

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

БАЙЕСОВСКИЕ СЕТИ В МАШИННОМ ОБУЧЕНИИ

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 248 группы

направления 09.04.03 — Прикладная информатика

механико-математического факультета

Рустамова Дмитрия Романовича

Научный руководитель

доцент, к. ф.-м. н.

Н. Ю. Агафонова

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2026

Введение. В современном машинном обучении все большее значение приобретают методы, способные не только строить точные прогнозы, но и явно учитывать неопределенность. Для задач распознавания образов это особенно важно: итоговая классификация зависит от шума измерений, ограниченности обучающей выборки и визуальной близости классов.

Байесовские сети представляют собой инструмент вероятностного моделирования, объединяющий:

- теорему Байеса как правило обновления знаний по данным;
- аппарат условной независимости;
- графовое представление совместных распределений;
- возможность интерпретируемого вывода при неполной информации.

Актуальность темы определяется требованиями прикладных областей: в диагностике, прогнозировании, анализе рисков и обработке изображений важны точность, объяснимость и оценка надежности прогноза. Классические нейронные сети хорошо выявляют сложные зависимости, но в стандартной форме не дают полноценного вероятностного описания неопределенности. Байесовский подход преодолевает это ограничение, рассматривая параметры модели как случайные величины.

Данная магистерская работа посвящена исследованию байесовских сетей и байесовских моделей в задачах машинного обучения и распознавания образов. В работе рассматриваются теоретические основы вероятностного подхода, графовые вероятностные модели, методы точного и приближенного вывода, а также практическая реализация двух моделей для классификации рукописных математических символов из набора данных HASYv2: дискретной наивной байесовской сети и байесовской сверточной нейронной сети.

Целью данной магистерской работы является исследование методов построения и применения байесовских сетей в задачах машинного обучения и распознавания образов, а также экспериментальное сравнение классической дискретной байесовской модели и байесовской глубокой нейронной сети.

Для достижения поставленной цели были сформулированы следующие

задачи исследования:

- провести анализ теоретических основ байесовского подхода к оцениванию параметров и моделированию неопределенности;
- рассмотреть задачу распознавания образов и особенности классификации рукописных математических символов;
- исследовать графовые вероятностные модели, включая байесовские сети и марковские случайные поля;
- изучить методы приближенного вывода, в том числе вариационный вывод и методы Монте-Карло по схеме марковских цепей;
- рассмотреть роль дивергенции Кульбака-Лейблера и EM-алгоритма в байесовском оценивании;
- реализовать дискретную байесовскую сеть и байесовскую сверточную нейронную сеть для распознавания символов HASYv2;
- провести сравнительный анализ реализованных моделей с точки зрения точности, интерпретируемости, вычислительной сложности и способности оценивать неопределенность.

Объектом исследования выступают байесовские сети и байесовские вероятностные модели, применяемые в задачах машинного обучения.

Предметом исследования являются методы построения, обучения и применения дискретных байесовских сетей и байесовских нейронных сетей в задаче распознавания образов.

Научная новизна данной работы заключается в следующем:

- выполнено единое рассмотрение классического и современного байесовских подходов к распознаванию образов;
- реализована дискретная байесовская сеть на основе PCA-признаков и дискретизации;
- реализована байесовская сверточная нейронная сеть с вариационным выводом;
- проведено сравнение моделей на наборе рукописных математических символов HASYv2;

- показан компромисс между точностью, интерпретируемостью и оценкой неопределенности.

Практическая значимость определяется возможностью использования результатов в интеллектуальных системах, работающих с неполными, шумными или неоднозначными данными:

- дискретные байесовские сети подходят для задач, где важны прозрачность и экспертная проверяемость;
- байесовские нейронные сети полезны там, где необходимы высокая точность и количественная оценка неопределенности.

Методы исследования: теория вероятностей, байесовский вывод, графовые вероятностные модели, методы снижения размерности, вариационный вывод, EM-алгоритм, методы машинного обучения, сверточные нейронные сети, программная реализация на языке Python с использованием библиотек `pgmpy`, `TensorFlow Probability`, `scikit-learn`.

Структура работы. Магистерская работа состоит из введения, семи глав, заключения, списка использованных источников и приложений. Теоретическая часть включает вероятностные основания байесовского подхода, постановку задачи распознавания, графовые модели и методы вывода. Практическая часть содержит реализацию моделей, анализ экспериментов и выводы о применимости каждого подхода.

Теоретическая значимость состоит в систематизации понятий байесовского машинного обучения и показе связи между:

- классической теорией вероятностей;
- графовыми вероятностными моделями;
- вариационным выводом;
- современными байесовскими нейронными сетями.

Первая глава посвящена вероятностным основаниям байесовского подхода к оцениванию параметров. В ней рассматриваются:

- пространство элементарных событий, события и вероятность;

- условная вероятность, независимость и формула полной вероятности;
- теорема Байеса как механизм пересмотра априорных гипотез;
- априорное распределение, функция правдоподобия, апостериорное и прогностическое распределения.

Вторая глава посвящена задаче распознавания образов. Классификация рукописных математических символов рассматривается как прикладная задача со следующими трудностями:

- большое число классов;
- неодинаковое количество примеров в классах;
- наличие визуально похожих символов;
- зависимость качества классификации от способа представления изображений.

В главе обсуждаются понятия функции ошибки, переобучения и регуляризации. Показано, что увеличение сложности модели без контроля обобщающей способности может приводить к ухудшению качества на новых данных. Рассматриваются способы борьбы с переобучением: регуляризация, увеличение объема обучающей выборки, выбор более устойчивых признаков. Эти рассуждения мотивируют переход к вероятностным моделям, которые позволяют не только находить наиболее вероятный класс, но и оценивать степень уверенности в предсказании.

Особое внимание уделено примеру аппроксимации полиномиальной кривой. На этом примере показывается различие между частотным и байесовским подходами: в частотной постановке параметры модели считаются фиксированными, а в байесовской постановке неопределенность относительно параметров описывается вероятностным распределением. Рассматривается роль нормального распределения, точности, регуляризации и априорных предположений о параметрах. Делается вывод, что байесовское оценивание естественно связывает обучение модели с контролем переобучения и количественным описанием неопределенности.

Третья глава содержит изложение графовых вероятностных моделей.

Основные элементы главы:

- ориентированные ациклические графы как основа байесовских сетей;
- неориентированные модели и марковские случайные поля;
- факторизация совместного распределения в произведение условных распределений;
- компактное описание зависимостей между случайными величинами.

В главе подробно анализируются понятия условной независимости, d -разделения и марковского покрытия. Рассматриваются базовые графовые конфигурации, демонстрирующие, как наличие или отсутствие наблюдений может блокировать или разблокировать пути зависимости между переменными. Показано, что графовая структура не только уменьшает число параметров модели, но и делает вероятностные зависимости более наглядными. Также обсуждается применение байесовской сети к задаче полиномиальной регрессии и вводятся порождающие модели, связывающие скрытые причины и наблюдаемые данные.

Четвертая глава посвящена приближенному выводу. Точный вывод во многих реалистичных моделях вычислительно сложен, поэтому рассматриваются два класса методов:

- вариационный вывод;
- методы Монте-Карло по схеме марковских цепей.

Вариационный вывод описывается как задача замены сложного апостериорного распределения более простым распределением из выбранного семейства. Вводится нижняя граница обоснованности (ELBO), максимизация которой эквивалентна приближению истинного апостериорного распределения в смысле дивергенции Кульбака-Лейблера. Отдельно обсуждается подход среднего поля и нормальные вариационные приближения. В контексте практической части подчеркивается, что именно вариационный вывод используется при обучении байесовской сверточной нейронной сети с вероятностными слоями.

Также отмечается различие между неопределенностью данных и неопре-

деленностью модели. Первая связана с шумом и неоднозначностью самих наблюдений, вторая - с недостатком информации для точного выбора параметров модели. В задачах распознавания образов это различие особенно важно: визуально похожие символы могут быть неоднозначными даже для хорошо обученной модели, тогда как редкие или плохо представленные классы увеличивают неопределенность параметров. Байесовский вывод дает естественный аппарат для учета обоих аспектов.

Пятая глава рассматривает дивергенцию Кульбака-Лейблера и EM-алгоритм. Дивергенция Кульбака-Лейблера используется как мера различия между вероятностными распределениями и связывает вариационный вывод с задачей приближения апостериорного распределения.

EM-алгоритм рассматривается как метод оценивания параметров моделей со скрытыми переменными. Описываются E-шаг, на котором вычисляется апостериорное распределение скрытых переменных, и M-шаг, на котором максимизируется ожидаемый логарифм полного правдоподобия. Показывается связь EM-алгоритма с вариационным выводом: EM можно интерпретировать как частный случай координатной оптимизации нижней границы обоснованности. Такой взгляд связывает классические алгоритмы обучения вероятностных моделей с современными методами байесовского машинного обучения.

Шестая глава посвящена программной реализации моделей для распознавания рукописных математических символов. Реализованы два подхода:

- дискретная наивная байесовская сеть: изображение преобразуется в числовое представление, размерность уменьшается методом PCA, признаки дискретизируются, параметры оцениваются байесовским способом;
- байесовская сверточная нейронная сеть: веса рассматриваются как случайные величины, а обучение сводится к вариационному приближению апостериорного распределения.

В главе приводятся результаты обучения и анализ ошибок классификации:

- для дискретной модели характерны ошибки, связанные с потерей информации при дискретизации и предположением условной независимости признаков;
- для байесовской нейронной сети наиболее сложными остаются пары визуально близких символов, различающихся тонкими деталями формы.

Анализ ошибок показывает, что дискретная модель чаще путает классы, различия между которыми теряются после PCA-преобразования и разбиения признаков на интервалы. Байесовская нейронная сеть лучше сохраняет информацию о форме символа, но также испытывает трудности на границах классов, где различия минимальны. Тем не менее вероятностная природа модели позволяет рассматривать не только итоговую метку, но и распределение уверенности по классам, что полезно при последующей экспертной проверке результата.

Отдельно подчеркивается, что для рукописных символов существенную роль играет способ извлечения признаков. Простое векторное представление изображения удобно для классических моделей, но плохо учитывает локальную пространственную структуру. Сверточные модели, напротив, способны выделять локальные элементы формы и объединять их в более сложные признаки. Это различие становится одним из ключевых факторов при дальнейшем сравнении дискретной байесовской сети и байесовской нейронной сети.

Седьмая глава содержит сравнительный анализ реализованных моделей. Основные результаты:

- дискретная байесовская сеть с PCA-дискретизацией достигла точности 56,6%;
- байесовская сверточная нейронная сеть с вариационным выводом показала точность 78,1%;
- разница объясняется сохранением пространственной структуры изображения в сверточной модели и потерей части информации при дискретизации в классической модели.

При этом дискретная байесовская сеть обладает важным преимуществом - высокой интерпретируемостью. Ее параметры можно непосредствен-

но анализировать, а решения модели легче объяснять через условные распределения признаков. Байесовская нейронная сеть значительно точнее, но требует большего времени обучения, больших вычислительных ресурсов и труднее поддается прямой интерпретации. Зато она позволяет оценивать неопределенность предсказаний за счет распределения по весам и усреднения результатов по выборкам из апостериорного распределения.

Выбор модели определяется требованиями прикладной задачи:

- если важны прозрачность, простота и экспертная проверка, предпочтительна дискретная байесовская сеть;
- если важны точность и оценка неопределенности в сложных данных, более эффективна байесовская глубокая нейронная сеть.

Заключение. В ходе выполнения магистерской работы были исследованы байесовские сети и байесовские модели в контексте машинного обучения и распознавания образов. Рассмотрены:

- вероятностные основы байесовского подхода;
- теорема Байеса и методы оценки параметров;
- графовые вероятностные модели;
- приближенный вывод, дивергенция Кульбака-Лейблера и EM-алгоритм.

Практическая часть включала реализацию двух моделей для набора `NASV2`:

- дискретная байесовская сеть на основе PCA-признаков и дискретизации: точность 56,6%;
- байесовская сверточная нейронная сеть с вариационным выводом: точность 78,1%.

В ходе исследования были решены поставленные задачи:

- изучены подходы к построению байесовских моделей;
- рассмотрены методы вывода и обучения;
- выполнена программная реализация двух моделей;

- проведено экспериментальное сравнение и проанализированы причины различий в качестве.

В результате работы показано, что байесовский подход остается важным направлением машинного обучения, поскольку позволяет объединять данные, априорные предположения и количественную оценку неопределенности. Полученные результаты могут быть использованы при разработке интеллектуальных систем распознавания образов, диагностических систем, систем анализа рисков и других приложений, где важны надежность, объяснимость и устойчивость к неоднозначным данным. Перспективными направлениями дальнейших исследований являются улучшение дискретизации признаков, использование более выразительных структур байесовских сетей, расширение набора экспериментальных данных и применение современных методов калибровки неопределенности для глубоких байесовских моделей.