

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра социальной информатики

**СМЕШАННЫЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ
ВЫБОРОВ В ОРГАНЫ ГОСУДАРСТВЕННОЙ ВЛАСТИ**

(автореферат бакалаврской работы)

студентки 4 курса 451 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
социологического факультета
Павловой Дарьи Александровны

Научный руководитель
кандидат физико-математических наук,
доцент

_____ М.Г.Плешаков
подпись, дата

Зав. кафедрой
кандидат социологических наук, доцент

_____ И.Г. Малинский
подпись, дата

Саратов 2023

ВВЕДЕНИЕ

В современном мире все чаще встречается необходимость прогнозирования, то есть суждения о возможных состояниях объекта в будущем, об альтернативных путях и сроках его осуществления. Эта необходимость усиливается потребностью определить будущее, предвидеть, как и когда страна выйдет из кризиса, какие тенденции возобладают, каковы будут результаты социальных, политических и экономических перемен и что конкретно они принесут каждому человеку. Реализовать алгоритм прогнозирования в государственные органы власти возможно с помощью искусственных нейронных сетей. Нейронная сеть — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей. По сути, в конечном итоге мы получаем продукт, имитирующий человеческое мышление.

Актуальность построения прогнозных моделей исхода выборов важна для исследования электорального процесса, его элементов и этапов, маркетинговой стратегии предвыборной кампании. Кроме того, микроуровневая модель электорального прогнозирования ставит под угрозу точность наблюдений в сфере экономики, где прогнозируются такие важные показатели, как, например, уровень безработицы или валовой национальный продукт.

Роль прогнозирования выборов в государственные органы власти достаточно велика. Актуальность прогнозирования усиливается потребностью определить будущее, предвидеть, как и когда страна выйдет из кризиса, какие тенденции возобладают, каковы будут результаты социальных, политических и экономических перемен и что конкретно они принесут каждому человеку.

Степень изученности темы. На данный момент модель не проработана достаточно детально, чтобы служить инструментарием для построения политической кампании кандидата на пост президента, однако сам факт возможности статистического прогноза итога выборов открывает нам огромные перспективы для дальнейшего анализа. Также существует несколько

технологий, которые позволят дополнить инструментарий. Очевидно, что нельзя забывать и об использовании в построении прогноза классических методов опросов.

Применению алгоритмов искусственных нейросетей в прогнозировании выборов посвятили свои работы А.А. Ослон¹, Ю.Ю. Петрунин².

Степень разработанности проблемы в зарубежных источниках значительно обширнее, чем в русскоязычных. Начальные наработки по проблеме ИИ относятся к трудам Г. Лейбница, Г. Буля. Формирование и оформление футурологических концепций в начале XX в. задало современный вектор развития ИИ, авторы К. Чапек, Н. Тесла, Л. Торрес сформулировали ожидания «умных машин» и автоматизированных экспертных систем.

В 1943 г. учёными В. МакКаллохом и В. Питтсом была предложена идея создания искусственной нейронной сети, по аналогу с принципами работы биологического мозга. Проблематика была углублена Э. Беркли, А. Тьюрингом. М. Марински и Д. Эдмудсон впервые симулировали передачу электронного импульса по нейронной сети.

Термин «Искусственный интеллект» был впервые сформулирован в 1955 г. на Дартмутской конференции учёными Д. Маккарти, М. Мински, Н. Рочестером и К. Шанноном. Первая компьютерная программа, имитирующая работу мозга, также была представлена на этой конференции учёными А. Ньюэллом, К. Шоу, Г. Симоном.

Дальнейшие разработки в сфере ИИ проводились Д. Боброу, Д. Вазенбаумом, Е. Фейгенбаумом, Б. Бухэненом, Д. Ледербергом, К. Джерасси.

¹ Александр Анатольевич Ослон (род. 19 марта 1952 года) — советский и российский социолог. Президент Фонда «Общественное мнение». Кандидат технических наук.

² Петрунин Юрий Юрьевич - доктор философских наук, профессор, заведующий кафедрой «Математические методы и информационные технологии в управлении» факультета государственного управления Московского государственного университета имени М.В. Ломоносова.

Большой вклад в технологии глубинного обучения внесли Д. Хопфилд и Р. Румельхарт, предложившие метод. Экспертные системы, как метод предложил Э. Хайгенбаум.

Современную проблематику ИИ разрабатывали Голдберг и Холланд, С. Бишоп, Т. Янг, Д. Хазарика, С. Пориа, А. Юсал, Д. Джун, Ц. ШтайнерТрелкелд, Р. Шаокин, К. Хе, Р. Гиршик и др.

Проблематикой ИИ в политической науке занимались Уоркман, Коллингвуд и Вилкерсон, Бойдстан, Д. Баттерс, Г. Кинг, Д. Пэн, М. Робертс, Б. Ю, К. Квинн, Б. Монро, К. Босалис, Т. Коэн, Дж. Фаррел, Дж. Гриммер, А. Тумасян, Т. Шпренгер, П. Санднер, Б. О'Коннор, Г. Ванг, Дж. Четано, А. Питерсон, А. Шпирлинг, М. Ансари, М. Азиз, М. Сиддики, М. Лавер, К. Бенуа, В. Лёве, Дж. Слапин, Г. Клувер, Д. Дирмер, Б. Лорендаль, П. Барбера, Дж. Вилкерсон, А. Гертель-Фернандес, К. Кашин, С. Нараян, А. Бильбао Джайо, С. Кудурунта, М. Ивьер, А. Рао, Дж. Юнгсеок и др.

Несмотря на очевидную актуальность данной темы, опубликованные работы российских социологов и политологов, направленные на прогнозирование результатов выборов, не позволяют говорить о существовании в научном мире России каких-либо моделей, основывающихся на точных данных и строгих статистико-математических методах анализа.

Объектом исследования является искусственный интеллект в политологии.

Предметом исследования является разработка смешанной модели прогнозирования выборов в государственные органы власти посредством алгоритмов искусственного интеллекта.

Цель исследования заключается в разработке смешанной модели прогнозирования выборов в государственные органы власти.

Для достижения поставленной цели работы были определены следующие **задачи**:

1. Изучить понятие нейрона и искусственных нейронных сетей.
2. Изучить подходы к прогнозированию электорального поведения.

3. Рассмотреть применение нейронных сетей в прогнозировании результатов выборов в государственные органы власти и, в частности, президентских выборов.

4. Разработать смешанную модель выборов в государственные органы власти.

Теоретико-методологические основания исследования содержат наработки в сфере исследования как ИИ в принципе, так и кейсов применения продуктов на основании ИИ в политологической практике, в основном, в электорально-аналитическом разрезе. Методологическую базу исследования составляет разбор кейсов применения ИИ в электоральных процессах.

Научная новизна исследования для отечественной политологии лежит в глубоком обзоре практик применения нейронных сетей в политологии и электоральных кампаниях и использовании программных продуктов на основе машинного обучения в прогнозировании результатов выборов. В настоящей ВКР актуализируется оценка разработанности проблематики искусственного интеллекта и перечисляются новые возможности продуктов на основе ИИ в разрезе их применимости в политологической науке.

Данная научно-квалификационная работа восполняет пробел в исследованиях практик применения искусственных нейронных сетей в политическом анализе и политических технологиях как минимум на уровне систематизации перечня подобных практик за рубежом через анализ кейсов применения искусственных нейронных сетей в прогнозировании исходов выборов, оптимизации ведения электоральных кампаний и возникающих возможностях по производству дезинформации в виде «дипфейк»-видеозаписей, подделки аудиозаписей и автоматическому массовому комментированию публикаций в Интернете.

В работе вдобавок уделяется внимание нестандартным вызовам для традиций демократического процесса, возникающим благодаря углубления проникновения ИИ-продуктов в политологию.

Для отечественной политической аналитики, новизна работы состоит в предложении междисциплинарного метода анализа политико-идеологических предпочтений электората, совмещающего наработки по моделированию электорального поля на основании исторических данных по голосованию с технологиями машинного обучения для прогнозирования результатов выборов.

Эмпирической базой выпускной квалификационной работы являются исходы выборов Президента России за 2006 г, второй – часть набора данных американских кандидатов, созданная при разделении данных на обучающую и тестовую выборки. Апробация результатов прогноза на результаты голосования по партийным спискам производилась через выявление соответствия прогноза, построенного с помощью инструментов ИИ на основании социологического прогноза и коэффициентов поддержки той или иной идеолого-политической структуры на основании исторических данных по голосованию на выборах реальным результатам выборов.

Теоретическая значимость исследования заключается в возможности использования основных положений и выводов данной работы для дальнейшего изучения выбранной проблематики.

Структура выпускной квалификационной работы представлена введением, тремя главами, заключением, списком использованных источников и приложением.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе «Разновидности нейронных сетей и их применение» описываются понятия биологического и искусственного нейрона, а также теоретические сведения о существующих на данный момент нейронных сетях и их применении.

Нейронная сеть — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей. По сути, в конечном итоге мы получаем продукт, имитирующий человеческое мышление.

Искусственные нейронные сети были построены по принципу биологических нейронных сетей, которые представляют собой сети нервных клеток, выполняющие определенные физиологические функции. Составным элементом нейронных сетей являются нейроны.

Искусственные нейронные сети были построены по принципу биологических нейронных сетей, которые представляют собой сети нервных клеток, выполняющие определенные физиологические функции. Составным элементом нейронных сетей являются нейроны.

Функция активации нейрона характеризует зависимость сигнала на выходе нейрона от суммы сигналов на его входах. Обычно функция является монотонно возрастающей и находится в области значений $[-1,1]$ (гиперболический тангенс) и $[0,1]$ (сигмоида). Для некоторых алгоритмов обучения необходимо, чтобы активационная функция была непрерывно дифференцируемой на всей числовой оси. Искусственный нейрон характеризуется своей активационной функцией (например, название "сигмоидальный нейрон").

Перцептрон — тип искусственного нейрона, разрабатываемый Фрэнком Розенблаттом в 1950-ых и 1960-ых годах. В современных работах чаще всего используют другую модель искусственного нейрона — 18 сигмоидальный нейрон.

Розенблатт предложил использовать веса — числа, выражающие важность вклада каждого входа в конечный результат. Взвешенная сумма (или веса) сравнивается с пороговым значением (threshold), и по результатам определяется, будет ли выдан 0 или 1. Пороговое значение также является параметром нейрона.

Нейроны можно разделить на группы в зависимости от их положения в сети:

- входные нейроны принимают исходный вектор данных;
- в промежуточных нейронах происходят основные вычислительные операции — обучение;

- выходные нейроны — результат работы сети.

Базовыми архитектурами нейронных сетей являются сети прямого распространения и персептроны. Данные нейронные сети очень прямолинейны. Информация по ним передается от входа сразу к выходу. Клетки слоя данных сетей не связаны между собой, в отличие от соседних слоев, которые обычно полностью связаны. Нейронные сети прямого распространения обычно обучаются по методу обратного распространения ошибки. При таком обучении нейронная сеть получает множество данных как на вход, так и на выход. Этот процесс называется обучением с учителем.

Еще одной разновидностью базовых нейронных сетей является нейронная сеть Хопфилда. Данная нейронная сеть используется для восстановления изображений. Нейронная сеть Хопфилда является полносвязной нейронной сетью с симметричной матрицей связей. Понятие полносвязной означает, что каждый нейрон передает свой выходной сигнал остальным нейронам, включая самого себя. Количество нейронов данной сети определяется количеством входов и выходов. Также нейронная сеть Хопфилда обучается только с учителем, так как для восстановления изображения ей необходимо знать исходные значения изображения, которое необходимо восстановить.

Еще одним видом нейронной сети является Машина Больцмана. Чаще всего данную нейронную сеть рассматривают как стохастический 13 генеративный вариант нейронной сети Хопфилда. Машина Больцмана используется для обучения алгоритм имитации отжига. Также данная нейронная сеть оказалась первой нейронной сетью, способной обучаться внутренним представлениям и решать сложные комбинаторные задачи.

Также существуют свёрточные нейронные сети. Данный вид нейронных сетей сильно отличается от остальных. Свёрточные нейронные сети обычно используют для классификации изображений. Главной особенностью данной нейронной сети является «сканер», который считывает изображение кусками, и процесс «свертки», который уменьшает размер матрицы признака изображения.

К более продвинутым конфигурациям нейронных сетей относятся рекуррентные нейронные сети. Особенностью этих нейронных сетей является то, что они получают информацию не только с предыдущего слоя, но также и от самих себя с предыдущего прохода. Из-за этой особенности становится важным порядок подачи данных для обучения нейронной сети. Сложностью в использовании нейронных сетей такого типа является проблема исчезающего градиента, которая определяется быстрой потерей информации с течением времени. Данная проблема влияет только на веса при обучении модели. Чаще всего нейронные сети такого типа используют для автоматического дополнения информации.

Нейронные сети Кохонена, также именуемые как самоорганизующиеся карты, применяют соревновательное обучение для классификации данных без учителя. Для данной нейронной сети подаются входные данные, которые нейронная сеть сама распределяет по нейронам, которые совпадают с ними в большей степени. После чего нейроны меняются для большей точности совпадения, двигая своих соседей.

В результате рассмотрения понятия нейрона было изучено строение и функционал биологического нейрона, который служит для приема, обработки и передачи информации. Также было рассмотрено строение искусственного нейрона, который является математической моделью биологического нейрона и выполняет схожие функции. Кроме рассмотрения искусственного нейрона в отдельности, были рассмотрены виды нейронных сетей – совокупности искусственных нейронов. Каждая искусственная нейронная сеть, описанная в данной главе, отличается как строением, так и сферой применения.

Говоря о практических возможностях и перспективах использования искусственного интеллекта в политической сфере, стоит отметить, что с появлением компьютерных технологий возникло направление на стыке кибернетики и политологии, изучающее возможности использования технологий ИИ для анализа и прогнозирования политических процессов. На данный момент, есть несколько приоритетных категорий исследований в

области ИИ в политологии: первая фокусирует внимание на анализе текстов, вторая делает упор на визуальную часть - анализ и классификация изображений.

Во второй главе «Методы и подходы к прогнозированию электорального поведения» речь идёт о научных и ненаучных методах прогнозирования в государственные органы власти.

Ненаучные модели прогнозирования выборов в госорганы не основаны на точно сформулированных причинах и гипотезах, на правильно собранных данных. Среди ненаучных моделей прогноза можно выделить несколько типов в зависимости от источника, из которого они исходят, — это модели, предлагаемые так называемыми предсказателями, политическими деятелями и общественными лидерами.

Говоря о предсказаниях политическими деятелями, одним из недостатков подобных моделей является предубежденность их авторов. Делая прогнозы, политические деятели опираются на свой собственный опыт и на информацию, полученную из ближайших к кандидатам источников. Когда они пытаются прогнозировать, то стремятся поддержать своих кандидатов, несмотря на объективные условия. В результате они почти всегда предсказывают результат с большим запасом голосов «за» и «против», чем выходит на самом деле.

Политические деятели – это не единственные эксперты, делающие прогнозы. Журналисты, обозреватели, ученые, критики также пытаются предсказать результаты выборов. Всех этих комментаторов избирательных кампаний можно отнести в группу общественных деятелей. Существенная разница между политическими деятелями и общественными авторитетами заключается в том, что последние, как правило, лишены прямой партийной предубежденности.

Однако, прогнозы, данные общественными авторитетами, ненадежны по своей сути, поскольку основаны на интуиции и не имеют четкого алгоритма, который можно использовать от выборов к выборам. Еще одна проблема, возникающая при построении моделей прогнозирования, заключается в том,

что в реальности предельная ошибка выборки превышает 3 процентных пункта, определенных теорией. Это происходит из-за не столько неверно построенной выборки или неверно выбранного инструмента исследования, а по причине трудности определения числа респондентов, в действительности голосующих на выборах.

В основе статистико-математических моделей прогнозирования результатов выборов принято использовать метод регрессионного анализа. Поскольку на сегодняшний день в политической науке развитых либеральных демократий отсутствует единая общепризнанная модель электорального поведения, то и для построения прогнозных моделей используются совершенно разные факторы влияния на результаты выборов. Сравнительный анализ, проведенный М. Льюис-Беком и Т. Райсом в работе «Прогнозирование выборов», показал наличие, по крайней мере, девяти моделей прогнозирования результатов президентских выборов, включая их собственную.

Хотя общие принципы построения этих моделей одинаковы, – все они основаны на общенациональных серийных опросах, обращаются к регрессионным моделям, – но в них используются различные переменные, объясняющие мотивы формирования электорального поведения. В то же время более детальный анализ предметной области показывает, что едва ли не во всех этих моделях используется т.н. экономический фактор голосования, берущий свои корни из рационально-инструментальной модели электорального поведения. Однако, как нет единства среди приверженцев рационально-инструментальной модели по вопросу, какой показатель брать за операционализированное понятие экономического фактора – оценку избирателем собственного экономического положения или результатов работы экономики страны в целом, так нет этого единства и среди авторов прогнозных электоральных моделей.

В электоральной сфере политической аналитики уже активно применяются возможности ИИ. Технологии ИИ создают возможности для

точного прогнозирования результатов выборов. На современном этапе, технологии ИИ применяются в следующих областях:

1. Политический анализ и прогнозирование, делится на две крупных подгруппы:

1) Прогнозирование результатов выборов на основании NLP-сетей. В этой области следует выделить работы Тумасяна, О'Коннора, Ванга, Ансари, Жоу, Максе.

2) Электоральное прогнозирование на основании анализа изображений. Разработками в этой области занимались Ванг, Ли и Луо, а также профессор Стенфордского университета М. Косински.

Существуют и коммерческие компании, специализирующиеся на аналитике и прогнозировании выборов с помощью ИИ, например, KCore Analytics, среди кейсов которой можно выделить точные прогнозы победы Д. Трампа, Ф. Фернандеса и Д. Байдена.

2. Использование ИИ для формирования политической повестки и агитации на основе анализа настроений

3. Использование ИИ и инструментов продвинутой аналитики для таргетирования политической рекламы в рамках предвыборной агитации.

Более половины избирателей в 2008 г., по данным Pewresearch (2009) использовали Интернет, как основной источник информации о кандидатах и ходе кампаний.

Команда штаба Обамы в 2008 г. впервые применила технологии машинного обучения для продвинутой аналитики данных и таргетирования рекламы по результатам анализа поведения пользователей сети.

4. Использование ИИ для создания ботов и Deep-fake. Исследованиями в этой области занимался Макс Вейсс.

5. Использование ИИ для распознавания дезинформации. В США, основные наработки по этому направлению создаёт Агентство перспективных исследовательских проектов (DARPA).

В третьей главе «Реализация смешанной модели прогнозирования выборов в государственные органы власти», посвященном практическому исследованию возможностей ИИ в отечественной политологии, предпринимается попытка создания гибрида двух моделей прогнозирования выборов: Л.Н. Ясницкого и А. Лихтмана. посредством пакета прикладных программ MATLAB.

The Matlab³ – язык программирования и пакет узкоспециализированных приложений. Это сокращенное название от Matrix Laboratory. Соответствующий комплекс представляет собой набор профессиональных инструментов для технических вычислений, а также компьютерного моделирования. Отличается строгими требованиями качества.

The Matlab – мультипарадигменный язык и среда числовых вычислений. Он поддерживает:

1. работу с матрицами;
2. отображение функций и данных;
3. реализацию различных алгоритмов;
4. создание пользовательских интерфейсов;
5. взаимодействие с программами, написанных на различных языках.

Пакет для работы с нейронными сетями Neural Network Toolbox⁴ представляет собой полноценную среду MATLAB для решения прикладных задач, обеспечивая поддержку проектирования, обучения и моделирования множества известных сетевых парадигм, от базовых моделей персептрона до самых современных ассоциативных и самоорганизующихся сетей. Пакет может быть использован для исследования и применения нейронных сетей к таким задачам, как обработка сигналов, нелинейное управление и финансовое моделирование.

³ Matlab: описание и особенности // [Электронный ресурс][сайт URL: <https://otus.ru/journal/matlab-opisanie-i-osobennosti/>] (дата обращения 8.02.2023) Загл. С экрана. Яз. Рус.

⁴ Нейронные сети. Реализация в Matlab: учебное пособие / С.Г. Николаева. – Казань: Казан. гос. энерг. ун-т, 2015. – 92 с.

Ясницкий в своей работе «Применение нейросетевых технологий в политологии» ставит вопрос «Можно ли при построении прогностической программы ограничиться учетом только личностных характеристик кандидата?». Он отмечает, что учесть и правильно оценить влияние всей палитры факторов на исход выборов чрезвычайно сложно, но есть и такие факторы, которые определяются только личностью претендента на президентское кресло. Это его возраст, пол, семейное положение, количество детей, род занятий, социальная принадлежность.

«Беспроегрывная» технология доктора Лихтмана и его коллеги-соавтора, русского геофизика Владимира Кейлиса-Борока, вошедшая в историю в начале 1980-х. Она очень проста. «Ключами к Белокипенному дому» представляются 13 пунктов, устанавливающих фаворита предвыборной гонки. Данными аспектами представляется протекающее и долгосрочное народнохозяйственное состояние, харизматичность оба кандидата, активы и ошибки во внешней политике, неимение общественных волнений, скандалы, дерзкие изменения в внутренней политике, существование эффективного кандидата через третьей партии, отношение в избраниях функционирующего президента, неимение соревновательности в верховодящей партии и количество мест в Палате поверенных сравнительно с предыдущими переходными выборами. Эти критерии выстроены в шкалу двоичных оппозиций (проще говоря, напротив любого из 13 критериев необходимо устанавливать плюсы и минусы).

Смешанная модель, реализованная в данном исследовании, показала результат, прогнозирующий не лучше модели Лихтмана. Однако по дереву решений можно заметить, что первым значимым признаком выделяется столбец из модели Ясницкого. Это говорит о том, что на другом тестовом примере набора данных модель может сделать решающий выбор в пользу истинного решения, поскольку признаки Ясницкого оказали положительное влияние на точность модели и оказали влияние на создание нового более

глубокого дерева решения.

Смешанная модель, обученная на наборе данных отечественных кандидатов, дала результат хуже, чем на наборе американских кандидатов. Это объясняется тем, что многие шкалы набора отечественных кандидатов имеют одинаковое значение в диапазоне года конкретных выборов, что не приносит вклад в точность модели, а иногда и создает ложные связи, обучаясь на наборе данных нескольких президентских выборов.

ЗАКЛЮЧЕНИЕ

Выпускная квалификационная работа посвящена разработке смешанного алгоритма прогнозирования в госорганы власти. В результате ее выполнения были спроектированы, реализованы и протестированы алгоритмы, позволяющие распознать победителя выборов в госорганы власти, первый из которых выдает число голосов за кандидата, а второй – определяет победу кандидата от правящей партии.

В результате рассмотрения понятия нейрона было изучено строение и функционал биологического нейрона, который служит для приема, обработки и передачи информации. Также было рассмотрено строение искусственного нейрона, который является математической моделью биологического нейрона и выполняет схожие функции. Кроме рассмотрения искусственного нейрона в отдельности, были рассмотрены виды нейронных сетей – совокупности искусственных нейронов.

Проводя обзор предметной области, были изучены методы прогнозирования в госорганы власти, рассмотрены самые успешные попытки внедрения искусственного интеллекта в электоральном исследовании.

Во время выполнения данной работы были успешно завершены следующие задачи:

- 1) Изучены виды нейронных сетей и их применение.
- 2) Изучены методы прогнозирования в госорганы власти: ненаучные и с использованием математического аппарата.
- 3) Проведен разведочный анализ наборов данных отечественных и

американских кандидатов.

4) Реализованы смешанные модели для определения числа голосов отечественных кандидатов (алгоритм Левенберга-Марквардта и алгоритм устойчивого обратного распространения) и модель для определения победы представителя правящей партии американских кандидатов (алгоритм Левенберга-Марквардта), проведено сравнение моделей и оценка точности их работы.

По результатам тестирования моделей были выявлены их недостатки, а также предложены варианты их использования.