

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к дипломной работе

«Машинное обучение для последовательного тестирования гипотез»

Домнин Даниил Дмитриевич

Научный руководитель – старший преподаватель кафедры дискретной математики и алгоритмики ФПМИ Молочко И. П.

Минск, 2025

АННОТАЦИЯ

Дипломная работа, 53 страницы, 2 таблицы, 12 иллюстраций, 11 формул, 12 источников.

Ключевые слова: ПОСЛЕДОВАТЕЛЬНО ТЕСТИРОВАНИЕ ГИПОТЕЗ, КОВАРИАТНАЯ КОРРЕКЦИЯ, SPRT, CUPED, МАШИННОЕ ОБУЧЕНИЕ

Объект исследования является процедура последовательного тестирования статистических гипотез.

Предметом исследования являются методы повышения эффективности последовательного тестирования гипотез за счет использования информации о ковариатах с помощью моделей машинного обучения.

Целью работы является разработка и исследование метода повышения эффективности последовательного тестирования гипотез за счет ковариатной коррекции на основе моделей машинного обучения для ускорения обнаружения статистически значимых эффектов при строгом контроле ошибок первого рода.

Методами исследования являются теоретический анализ, разработка алгоритма, моделирование и вычислительные эксперименты.

Полученные результаты и их новизна: разработан метод ковариатной коррекции на основе ML-моделей для последовательного тестирования. Экспериментально подтверждено значимое сокращение среднего числа наблюдений при контроле ошибки первого рода. Эффективность зависит от качества ML-модели; нелинейные модели эффективнее при сложных зависимостях. Новизна – в сочетании однократного обучения ML-модели с анализом остатков в последовательных процедурах.

Достоверность материалов и результатов дипломной работы: использованные материалы и результаты дипломной работы являются достоверными. Работа выполнена самостоятельно.

Областью возможного практического применения является повышение эффективности онлайн-экспериментов в технологических компаниях, интернет-маркетинге и других сферах, применяющих последовательный анализ данных.

АНАТАЦЫЯ

Дыпломная праца, 53 старонкі, 2 табліцы, 12 ілюстрацый, 11 формул, 12 крыніц.

Ключавыя словы: ПАСЛЯДОЎНАЕ ТЭСЦІРАВАННЕ ГІПОТЭЗ, КАВАРЫЯТНАЯ КАРЭКЦЫЯ, SPRT, CUPED, МАШЫННАЕ НАВУЧАННЕ.

Аб'ектам даследавання з'яўляецца працэдура паслядоўнага тэсціравання статыстычных гіпотэз.

Прадметам даследавання з'яўляюцца метады павышэння эфектыўнасці паслядоўнага тэсціравання гіпотэз за кошт выкарыстання інфармацыі аб каварыятах з дапамогай мадэляў машыннага навучання.

Мэтай даследавання з'яўляецца распрацоўка і даследаванне метаду павышэння эфектыўнасці паслядоўнага тэсціравання гіпотэз за кошт каварыятнай карэкцыі на аснове мадэлей машыннага навучання для паскарэння выяўлення статыстычна значных эфектаў пры строгім кантролі памылак першага роду..

Метадамі даследавання з'яўляюцца тэарэтычны аналіз, распрацоўка алгарытму, мадэляванне і вылічальныя эксперыменты.

Атрыманыя вынікі і іх навізна: распрацаваны метады каварыятнай карэкцыі на аснове ML-мадэлей для паслядоўнага тэсціравання. Эксперыментальна пацверджана значнае скарачэнне сярэдняй колькасці назіранняў пры кантролі памылак першага роду. Эфектыўнасць залежыць ад якасці ML-мадэлі; нелінейныя мадэлі больш эфектыўныя пры складаных залежнасцях. Навізна – у спалучэнні аднаразовага навучання ML-мадэлі з аналізам астач у паслядоўных працэдурах.

Даставернасць матэрыялаў і вынікаў дыпломнай працы: выкарыстаныя матэрыялы і вынікі дыпломнай Працы з'яўляюцца даставернымі. Праца выканана самастойна.

Вобласцю магчымага практычнага прымянення з'яўляецца павышэнне эфектыўнасці анлайн-эксперыментаў у тэхналагічных кампаніях, інтэрнэт-маркетынгу і іншых галінах, дзе выкарыстоўваецца паслядоўны аналіз даных.

ANNOTATION

Diploma work, 53 pages, 2 tables, 12 illustrations, 11 formulas, 12 sources.

Keywords: SEQUENTIAL HYPOTHESIS TESTING, COVARIATE ADJUSTMENT, SPRT, CUPED, MACHINE LEARNING.

The object of the research is the procedure of sequential hypothesis testing.

The subject of the research is a method for improving the efficiency of sequential hypothesis testing by exploiting covariate information using machine learning models.

The purpose of the research is development and study of a method to enhance the efficiency of sequential hypothesis testing through covariate adjustment based on machine learning models to accelerate the detection of statistically significant effects under strict control of type I errors.

Methods of research are theoretical analysis, algorithm development, simulation, and computational experiments.

The results of the work and their novelty: a covariate adjustment method based on ML models for sequential hypothesis testing has been developed. Experimentally confirmed significant reduction in the average number of observations while controlling the type I error. The effectiveness depends on the quality of the ML model; nonlinear models are more effective for complex dependencies. The novelty lies in combining one-time ML model training with residual analysis in sequential procedures.

Authenticity of the materials and results of the diploma work: the materials used and the results of the diploma work are authentic. The work has been put through independently.

Recommendations on the usage. The results of the work can be used to improve the efficiency of online experiments in technology companies, internet marketing, and other fields utilizing sequential data analysis.