АННОТАЦИЯ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Наименование темы: Исследование временных изменений стилистики А.С. Пушкина методами машинного обучения

Выполнена студентом Гудковым Степаном Алексеевичем
Факультет информационных технологий, Новосибирский государственный университет
Кафедра общей информатики
Группа 20206
Направление подготовки 09.03.01 Информатика и вычислительная техника
Направленность (профиль): Программная инженерия и компьютерные науки

Объем работы: 35 страниц
Количество иллюстраций: 1
Количество таблиц: 2
Количество литературных источников: 36
Количество приложений: 3

Ключевые слова: стилеметрия, анализ поэтических текстов, особенности авторского стиля, машинное обучение, нейронные сети.

Объектом исследования являются поэтические тексты А.С. Пушкина.

Цель: разработка алгоритмов и методов по анализу стихотворных текстов полного собрания сочинений А.С. Пушкина для выявления в них признаков, описывающих изменение авторской стилистики. Для достижения цели выполнены задачи: проанализированы способы статистического описания текста; проведена предобработка стихотворных текстов собрания сочинений А.С. Пушкина; разработан алгоритм извлечения признаков из текста; найдено эмпирическое распределение признаков, определено наличие зависимости распределения от года создания текста; разработан классификатор текстов по периодам творчества; оценена работа классификатора, полученные результаты визуализированы. Использованы следующие методы исследования: анализ и сравнение способов статистического описания текста; синтез технологических решений для разработки алгоритма извлечения признаков; формализация извлечённых признаков путём построения эмпирического распределения; моделирование исследуемых текстов с использованием выделенных признаков, анализ полученных моделей методами машинного обучения; проведение экспериментов с применением разработанных программных средств, визуализация и анализ полученных результатов.

Актуальность. Использование информационных систем при анализе текстов на естественном языке позволяет применять в исследовании методы статистики и машинного обучения. Происходящее со временем изменение стилистики автора отражается на статистических характеристиках текста. Определить зависимость распределения признаков от периода создания произведения возможно с помощью методов машинного обучения. Для этого решается задача классификации произведений по периодам творчества автора. Те признаки, на основании которых проведена классификация, отражают изменение авторской стилистики.

В результате разработан программный модуль, предоставляющий возможность извлекать признаки из стихотворного текста, строить их эмпирическое распределение, преобразовывать текст в числовой вектор признаков. Получен классификатор стихотворных текстов А.С. Пушкина по периоду создания. Применение метода Шепли для интерпретации предсказаний нейронной сети позволило определить значимость каждого признака. Полученные результаты могут найти применение в филологических и лингвистических исследованиях.

Научная новизна работы заключается в применении методов статистического анализа стихотворных текстов А.С. Пушкина с целью исследования временных изменений авторской стилистики и разработке алгоритмов, позволяющих эти методы применить.

Гудков Степан Алексеевич / _____ 20 мая 2024 г.