

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

**«Зависимость результатов соревнований по программированию от
различных факторов»**

Василевский Виктор Владимирович

Научный руководитель – доктор физико-математических наук,
профессор Котов В.М.

Минск, 2019

Реферат

Магистерская диссертация, 43 страницы, 8 рисунков, 16 таблиц, 5 источников, 1 приложение.

СПОРТИВНОЕ ПРОГРАММИРОВАНИЕ, СТАТИСТИЧЕСКИЙ АНАЛИЗ, МАШИННОЕ ОБУЧЕНИЕ, МЕТОДЫ МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ.

Объект исследования — данные сайта codeforces.com. В частности, в работе использовались статистические данные результатов соревнований по спортивному программированию, метаинформация про участников соревнований, исходные коды решений задач и информация о тэгах задач.

Цель работы — исследование подхода машинного обучения для автоматического определения тэгов задач и решений спортивного программирования по исходным кодам решений участников, оценка и сравнительный анализ различных методов машинного обучения; исследование статистических данных о зависимости результатов соревнований по спортивному программированию от различных факторов.

В ходе работы были исследованы методы информационного поиска TF-IDF, BM-25, а также методы многоклассового машинного обучения, такие как мультиклассовый метод k ближайших соседей, EmbeddingClassifier.

Результатом является предложенный подход к решению задачи автоматического определения тэгов задач по спортивному программированию, веб-интерфейс, позволяющий визуализировать результаты. Выполнена реализация различных методов автоматического определения тэгов задач и проведён сравнительный анализ этих методов. Так же результатом являются данные полученные в результате проведённого статистического анализа зависимостей результатов соревнований по программированию от различных факторов.

Областью применения является спортивное программирование, автоматическое определение тэгов задач и решений по спортивному программированию.

Abstract

Master's thesis, 43 pages, 8 figures, 16 tables, 5 references, 1 appendix.

SPORTS PROGRAMMING, STATISTICAL ANALYSIS, MACHINE LEARNING METHODS MNOGOGOLOSNOY CLASSIFICATION.

Object of research — site data codeforces.com ahhh! In particular, the paper used statistical data on the results of the competition in sports programming, meta information about the participants, the source code of the problem solutions and information about the task tags.

The aim of the work is to study the approach of machine learning to automatically determine the tags of problems and solutions of sports programming on the source code of decisions of participants, evaluation and comparative analysis of various methods of machine learning; study of statistical data on the dependence of the results of sports programming on various factors.

In the course of the work the methods of information retrieval TF-IDF, BM-25, as well as methods of multi-class machine learning, such as multi-class method k nearest neighbors, EmbeddingClassifier were investigated.

The result is the proposed approach to solving the problem of automatic identification of task tags for sports programming, a web interface that allows you to visualize the results. The implementation of various methods of automatic identification of task tags is carried out and a comparative analysis of these methods is carried out. The same result is the data obtained as a result of the statistical analysis of the dependence of the results of programming competitions on various factors.

The field of application is sports programming, automatic identification of task tags and solutions for sports programming.