mainly due to the WR's trajectory direction as shown in Fig. 5(d,e,f). Thus, these templates capture salient aspects of the play types well, despite large intra-class differences. In the next section, we show that our sports expert confirms these observations.

5.2 Quantitative analysis - play type classification

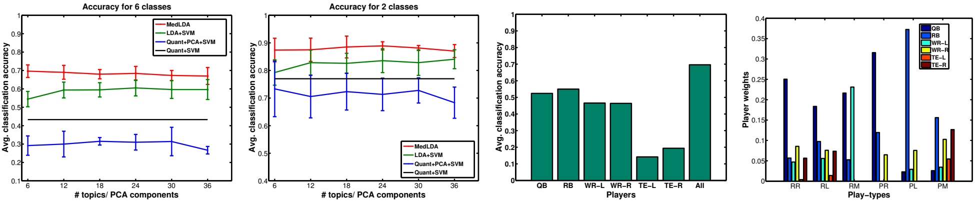
We compare our supervised topic model approach for play type classification against three baseline methods.

Baseline 1: This baseline is inspired by the bag-of-words+SVM approach for action classification as in [14]. Here, the raw word counts from the documents are directly used as features to learn a multi-class SVM classifier [5] with RBF kernel. We refer to this method as Quant+SVM.

Baseline 2: Documents represented as raw word counts lie in very high dimensions, *i.e.* 2700 dimensions. We applied PCA to reduce the dimensions and the top K principal components were selected to create our feature set. Therefore, every training (and test) sample was projected onto the subspace spanned by the selected principal components and their coefficients were used as features to train a SVM classifier with RBF kernel. We refer to this as Quant+PCA+SVM.

Baseline 3: This baseline employs the unsupervised LDA method. Here, we first applied LDA on training documents from plays with K number of topics. The topics weights $P(z|W_{train})$ from training documents were then used as features to learn a multi-class SVM classifier. For the testing phase, the topic weights $P(z|W_{test})$ from a folding-in procedure form our test samples. This procedure is referred to as LDA+SVM.

We compare our method with the three baselines for the task of 6 class and 2 class (run vs pass) classification tasks for varying number of principal components (in baseline 2) or number of topics in LDA+SVM and MedLDA, *i.e.*, $K = 6, 12, .., 36$. In all our experiments, SVM regularization parameters for the baselines (and κ in MedLDA) were chosen using a cross validation data within the training set. For each model size, the average accuracy was



(a) 6 classes (b) 2 classes (c) MedLDA per player (d) MKL weights

Figure 6: Average classification accuracy of our method versus three baselines in predicting a) six classes and b) two classes (run vs pass); (c,d) Studying importance of players: c) classification performance obtained on documents created from a single player. d) player weights from MKL obtained by training one-vs-all classifiers for each play type.

then obtained by five-fold cross validation. Figure 6(a) and (b) show the performance on the 6 class and 2 class tasks. Firstly, we note that the simplest baseline method Quant+SVM achieves an average accuracy of 43% on six classes and 76.7% on two classes, which is better than a random guess classifier that gives 16.7% or 50% for 6 and 2 classes respectively. The second baseline, Quant+PCA+SVM with $K = 6, 12, \cdots, 36$, results in only 30% accuracy. However, when we consider 193 (out of 2700) principal components covering 95% of the training data variance, we achieve 43.3% and 78% accuracy for 6 and 2 classes respectively, which is comparable to Quant+SVM. This shows that large variations exist in the data that could not be captured with few principal components. The third baseline LDA+SVM achieves 60% and 82% for the 6 and 2 classes respectively. Classification accuracy improves as the number of topics is increased, saturating at 60% for the 6-class classification. This shows that LDA (co-occurrence) based dimensionality reduction is more effective in modeling play types. Finally, our approach based on MedLDA achieves 70% accuracy on 6-class and 88% on 2-classes, outperforming all three baselines.

Expert Annotation: To test our hypothesis that player trajectories have vital cues to describe a play type, we engaged a sports expert (with 7 years of experience in the sport) to label the plays by viewing *only* videos of player trajectories with no access to the original play clip. The expert was given videos displaying the trajectories of the six offense players. Each player trajectory was given a unique color (as in Fig. 5) to identify their roles. The expert was then asked to provide up to three guesses for each play. The confusion matrix of the expert's guesses is shown in Fig. 4(c). The average accuracy based on the expert's first guess is 68.7% for 6 classes and 92.9% for run vs pass. This shows that it is possible to differentiate a run vs pass (2 class) quite accurately using only trajectories while predicting the 6 classes is more challenging. Lastly, comparing the confusions of the expert with MedLDA in Fig. 4(b,c), we see that MedLDA classifies marginally better (70%) than the expert while mostly agreeing on 4 of 6 classes. The "pass-mid" class is predicted with the lowest classification accuracy because it shares many similarities with pass left and pass right.

5.3 Discriminative Players

In addition to classifying plays, we investigate which of the player roles is most discriminative in general and for a given play type in particular. First, to analyze the importance of each player, we apply the MedLDA model on documents created from only a single player. In other words, to analyze player r', document $\mathbf{W}_d$ is created from trajectory $\{X^d_{r=r'}(t)\}^T_{t=1}$. MedLDA is then applied by setting $K = 6$ for each player to classify play types. The bar-plot in Fig. 6(c) shows the average classification accuracy obtained for each player. Clearly, QB and RB give the highest accuracy. This is because the actions of QB and RB largely determine if the play is a run or pass. The WRs also give 40%, since their roles are discriminative on pass plays. TEs give poor classification results because their trajectory variation is very low across all play types. Lastly, we also note that combination of all the players, as shown in Fig. 6(a), is the most discriminative.

We also investigate whether specific players are more discriminative in classifying certain play types. We address this question by learning a one-vs-all classifier for each play type using per-player topic weights as features to Multiple Kernel Learning (MKL) [18]. The intuition is that the optimal set of weights for each player, learned using MKL will be a good indicator of the player's importance in classifying a play type. To do this, we first apply the MedLDA model on documents created from individual players with K=6 for each player as described in the previous paragraph. MKL is then applied on documents described by per-player topic weights. In other words, the feature set used to train an MKL classifier is given by $\mathcal{Z} = \{\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_R\}$, where $\bar{z}_r$ are the topic weights corresponding to player r. The weights learned for each player and for each play type using MKL are shown in Fig. 6(d), where higher weights correspond to the most discriminative players. From the bar-plot for each play type, we see that the QB and RB are the most discriminative players in run plays. WRs tend to have higher weights in pass plays than run plays. In pass plays, their trajectories vary whereas in run plays they typically block. We see that TEs do not play a major role, except in pass mid plays. In these plays, their trajectory variation increases because they are expected to both receive the ball and block.

6 Conclusion

In this paper, we address the problem of representing and classifying football plays. We apply a max-margin based supervised topic model on plays, where player trajectories are broken down into words. We learn topics that resemble common player trajectories in six play types. Our qualitative and quantitative results (average accuracy of 70% in classifying six classes, and 88% on two classes (run vs pass)) clearly support our claim that player trajectories are powerful in representing and classifying play types. Finally, we learn that QB and RB are very discriminative in play type classification.

7 Acknowledgements

This study is supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sciences Center from Singapore's Agency for Science, Technology and Research (A*STAR).

References

[1] http://publish.illinois.edu/videoadsc/publications/football-trajectory-dataset/.

[2] I. Atmosukarto, B. Ghanem, S. Ahuja, K. Muthuswamy, and N. Ahuja. Automatic recognition of offensive team formation in american football plays. In *1st IEEE International Workshop on Computer Vision in Sports*, 2013.

[3] David M. Blei and Jon D. McAuliffe. Supervised topic models. In *NIPS*, 2007.

[4] David M Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *JMLR*, 3(4-5):993–1022, 2003.

[5] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[6] R. Emonet, J. Varadarajan, and J.-M. Odobez. Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model. In *CVPR*, 2011.

[7] Bernard Ghanem, Tianzhu Zhang, and Narendra Ahuja. Robust video registration applied to field-sports video analysis. In *ICASSP*, 2012.

[8] T. Hofmann. Unsupervised learning by probability latent semantic analysis. *JMLR*, 42: 177–196, 2001.

[9] Stephen Intille and Aaron F. Bobick. Recognizing planned, multiperson action. *CVIU*, 81:441–445, 2001.

[10] Alexander Kläser, Marcin Marszałek, Cordelia Schmid, and Andrew Zisserman. Human focused action localization in video. In *ECCV-(SGA)*, 2010.

[11] Ruonan Li, R. Chellappa, and S.K. Zhou. Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition. In *CVPR 2009.*, pages 2450–2457, 2009.

[12] Chen Change Loy, Tao Xiang, and S. Gong. Multi-camera activity correlation analysis. In *CVPR*, 2009.

[13] Karthik Prabhakar, Sangmin Oh, Ping Wang, Gregory Abowd, and James Rehg. Temporal causality for the analysis of visual events. In *CVPR*, 06 2010.

[14] Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing human actions: A local svm approach. In *ICPR*, 2004.

[15] Behjat Siddiquie, Yaser Yacoob, and Larry S. Davis. Recognizing plays in american football videos. Technical report, University of Maryland, 2009.

[16] Eran Swears and Anthony Hoogs. Learning and recognizing complex multi-agent activities with applications to american football plays. In *WACV*, 2012.

[17] J. Varadarajan, R. Emonet, and J.-M. Odobez. A sequential topic model for mining recurrent patterns from long term video logs. In *IJCV*, 2012.

[18] Manik Varma and B. R. Babu. More generality in efficient multiple kernel learning. In *ICML*, 2009.

[19] Yang Wang and Greg Mori. Max-margin latent dirichlet allocation for image classification and annotation. In *BMVC*, pages 112.1–112.11, 2011.

[20] Tianzhu Zhang, Bernard Ghanem, and Narendra Ahuja. Robust visual tracking via multi-task sparse learning. In *CVPR*, 2012.

[21] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. Random field topic model for semantic region analysis. In *CVPR*, 2011.

[22] Jun Zhu, Amr Ahmed, and Eric P. Xing. Medlda: Maximum margin supervised topic models. *JMLR*, 13:2237–2278, 2012.