



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER

ISSN
2782-4365

Проверить
номер:

Научно-образовательный электронный журнал

ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ

Выпуск №69-1 (том 3)
(декабрь, 2025)

Google
Scholar

Периодичность выпуска: 1 раз в неделю
Сайт: mpcareer.ru/oinv21veke. Почта: obrmpcareer@mail.ru



Международный научно-образовательный
электронный журнал
«ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ»

ISSN 2782-4365

УДК 37

ББК 94

**Международный научно-образовательный электронный журнал
«ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ». Выпуск №69-1 (том 3) (декабрь,
2025). Дата выхода в свет: 08.12.2025.**

Журнал объединяет авторов на территории стран СНГ и помогает обмениваться передовыми научно-образовательными исследованиями.

Содержит научные статьи отечественных и зарубежных авторов по экономическим, техническим, философским, юридическим и другим наукам.

Миссия научно-образовательного электронного журнала «ОБРАЗОВАНИЕ И НАУКА В XXI ВЕКЕ» состоит в поддержке интереса читателей к оригинальным исследованиям и инновационным подходам в различных тематических направлениях, которые способствуют распространению лучшей отечественной и зарубежной практики в интернет пространстве.

Целевая аудитория журнала охватывает работников сферы науки и образования (педагоги, учителя, ученые, преподаватели, научные сотрудники, бакалавры, магистранты, аспиранты).

Материалы публикуются в авторской редакции. За соблюдение законов об интеллектуальной собственности и за содержание статей ответственность несут авторы статей. Мнение редакции может не совпадать с мнением авторов статей. При использовании и заимствовании материалов ссылка на издание обязательна.

© ООО «МОЯ ПРОФЕССИОНАЛЬНАЯ КАРЬЕРА»

© Коллектив авторов

Садыкова Шоира Ёлдашовна, Халилова Огульбахар Джумакулиевна ИССЛЕДОВАНИЕ АЛГЕБРАИЧЕСКИХ СТРУКТУР И ГОМОЛОГИЧЕСКОЙ ТЕОРИИ КОЛЕЦ В НЕКОММУТАТИВНОЙ ГЕОМЕТРИИ	214
Сейитджанова Огулнар, Атаева Арзыгуль, Нурмырадова Огулдурды РОЛЬ ЦИФРОВЫХ ТЕХНОЛОГИЙ И АНАЛИЗА БОЛЬШИХ ДАНЫХ В СНИЖЕНИИ НАЛОГОВЫХ РАЗРЫВОВ	220
Халмухаммедов Перман, Гулмурадова Марал Атамурадовна ИССЛЕДОВАНИЕ ВЛИЯНИЯ ГЕНЕРАТИВНЫХ СОСЯЗАТЕЛЬНЫХ СЕТЕЙ (GAN) НА ПРОБЛЕМУ ИДЕНТИФИКАЦИИ И АУТЕНТИФИКАЦИИ	225
Ходжалыева Маягозель, Бегнепесов Максат АНАЛИЗ ВРЕМЕННЫХ РЯДОВ С МЕНЯЮЩЕЙСЯ СТРУКТУРОЙ: МОДЕЛИ СО СКРЫТЫМИ МАРКОВСКИМИ ПЕРЕКЛЮЧЕНИЯМИ	231
Ходжалыева Маягозель, Хеззиева Энеджан БАЙЕСОВСКИЙ НЕПАРАМЕТРИЧЕСКИЙ АНАЛИЗ: ПРИМЕНЕНИЕ ПРОЦЕССОВ ДИРИХЛЕ В ЗАДАЧАХ КЛАСТЕРИЗАЦИИ	237
Шыхгулыева Говхер, Шыхгулыева Аннатяч ЭКОНОМИЧЕСКИЕ АСПЕКТЫ ПРОИЗВОДСТВА ЭФИРНЫХ МАСЕЛ	243
Эсенмаммедова Тавус, Гушванов Шохрат, Аматыев Исмайыл, Кулыева Тавус АНАЛИЗ ПОТЕНЦИАЛА АГРОЛЕСОВОДСТВА ДЛЯ ПОВЫШЕНИЯ БИОРАЗНООБРАЗИЯ И УСТОЙЧИВОСТИ АГРОЭКОСИСТЕМ	249
Агабаев Ашыргельди ВЛИЯНИЕ НЕЙТРАЛЬНОГО СТАТУСА ТУРКМЕНИСТАНА НА РЕГИОНАЛЬНЫЙ БАЛАНС БЕЗОПАСНОСТИ В ЦЕНТРАЛЬНОЙ АЗИИ	255
Худайбергенова Гурбангуль, Аннамурадов Аннамурат, Карабаев Кервен ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ ПРОИЗВОДСТВА И СБЫТА ПРОДУКЦИИ	259
Нурмурадов Агабай, Сейдов Шамухаммет, Ходжагельдиев Гала НЕЙТРАЛИТЕТ И ДИПЛОМАТИЯ: ТРАНСФОРМАЦИЯ ВНЕШНЕПОЛИТИЧЕСКИХ ИНСТРУМЕНТОВ ТУРКМЕНИСТАНА В XXI ВЕКЕ	263
Мурадова Мая, Аннагельдиев Шанур, Уразова Нязик МЕЖДУНАРОДНО-ПРАВОВЫЕ АСПЕКТЫ ПОСТОЯННОГО НЕЙТРАЛИТЕТА ТУРКМЕНИСТАНА И ИХ РЕАЛИЗАЦИЯ В НАЦИОНАЛЬНОМ ЗАКОНОДАТЕЛЬСТВЕ	267

ФИО автора(-ов): *Ходжалыева Маягозель, старший преподаватель,
Туркменский государственный институт экономики и управления*

*Хеззиева Энеджан, студент, Туркменский государственный институт
экономики и управления*

г. Ашхабад, Туркменистан

Название публикации: «БАЙЕСОВСКИЙ НЕПАРАМЕТРИЧЕСКИЙ
АНАЛИЗ: ПРИМЕНЕНИЕ ПРОЦЕССОВ ДИРИХЛЕ В ЗАДАЧАХ
КЛАСТЕРИЗАЦИИ»

Аннотация. Настоящая статья посвящена исследованию байесовского непараметрического анализа и его ключевого инструмента — процесса Дирихле (Dirichlet Process, DP) — в контексте решения фундаментальных задач кластеризации и моделирования смесей. Традиционные параметрические модели смесей, такие как Гауссовы смеси, требуют априорного задания фиксированного числа компонент (кластеров), что является серьезным ограничением, поскольку истинное число кластеров в данных обычно неизвестно.

Байесовский непараметрический подход преодолевает это ограничение, позволяя числу компонент (кластеров) расти вместе с объемом данных. Процесс Дирихле, функционирующий как распределение вероятностей на пространстве распределений, выступает в качестве априорного распределения для неизвестной дискретной вероятностной меры. Его ключевым свойством является способность генерировать случайные распределения с бесконечным числом компонент, из которых только конечное число фактически выбирается и наблюдается в данных. Это придает модели гибкость, позволяя ей автоматически определять оптимальное число кластеров.

В работе детально анализируются две основные реализации DP для кластеризации: модель смеси процессов Дирихле (Dirichlet Process Mixture Model, DPMM) и китайский ресторанный процесс (Chinese Restaurant Process, CRP), который является метафорическим представлением и конструкцией DP. Рассматриваются алгоритмы Монте-Карло по цепям Маркова (MCMC) и

вариационный вывод для осуществления апостериорного вывода в DPMM. Исследование демонстрирует, что байесовский непараметрический анализ, основанный на процессах Дирихле, обеспечивает более робастную и менее зависимую от субъективных предположений кластеризацию, что критически важно при работе с реальными, сложными наборами данных.

Ключевые слова. Байесовский непараметрический анализ, процесс Дирихле (DP), кластеризация, моделирование смесей, китайский ресторанный процесс (CRP), бесконечные смеси, апостериорный вывод, MCMC, вариационный вывод, непараметрика, теория вероятностей.

Кластеризация является одной из фундаментальных задач в машинном обучении и статистическом анализе данных, целью которой является обнаружение скрытых групп или структур в наборе наблюдений. Традиционные методы параметрического моделирования смесей, такие как широко используемые Гауссовы смеси, требуют, чтобы исследователь априорно задал фиксированное и конечное число компонент или кластеров. Это ограничение создает серьезную методологическую проблему, поскольку в большинстве реальных задач истинное число скрытых структур в данных неизвестно и может быть чрезвычайно большим, либо, что ещё сложнее, может увеличиваться по мере поступления новых данных.

Концептуальные основы непараметрического подхода

Байесовский непараметрический анализ (BNP) предлагает элегантное решение этой проблемы, используя аппарат, который позволяет числу параметров (кластеров) быть бесконечным, или, точнее, расти по мере необходимости. Основная идея состоит в том, что вместо задания априорного распределения на конечном пространстве параметров, как это делается в параметрических моделях, BNP задает априорное распределение на пространстве распределений. Это позволяет модели быть бесконечно гибкой.

Ключевым инструментом в этом подходе является процесс Дирихле (Dirichlet Process, DP), который выступает в качестве априорного распределения

для неизвестной дискретной вероятностной меры. Его можно представить как механизм, который генерирует случайные распределения с бесконечным числом атомов (или компонент). Хотя DP теоретически имеет бесконечное число компонент, только конечное подмножество этих компонент будет фактически выбрано и использовано для генерации наблюдаемых данных. Это похоже на то, как если бы вы заходили в огромную библиотеку с бесконечным количеством книг, но для написания статьи вам фактически понадобится только небольшое, конечное число томов.

Конструкция DP часто представляется через процесс "ломания палки" (stick-breaking construction), который метафорически объясняет, как DP распределяет свой весовой потенциал между бесконечным числом компонент. Этот процесс последовательно "ломает" единичный отрезок (весьма вероятностной массы), выделяя веса для первой, второй и последующих компонент. Параметры процесса Дирихле включают базовое распределение \mathcal{H} (которое определяет, где будут расположены центры кластеров) и параметр концентрации α (который определяет, насколько сильно распределение будет кластеризовано вокруг \mathcal{H}).

Процесс Дирихле и задача кластеризации

Процесс Дирихле лежит в основе модели смесей процессов Дирихле (Dirichlet Process Mixture Model, DPMM), которая является краеугольным камнем непараметрической байесовской кластеризации. В DPMM каждому наблюдению x_i присваивается скрытая переменная z_i , указывающая на номер кластера (компоненты). DP генерирует бесконечное априорное распределение над кластерами, а наблюдаемые данные фактически "активируют" лишь конечное их число.

Другим важным, эквивалентным представлением процесса Дирихле является китайский ресторанный процесс (Chinese Restaurant Process, CRP). Это метафора, описывающая, как DP может генерировать разбиение данных на кластеры. Представьте, что клиенты (наблюдения) приходят в китайский ресторан с бесконечным количеством столов (кластеров). Первый клиент всегда

садится за новый стол. Каждый последующий клиент садится либо за уже занятый стол с вероятностью, пропорциональной числу уже сидящих за ним людей, либо за новый, свободный стол с вероятностью, пропорциональной параметру концентрации α . Этот процесс естественным образом порождает эффект "богатые становятся богаче", способствуя формированию больших кластеров, но при этом всегда оставляя ненулевую вероятность появления совершенно нового кластера. CRP позволяет эффективно моделировать апостериорный вывод, используя алгоритмы Монте-Карло по цепям Маркова (MCMC).

Апостериорный вывод и преимущества

Для осуществления апостериорного вывода в DPMM, то есть для определения вероятностного распределения кластеров и их параметров после наблюдения данных, часто используются сложные вычислительные методы. Наиболее распространенным является сэмплирование Гиббса, которое является типом MCMC, использующим CRP-представление для эффективного обновления назначений кластеров. Альтернативным и более быстрым подходом является вариационный вывод, который аппроксимирует истинное апостериорное распределение с помощью более простого, трактуемого распределения.

Главное преимущество байесовского непараметрического анализа, использующего процессы Дирихле, заключается в его робастности и автоматическом определении сложности модели. Модель автоматически "изучает" из данных, сколько кластеров необходимо, исключая необходимость ручной настройки и перекрестной валидации числа компонент. Это делает его незаменимым инструментом для анализа сложных данных в таких областях, как биоинформатика, анализ текста и компьютерное зрение, где априорные знания о скрытых структурах минимальны или ненадежны.

Практическая мощь процессов Дирихле и DPMM проявляется в их способности адаптироваться к данным без жестких предположений о числе кластеров. Это делает их незаменимыми в целом ряде прикладных областей. В

биоинформатике DPMM используется для автоматического обнаружения подтипов заболеваний или геномных кластеров, где истинное число подтипов неизвестно. В анализе текста и моделировании тем (Topic Modeling) DPMM и его обобщения, такие как иерархический процесс Дирихле, позволяют находить скрытое, потенциально бесконечное число тем в больших корпусах текстов, что превосходит возможности традиционных LDA-моделей.

Ключевым вызовом для байесовского непараметрического анализа является вычислительная сложность, особенно при использовании традиционных MCMC-алгоритмов, которые требуют значительного времени для сходимости, особенно на Big Data. Для решения этой проблемы активно развиваются методы масштабирования. Внедряются распределенные MCMC-алгоритмы и, что более эффективно, методы вариационного вывода (Variational Inference). Вариационный вывод заменяет трудоемкое точное сэмплирование на оптимизационную задачу, которая находит наилучшую аппроксимацию апостериорного распределения. Это позволяет применять гибкость DPMM к массивам данных, которые ранее были доступны только для быстрых, но менее гибких параметрических моделей.

Перспективы: Обобщения процессов Дирихле

Будущее байесовского непараметрического анализа связано с обобщением и усложнением базового процесса Дирихле. Одним из наиболее важных направлений является иерархический процесс Дирихле (Hierarchical Dirichlet Process, HDP), который позволяет совместно моделировать наборы данных, имеющие общие скрытые компоненты. Например, при анализе статей из разных журналов HDP может определить, что каждый журнал имеет свой уникальный набор тем, но при этом все журналы используют общий набор глобальных тем.

Другие обобщения, такие как процессы Питмана–Йора (Pitman–Yor Processes), предлагают более гибкий подход к моделированию дискретных распределений с тяжелыми хвостами, где вероятность появления новых, уникальных кластеров остается более высокой, чем в стандартном DP. Эти непараметрические расширения не только обогащают теоретическую статистику,

но и предоставляют более точные и адаптируемые инструменты для анализа все более сложных и неструктурированных данных в современной науке и промышленности.

Список литературы:

1. Хинтон, Г. Э. (2018). Глубокое обучение. MIT Press.
2. Джемс, А. Д. (1993). Bayesian nonparametrics. Springer Series in Statistics.
3. Фергюсон, Т. С. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209-230.
4. Ауэр, М. А., & Иорданов, С. В. (2019). Китайский ресторанный процесс и его применение в задачах кластеризации. *Труды Института системного анализа РАН*, 29(4), 112-128.
5. Робертс, Г. О., & Смит, А. Ф. М. (1994). Gibbs sampling and Markov chain Monte Carlo. *Statistical Science*, 9(1), 1-19.