

Заключение по содержанию диссертации

Гасников Александр Владимирович

(Ф.И.О. члена диссертационного совета)

Волков Никита Алексеевич

(Ф.И.О. соискателя ученой степени)

О некоторых свойствах вероятностных распределений

и их применении в задачах машинного обучения

(Название диссертации, ученая степень, на которую представлена диссертация, специальность)

представлена на соискание степени кандидата физико-математических наук по специальности 05.13.17 (теоретические основы информатики)

Дата защиты 15.12.20

Оценка соответствия диссертации требованиям Положения о присуждении ученых степеней кандидата наук, доктора наук в МФТИ (далее - Положение):

1. Актуальность тематики диссертации:

В диссертации предлагается новый метод оценивания параметров смеси многомерных распределений Стьюдента, служащий для решения двух распространенных проблем – наличия выбросов и пропущенных значений. Разработанный итерационный метод является обобщением метода, предложенного в 2006 году в работе А. Vellido. Кроме того, в диссертации предлагается, с использованием полученных оценок, улучшить известные методы оценивания представительности проб пластовых флюидов. Наконец, в диссертации изучается монотонности по b вероятности того, что биномиальная случайная величина с параметрами n и $b/(n+c)$ меньше чем b . Эта задача имеет отношение к известному вопросу Рамануджана о поведении функции Раманаджуна, а также к гипотезе Самуэляса и минимальной возможной вероятности отклонения на константу суммы независимых случайных величин от математического ожидания.

2. Научная новизна выносимых на защиту результатов:

Результаты диссертации являются новыми и подтверждены строгими и верными математическими доказательствами. Основными результатами являются следующие: 1) изучена монотонность по b вероятности того, что биномиальная случайная величина с параметрами n и $b/(n+c)$ меньше чем n при c из $[0,1]$ (ранее ответ был известен лишь для $c=0$), 2) разработан новый итерационный метод оценивания параметров смеси многомерных распределений Стьюдента по выборкам с пропусками, основанный на EM-алгоритме, 3) разработан метод машинного обучения (и соответствующий программный продукт), позволяющий решать ряд важных задач анализа данных.

3. Теоретическая и практическая значимость диссертационной работы:

Теоретические результаты диссертации могут быть использованы в теории вероятностей (например, для уточнения гипотезы Самуэляса в случае независимых одинаково распределенных случайных величин с двумя атомами) и математической статистике (при изучении итерационных методов оценивания параметров смеси

распределений). Практические результаты третьей главы диссертации могут быть использованы для решения некоторых задач анализа данных (особенно в работе инженеров геологов).

4. Полнота опубликования основных результатов диссертации в рецензируемых научных изданиях в соответствии с требованиями Положения:

Основные результаты диссертации опубликованы в 5 работах (три в журналах, индексируемых Scopus, еще две – RSCI). Результаты также докладывались на ряде семинаров и конференций.

5. Вопросы и замечания (в соответствии с п. 4.13 Положения соискатель отвечает на сформулированные здесь вопросы и замечания на заседании по защите диссертации):

Возможно ли, что вероятность $P(X < b)$ меняет монотонность несколько раз вблизи порога? Возможно ли для каких-то значений c получить настолько же точный результат, как тот, который известен при $c=0$?

Возникает и ряд других вопросов, на которых в первой главе не содержится ответ. Стоило добавить в первую главу некоторую дискуссию, в которой такие вопросы и возможность ответа на них обсуждается.

В диссертации есть незначительное количество орфографических ошибок и опечаток, которые не влияют существенно на высокий уровень результата и презентации.

6. Общая характеристика диссертации (не включает резолютивную часть):

Диссертационная работа производит положительное впечатление. В ней решаются три задачи: 1) монотонность по b вероятности $P(X < b)$, где X – биномиальная случайная величина с параметрами n и $b/(n+c)$, 2) оценивание параметров смеси многомерных распределений Стьюдента, 3) построение метода машинного обучения, позволяющего решать ряд задач при наличии пропусков в данных. Хочется отметить, что решение первой задачи имеет некоторые следствия, уточняющие гипотезу Самуэляса, а метод машинного обучения, предложенный в третьей главе, был использован диссертантом для оценки достоверности данных исследований свойств пластовых флюидов. Наконец, проведенные эксперименты показали преимущество разработанного метода в сравнении с широко известными методами машинного обучения. Все теоретические результаты диссертации сопровождаются строгими математическими задачами. Диссертация является высококвалифицированной работой и удовлетворяет всем требованиям Положения о присуждении ученых степеней.

Дата _____

Подпись АУ / Гасников Александр Владимирович

