**Математическая оценка эволюции стиля игроков Барселоны**

Теги [Пеп Гвардиола](https://www.sports.ru/josep-guardiola/) [Алексис Санчес](https://www.sports.ru/alexis-sanchez/) [Луис Энрике](https://www.sports.ru/luis-enrique/) [Хави](https://www.sports.ru/xavi/) [Барселона](https://www.sports.ru/barcelona/) [Андрес Иньеста](https://www.sports.ru/iniesta/) [Лионель Месси](https://www.sports.ru/messi/) [Футбол](https://www.sports.ru/football/)

В большинстве клубов за 10 сезонов меняется все по несколько раз и без какой-либо структуры. Барселона, в каком-то смысле, идеальный клуб для анализа - с начала карьеры Месси и до сезона 15/16 она прошла путь от, возможно, лучшего позиционного футбола (Гвардиола) с небольшим вступлением в виде Райкарда до элитного контратакующего (Энрике). И хочется сказать, что переход был достаточно плавным. Я не являюсь экспертом по игре Барселоны (тем более, я начал осознанно смотреть футбол года с 2015-го), поэтому буду исходить от результатов модели. В конце выложу максимум визуализаций, чтобы все желающие смогли поправить мою интерпретацию или добавить что-то, чего в тексте не будет.

Стоит сразу отметить два момента:

* модель не скажет - какой игрок стал лучше или хуже, потому что она показывает стиль паса игрока (т.е. из каких зон продвигает, тип передач, количество и т.д.), но поможет проследить изменение стиля или, возможно, позиции;
* неправильно делать выводы о стиле игрока исключительно по тому, как он пасует (но именно об этом весь этот текст);

**Как это работает**

Сначала я в двух словах опишу модель и способ ее применения в этой задаче, затем подробнее остановлюсь на применении. Подробное описание модели и ссылку на ее реализацию можно найти в [первом тексте](https://www.sports.ru/tribuna/blogs/divfootballfield/2526764.html) в этом блоге. Модель использует [открытые данные StatsBomb](https://statsbomb.com/2019/07/welcome-to-the-messi-data-biography/).

Принцип обучения заключается в том, чтобы для данных по каждому пасу (координаты, id части тела, длина, высота и т.д.) модель пыталась понять кто из игроков отдал этот пас. После обучения берутся ее веса и с помощью несложных преобразований получается вектор для каждого игрока, описывающий тип передач, которые отдает этот игрок. Отдельно в компонентах вектора смысла нет (или его очень сложно понять), но в сравнении с векторами других игроков складывается интересная картина. Модель впервые была [представлена](https://www.youtube.com/watch?v=t1y5incr5Gw) на OptaPro форуме 2018 Бенжамином Товарни.

Для обучения такой модели нужны данные об очень большом количестве пасов для каждого игрока. Зная все это способ применения напрашивается: обучить модель на каждом сезоне Ла Лиги с 04/05 до 15/16 и сравнить получившиеся вектора игроков. На этом моменте можно пропускать часть с подробным объяснением и переходить к анализу результатов и визуализациям.

Подробную часть начну с того, что на деле модель не будет обучаться заново на каждом сезоне - она обучится на одном и далее будет дообучаться на остальных по очереди, сохраняя промежуточный результат. И обучение начнется с конца (с сезона 15/16), т.к. по первым сезонам значительно меньше данных.

// изображение 1.png

*Количество матчей в датасете по сезонам.*

На последнем сезоне модель обучалась 50 эпох, на остальных по 10. Из датасета также отсеялись все игроки не из Барселоны.

// изображение 2.png

*График функции потерь по эпохам для последнего сезона.*

По графику выше видно, что модель обучилась насколько это возможно - высокая точность от нее не требуется, нужны лишь контекстные векторы. Процесс получения векторов из модели также описан в первом посте, поэтому остановлюсь на вычислении "движения" этих векторов на протяжении всех сезонов. В Word2Vec (контекстная модель векторизации слов, по аналогии с которой создана Player2Vec) для близости векторов принято использовать косинусную близость:

// изображение 3.png

*Получившееся значение по формуле будет в диапазоне [0, 1] - чем ближе вектора друг к другу, тем ближе значение к нулю. Опять же, в Word2Vec принято использовать обратное значение (то есть 1 - косинус), поэтому в нашем случае в модели чем ближе игроки, тем ближе значение к единице.*

На деле есть еще один нюанс. После получения векторов по каждому сезону нужно будет смотреть разницу между ними, тогда нам интересно именно расстояние, а не близость, поэтому из единицы значение вычитаться не будет. То есть на визуализации изменения вектора на протяжении сезонов по мере увеличения расстояния, число также будет увеличиваться. Вообще, именно косинусная мера используется в Word2Vec потому, что все вектора нормализованы (их длина одинакова и равна единице), поэтому имеет смысл ее не учитывать и полагаться только на угол.

Для визуализации использовался PCA из-за своей простоты и отсутствия необходимости в последующей кластеризации. Хотя в первом посте для визуализации обучался t-SNE из-за большего количества игроков и более наглядной картинки.

Последний момент перед визуализацией: на графиках часто будет в углу несколько точек (игроков), чьи имена не получится разглядеть - это потому что в тех сезонах данные игроки не играли и модель их сгруппировала фактически в одну точку.

**Результаты**

Чтобы не перебарщивать с информацией, я не буду визулизировать игроков по каждому сезону, выберу наиболее показательные для каждого тренера - ознакомиться с остальными можно будет по ссылке на код в конце статьи.

// изображение 4.png

Франк Райкард, сезон 06/07.

// изображение 5.png

Первая Лига чемпионов Пепа.

// изображение 6.png

Вторая.

// изображение 7.png

Тито Виланова, сезон 12/13 (единственный).

// изображение 8.png

Херардо Мартино, сезон 13/14 (единственный).

// изображение 9.png

Лига Чемпионов Энрике (единственная: 14/15).

**Подготовка к выводам**

Прежде чем окончательно перейти к выводам, стоит дать краткое описание стиля барсы при тренерах, которые ее возглавляли в период, представленный в датасете. Так как я начал смотреть под конец этого периода, мне помогли знающие люди.

Райкард (04/05-07/08): 433, контроль и доминирование, но не на уровне Пепа. Уклон в импровизацию игроков, а не в четкую структуру игры.

Гвардиола (08/09-11/12): революционные контроль, доминирование и структура позиционной атаки.

Виланова (12/13): автопилот имени Гвардиолы.

Мартино (13/14): попытка сделать барсу более вертикальной и тотально прессингующей, но проблема в постоянном метании от схемы к схеме.

Энрике (14/15-15/16): постепенный переход на более контратакующую и вертикальную игру с четкой атакующей тройкой.

**Выводы**

Тут начинается наиболее субъективная часть текста и я уверен, что многие с ней справятся лучше меня, и если у вас есть мысли по изложенным графикам, то с удовольствием обсужу в комментариях.

Первое, что меня заинтересовало - насколько сильно менялась игра команды в пас на протяжении сезонов, и особенно при смене тренера. Для этого приведу еще один график. Для того, чтобы получить разницу между игрой команды в разные сезоны я построил массивы по сезонам для каждого игрока с помощью косинусного расстояния между текущим и предыдущим вектором и потом просуммировал расстояние по всем игрокам в каждый сезон.

// изображение 10.png

*По x - номер сезона (начиная с 05/06), по y - разница между текущим и предыдущим сезоном.*

На графике можно разглядеть очень интересную тенденцию - Райкард и Пеп в свои первые 2 сезона играли похожим составом, и, не смотря на разный подход к владению, на выходе стиль и расположение игроков было немного похоже. Но с третьего сезона

Гвардиолы начались изменения.

// изображение 11.png

*2 и 3 сезон Гвардиолы. Месси на обеих картинках не поместился, но его позиция практически не изменилась.*

Сразу бросается в глаза, что на первой картинке почти все центральные защитники находятся очень кучно, а на второй разброс намного больше. Вероятно, это связанно с более выстроенным билдапом с лучшим распределением ролей. Хотя при кучности на первой картинке очень странно, что Бартра находится далеко от всех игроков - если когда-нибудь буду пересматривать матчи Барселоны во второй сезон Пепа, то обязательно обращу на это внимание. Также на первой картинке видно более явное разделение на центральных защитников, крайних, центральных и опорных полузащитников, атакующую группу и Педро с Месси. На второй картинке таких явных кластеров уже нет, хотя в среднем по позициям распределение похожее.

Выше я уже написал про Тите, который пытался продолжить путь Пепа, график это как раз подтверждает - сразу после роста значения в сезоны Гвардиолы идет резкий спад. Этот спад как раз отражает относительно небольшую разницу в игре Барсы между сезонами 11/12 и 12/13.

Еще интересно посмотреть на разницу между сезонами при Тите и Мартино, ведь по сути именно при Мартино начал меняться стиль клуба.

// изображение 12.png

Опять же, на втором изображении векторы игроков располагаются ближе к диагонали, что, мне кажется, говорит о меньшем разнообразии и более прямолинейном стиле. Это вполне может быть следствием того, что Мартино пытался сделать Барселону тотально прессингующей. С другой стороны, в отличие от предыдущих тренеров, Херардо часто менял схемы, и именно поэтому немного странно видеть такую кучность. Но очень часто команды, заточенные на прессинг подбирают схему для того, чтобы было удобнее подстраиваться под розыгрыши соперника, не меняя свою игру при владении.

Теперь попробую пройтись по конкретным игрокам. Начну с таблицы с топ-10 игроков, по тому, насколько их векторное представление менялось в среднем по количеству сезонов:

// изображение 13.png

*Таблица с игроками, отсортированными по средней "дистанции". В третьей колонке суммарная "дистанция" по всем сезонам.*

Я писал ранее про то, что расположение векторов защитников сильно меняется, особенно это было явно после прихода Гвардиолы. Таблица доказывает это предположение - в ней только два игрока не из линии обороны. Поэтому остановлюсь подробнее на аномальном Серхи Роберто и на Мунире с Санчесом. Ситуация Роберто объясняется тем, что, в разные сезоны его ближайшими соседями были либо Ракитич и Иньеста, либо Алвес и Альба. Причем второй вариант был именно в сезоны при Энрике.

~~Изменения Санчеса намного интереснее - при Пепе он явно играл в атакующей группе и, благодаря отлаженной структуре его пасовая активность была достаточно высоко и ему не приходилось часто опускаться и помогать команде продвигать мяч. Но сразу после ухода Гвардиолы его вектор начал двигаться в сторону центральных полузащитников. Вероятно, это говорит о том, что, не смотря на желание продолжить работу предшественника, Тите давал Алексису намного больше свободы. Также вероятно нарушилась структура владения и свобода Санчеса могла быть вынужденным ходом - чтобы узнать первопричину, нужно углубляться в статистику (а еще лучше посмотреть несколько матчей), но сам факт очевиден. Кстати, при Мартино Санчес опять стал ближе к атакующей группе, но тут скорее причиной стало общее упрощение игры в пас.~~

Изменения Санчеса и Педро намного интереснее – при Пепе и Тите они располагались достаточно высоко и играли явно в атакующей группе, то есть им не нужно было постоянно опускаться и помогать команде продвигать мяч. Но с приходом Мартино и резким изменением стиля и Педро, и Санчес стали ближе к полузащитникам, а расстояние до Месси (а потом и Неймара с Суаресом) только увеличилось. Скорее всего это говорит о сильном нарушении структуры владения – вингеры стали получать намного ниже. Это подтверждается не только векторами Педро и Санчеса, но и игроками ротации, выходящими на те же позиции. При Энрике в этом плане ситуация только усугубилась, инетерсно было бы посмотреть на картину после прихода Вальверде.

По Муниру я не нашел какой-то закономерности - думаю его метания говорят о том, что тренеры Барселоны так и не смогли его оптимально разместить, хотя все понимали, что он очень талантлив. Также влияние оказала дикая конкуренция.

Есть еще несколько интересных наблюдений, например:

* в отличие от Альбы, Д. Алвес всегда был близко к центральным полузащитникам, что визуально подтверждалось его игрой;
* явное выделение Бускетса во все сезоны, но особенно в период Энрике (для меня стало большим удивлением);
* отстраненность (если так можно про него говорить) Месси во все сезоны до прихода Гвардиолы - при нем благодаря структуре пропасть между остальной частью атакующей группы и ним сократилась (пропасть исключительно в стиле паса), а при Энрике Неймар оказался даже более отстраненным.

В заключение приведу два визуализации, на которых я поместил усредненные векторы всех игроков Барселоны за весь период, по которому есть данные. На первом обычная визуализация PCA, на втором более сложный алгоритм, который на больших данных лучше группирует элементы.

 // изображение 14.png

*Левое изображение - PCA, правое - T-SNE.*

T-SNE подтверждает объективность модели, так как на увеличенной выборке игроки сгруппировались пратически по своим позициям, а PCA - величие Хави и Иньесты, которые выделились из общего тренда.

**Резюме**

Не смотря на мою низкую осведомленность об игре клуба в рассматриваемый период, у меня сложилось более целая картина, подтвердились некоторые предположения и нашлось много интересных моментов. Было действительно интересно разбираться во всех графиках и искать закономерности и соответствия с уже известной информацией. А теперь представьте, насколько более глубокие выводы можно сделать, если предоставить доступ к этой модели профессионалу.

Спасибо за внимание, весь код, визуализации и обученные модели по каждому сезону по [ссылке на github](https://github.com/dailydaniel/Barca_Evolution).