

Историческая информатика

Правильная ссылка на статью:

Латонов В.В., Латонова А.В. Определение авторства «Записок декабриста И.И. Горбачевского» методами машинного обучения // Историческая информатика. 2025. № 1. DOI: 10.7256/2585-7797.2025.1.72805 EDN: QALGAU URL: https://nbpublish.com/library_read_article.php?id=72805

Определение авторства «Записок декабриста И.И. Горбачевского» методами машинного обучения

Латонов Василий Васильевич

ORCID: 0000-0002-7810-8033

кандидат физико-математических наук

Руководитель направления по исследованию данных, ПАО «Сбербанк»

117342, Россия, г. Москва, ул. Островитянова, 41, к.1, кв. 172

✉ WLatonov@gmail.com



Латонова Анастасия Вячеславовна

кандидат исторических наук

независимый исследователь

117342, Россия, г. Москва, ул. Островитянова, 41, к.1, кв. 172

✉ iskrenne_vasha_aa@mail.ru



[Статья из рубрики "Искусственный интеллект и наука о данных"](#)

DOI:

10.7256/2585-7797.2025.1.72805

EDN:

QALGAU

Дата направления статьи в редакцию:

22-12-2024

Аннотация: В представленной работе объектом исследования являются «Записки декабриста И.И. Горбачевского» – яркий образец декабристской мемуаристики, несущий отпечаток исторического самосознания участников движения. Данный источник предлагает ценные сведения о перипетиях взаимоотношений между участниками таких декабристских организаций, как Общество соединенных славян и Южное общество, содержит взгляд изнутри на ход и причины поражения восстания Черниговского полка,

предоставляет фактологический материал о судьбе заговорщиков после суда над ними и отправки в Сибирь. Вместе с тем, начавшись еще в советской историографии, по сей день остается до конца не завершенным спор об авторстве этих "Записок": фигура декабриста Горбачевского в качестве автора рядом исследователей считается чисто номинальной. Вполне очевидно при этом, что личность автора определяет специфику изложенных в "Записках" суждений и привносит в изложение неизбежный субъективный налет, а потому должна приниматься во внимание при работе с источником. Предметом исследования в представленной работе, таким образом, является не разрешенный до сих пор вопрос об авторстве «Записок». Авторами предложено решение задачи определения авторства «Записок декабриста И.И. Горбачевского» при помощи методов машинного обучения. В качестве возможных авторов рассмотрен сам И.И. Горбачевский, а также декабрист П.И. Борисов. Новизна исследования заключается в том, что для определения авторства «Записок» были применены методы машинного обучения. Авторы обучили четыре типа моделей для предсказания авторства каждого из предложений «Записок». В результате большинство предложений «Записок» были оценены, как написанные Горбачевским. Наибольший процент предложений, 69.2 %, был отнесён к Горбачевскому моделью Count Vectorizer + SVC. Точность всех моделей в среднем превышала 80 %, а у основанных на кодировании при помощи BERT в среднем была близка к 90 %. Основным выводом работы, таким образом, можно считать, что «Записки» более вероятно были написаны И.И. Горбачевским, чем П.И. Борисовым. Примененные в рамках представленного исследования методы дают еще один аргумент в пользу этой версии. Код и датасет доступны по ссылке: https://github.com/WLatonov/Gorbachevskiy_notes.

Ключевые слова:

Определение авторства, Атрибуция, Стилметрия, Машинное обучение, Нейронные сети, Бинарная классификация, B E R T, Декабристы, Записки Горбачевского, Письма Горбачевского

Введение

«Записки» И.И. Горбачевского являются одним из ценнейших источников по истории декабристского движения, созданных самими его участниками. В них освещается становление и развитие такой декабристской организации, как Общество соединенных славян, примкнувшее впоследствии к Южному обществу декабристов. Написанные в сибирской ссылке, эти записки представляют собой не только источник фактологического материала, но и оригинальную концепцию развития тайного общества, и ретроспективный «взгляд изнутри» на допущенные заговорщиками ошибки. Красной нитью через всю концепцию «Записок» проходит обличение промахов Южного общества, взгляды и поступки участников которого противопоставляются взглядам и поступкам самих «соединенных славян» (не без некоторой идеализации последних).

«Записки» в своей печатной версии занимают не многим более 100 страниц и распадаются на три раздела. Первый из них сообщает читателю окончание истории Общества соединенных славян, когда оно сливается с Южным обществом декабристов. Во втором освещается восстание Черниговского пехотного полка, заметную роль в событиях которого сыграли экс-«славяне». Третий раздел повествует о дальнейшей судьбе повстанцев – о суде над ними, отправлении в ссылку, неудачном бунте каторжан,

организованном И.И. Сухиновым. Таким образом, являясь свидетельством несомненного исторического самосознания автора, по содержащихся в них фактах «Записки Горбачевского» выходят далеко за пределы биографии его самого, кем бы он ни был.

Последняя оговорка неслучайна. Вопреки устоявшемуся в литературе названию «Записок», мы не можем однозначно утверждать, что их автором из числа декабристов был непосредственно И.И. Горбачевский. Дело в том, что первая публикация этого источника – в журнале «Русский архив» в 1882 г., вып. 2 – была представлена под заголовком «Записки Неизвестного из Общества Соединенных Славян». Издатель (П.И. Бартенев) напечатал их по полученной из Сибири анонимной рукописи, сопроводив примечанием: «Кажется, что „Записки“ эти составлены бывшим подпоручиком 8-й артиллерийской бригады Иваном Ивановичем Горбачевским; но ручаться в этом нельзя».

Обзор существующих версий авторства «Записок»

Анонимность первой публикации и породила сомнения в авторстве Горбачевского, которые подробно сформулировала одна из основательниц советского декабристоведения – М.В. Нечкина. Продemonстрировав в ходе своего анализа, что «Записки» могли точно так же принадлежать перу товарищей Горбачевского (имя которого возникло изначально как предположение обретшего рукопись издателя), она пришла к выводу, что автором источника наиболее вероятно был другой член Общества соединенных славян – П.И. Борисов [\[1, С. 136-143\]](#).

Биограф И.И. Горбачевского Г.П. Шатрова также не считала его автором «Записок» (во всяком случае, автором единственным). В монографии «Декабристы и Сибирь» (1962) [\[2\]](#) она позиционировала в качестве создателя «Записок» П.И. Борисова, апеллируя при этом ко мнению М.В. Нечкиной; в монографии о лично Горбачевском же пришла к выводу, что «Записки» составлены на основе коллективных воспоминаний с последующей литературной обработкой одним автором, а И.И. Горбачевский принимал активное участие в этом общем труде. Главным аргументом против исключительного авторства Горбачевского Г.П. Шатрова находила существенное расхождение между позицией «Записок» и взглядами Горбачевского, выраженными им в письмах 1850-х–1860-х гг. на поселении. [\[3, С. 75\]](#). О «Записках» как о результате совместно выработанной концепции вследствие обмена мнениями внутри коллектива писала и другая исследовательница, Н.П. Матханова, у которой по этому поводу читаем: «в казематском сообществе произошли превращение индивидуальной памяти в социальную, идентифицирующую группу, выработка “коллективной истории” – той общей версии прошлого, условной схемы, общей идеи, которая состояла из образов событий» [\[4, С. 160\]](#).

Вместе с тем, такая точка зрения на авторство «Записок» Горбачевского (и сами изначальные доводы М.В. Нечкиной) была подвергнута критике со стороны ряда других историков. Б.Е. Сыроечковский, Л.А. Сокольский, И.В. Порох несостыковки между «Записками» и другими текстами Горбачевского (следственными показаниями и письмами) находили несущественными и объяснимыми, а авторство Горбачевского – совершенно несомненным. (Подробную аргументацию см.: [\[5\]](#); также см. [\[6\]](#)) К аналогичным выводам в результате текстологического анализа позднее пришел также Злобин Е.В. [\[7\]](#). Уже в постсоветский период в качестве автора «Записок» И.И. Горбачевский официально был указан, например, в Большой российской энциклопедии [\[8\]](#). В то же время, в историографическом поле по-прежнему продолжает бытовать и

противоположная точка зрения, по которой Горбачевский не является автором одноименных «Записок» (см.: [\[9\]](#)).

Таким образом, можно утверждать, что вопрос с авторством «Записок» Горбачевского не решен до конца к настоящему моменту. Однако для исследования его теперь появляются новые возможности в силу развития информационных технологий. Соглашаясь с предшественниками в том, что рассмотрение вопроса об авторстве «Записок» едва ли возможно без сопоставления их с другими текстами Горбачевского, попробуем проанализировать все эти тексты при помощи стилометрии [\[10\]](#). В отличие от традиционного подхода, при котором адекватно лишь сравнение перекликающихся хоть в чем-либо источников, в нашем случае мы можем вовлечь в оборот любые тексты, гарантированно принадлежащие перу интересующих нас лиц. Так, если Б.Е. Сыроечковский, Л.А. Сокольский и И.В. Порох считали достойным сопоставления с «Записками» только письмо Горбачевского к М.А. Бестужеву 12 июня 1861 г., содержащее большой материал о декабристском движении в целом и об отдельных декабристах, то мы теперь можем привлечь к сравнению даже те письма Горбачевского, в которых нет никаких суждений по декабристской тематике. То же относится и к материалам других членов Общества соединенных славян, которые могли принимать участие в создании «Записок», и прежде всего – к материалам П.И. Борисова.

Существует также версия, что «Записки Горбачевского» были написаны коллективом авторов. Однако достоверных сведений о том, какие именно авторы и в какой пропорции могли принимать участие в создании «Записок», нет. Следовательно, в этом дискурсе не может быть сформулирована корректная задача стилометрии и не могут быть сделаны сколь угодно значащие выводы. Поэтому в рамках данной работы мы сосредоточим свое внимание на проверке версий авторства «Записок» в лице И.И. Горбачевского и П.И. Борисова.

Обзор работ по стилометрии и постановка задачи

Стилометрия — это дисциплина, занимающаяся измерением стилистических характеристик с целью упорядочивания и систематизации текстов [\[10\]](#). Эти характеристики могут быть вычислены для любого достаточно большого авторского текста, и для каждого авторского стиля эти характеристики будут уникальны. Таким образом, стилометрия может быть использована для определения авторства текста при наличии образцов текстов возможных авторов, достаточно больших для вычисления характеристик.

Сама по себе задача определения авторства существует уже много столетий, однако впервые формализована она была, по-видимому, в работе Н.А. Морозова [\[11\]](#). В этой работе был предложен метод идентификации авторства через графики частот употребления слов. Также в числе первых работ по математическому исследованию стилистики авторов следует назвать статью [\[12\]](#) А.А. Маркова. В этой статье А.А. Марков применил статистический анализ, ранее изложенный им же в другой работе [\[13\]](#), где оценивалась вероятность буквы быть гласной в зависимости от цепи из двух предыдущих букв.

В работе [\[14\]](#) была применена идея цепей из двухбуквенных сочетаний из работы [\[13\]](#) и было показано, что этот подход позволяет определять истинного автора с вероятностью 84% при рассмотрении 80 возможных вариантов. Обобщение этого метода было предложено в работе [\[15\]](#), где в качестве единиц анализа рассмотрены не только двухбуквенные сочетания, но и одиночные грамматические классы слов, а также пары

слов. Особый интерес эта работа вызывает в связи с тем, что, согласно выводам авторов, определение авторства текста по двухбуквенным сочетаниям более точно, чем по одиночным словам и парам слов. Аналогичный результат был получен и в работе [16], где было показано, что трёхбуквенные сочетания позволяют устанавливать авторство точнее, чем слова.

В работах [17, 18] для установления авторства были рассмотрены вероятности появления разных n -буквенных сочетаний для $n > 2$, называемых n -граммами. В некоторых работах n -граммами также называют последовательности из n слов, которые также подходят для определения авторства, например в статье [19]. Авторы выяснили, что для английского языка 6-граммы дают наилучший результат распознавания автора, в то время, как, например, для греческого наилучший результат достигается при использовании 3-грамм.

Ещё один известный подход к атрибуции авторства – Дельта Бёрроуза [20] – метод, опубликованный в 2002-м году. В своей работе Джон Бёрроуз ввёл метрику, называемую Дельтой, используемую для измерения расстояния между текстами [21]. Дельта рассчитывается на основе всего словаря слов, используемых во всех текстах, между которыми считается расстояние. Эта метрика учитывает частоту использования каждого слова в отдельно взятом тексте и частоту использования слова во всём наборе текстов. Дельта Бёрроуза широко применяется в лингвистических исследованиях [22, 23] в том числе для определения меры проявления стиля авторов в совместной работе [24].

Наряду с этим в отечественной историографии существует ряд работ, посвященных определению особенностей авторского стиля. Результатом трудов Л.В. Милова, Л.И. Бородкина и других историков в 1970-1980-е годы стали исследования [25-27], где для задач анализа стиля и атрибуции использовался сетевой анализ структур грамматических связей. Во многом объектом этих исследований явились средневековые русские тексты.

В современных работах по определению авторства часто используются методы машинного обучения (ML). Так, в работе [28] авторы применили ряд классических методов в задаче атрибуции авторства на примере литературных текстов. В числе прочих были опробованы метод k ближайших соседей (KNN) [29] и метод опорных векторов (SVC) [30]. Отдельное внимание авторы уделяют вопросу предобработки текстов: рассмотрено обучение как на исходном тексте, так и на тексте с удалёнными редко встречающимися словами.

Также находит применение в задаче атрибуции авторства и глубокое обучение, например, авторы [31] использовали свёрточную нейронную сеть (CNN) и сравнили её точность с другими подходами, в частности, многослойным линейным персептроном (MLP) [32], а также с уже упомянутыми выше KNN и SVC. В работе [33] авторы решили задачу определения авторства русскоязычных текстов, строя процесс обучения на текстах из классической литературы и коротких публикациях в социальных сетях. Наряду с классическими ML-методами, авторы применили нейронные сети, включающие такие архитектуры, как LSTM [34] и BERT [35].

В настоящей работе решена задача определения авторства «Записок Горбачевского». В качестве материала для обработки при этом использована публикация «Записок» 1963 г. в «Литературных памятниках», для которой существует цифровая копия в сети Интернет, что делает ее использование более удобным. От оригинальной публикации 1882 г. эта

версия отличается местами иным членением на абзацы, что не является существенным для нашей задачи, а также наличием кратких аннотаций перед главами (эти аннотации нами при обработке текста удалялись). В качестве кандидатов на авторство «Записок» рассмотрены сам И.И. Горбачевский и П.И. Борисов, для решения использованы образцы текстов, написанных двумя этими декабристами. Для изучения стиля Горбачевского использовались его письма – источник, прежде всего, хранящий информацию об отношениях между ссыльными декабристами и об условиях их жизни на поселении. Вместе с тем, в этих письмах содержится достаточно суждений Горбачевского о декабристском заговоре. Для анализа доступно 81 его письмо за период 1839–1868 гг. Для исследования стиля Борисова, как и в случае с Горбачевским, также использовались письма. Писем Борисова в нашем распоряжении сравнительно немного, 20. Все они написаны в 1838–1847 гг. в период его жизни на поселении и по своей тематике весьма близки к письмам Горбачевского.

Решение задачи

Для решения поставленной задачи мы использовали классические ML-методы: SVC и логистическая регрессия (LR) [36]. Последняя была выбрана постольку, поскольку ориентирована на решение задачи бинарной классификации, а в нашей постановке возможных авторов лишь двое. Стоит отметить, что эти методы не могут быть применены сами по себе к текстовым данным, поэтому каждое предложение кодируется в вектор числовых признаков.

Для кодирования были использованы два метода. Первый из них, Count Vectorizer (CV), относится к классическим ML-подходам. Каждое предложение перед кодированием с помощью CV проходило предобработку:

1. Замена верхнего регистра на нижний;
2. Удаление знаков препинаний, круглых и квадратных скобок, а также кавычек;
3. Удаление служебных слов: предлогов, союзов, частиц, междометий.

Второй выбранный метод кодирования – предобученная на русскоязычных текстах модель BERT (использовалась модель, доступная по ссылке: <https://huggingface.co/papers/2408.12503>). Этот подход позволяет добиться большей точности при обучении, чем при кодировании классическими методами, поскольку, будучи предобученным, уже содержит информацию о русскоязычных текстах и том, как их эффективно кодировать. Таким образом, были обучены модели предсказания:

1. Count Vectorizer + SVC;
2. Count Vectorizer + LR;
3. BERT + SVC;
4. BERT + LR;

Схема моделей 1-2 изображена на рисунке 1, моделей 3-4 – на рисунке 2.

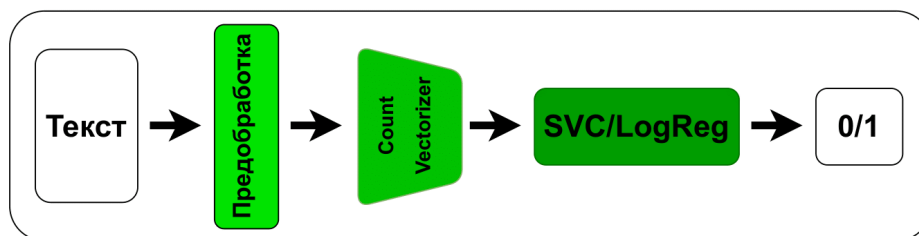


Рисунок 1.

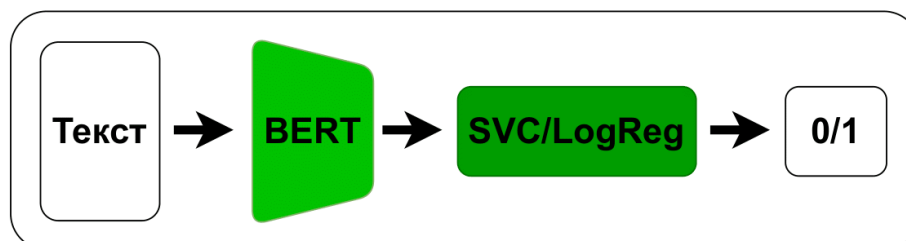


Рисунок 2.

Каждая модель на вход принимает одно предложение и классифицирует его, как принадлежащее авторству Горбачевского или Борисова. Предобработка текстов и обучение моделей были реализованы на языке Python. Реализация моделей SVC, LR, а также кодировщика Count Vectorizer были взяты из библиотеки sklearn, для предобработки использовались библиотеки pandas и numpy.

В письмах Борисова насчитывается в общей сложности 411 предложений, в то время, как в письмах Горбачевского их 2620. Для обучения необходимо равное число предложений обоих авторов, поэтому из писем Горбачевского требовалось выбрать 411 предложений. Выборка определялась генератором случайных чисел библиотеки numpy, причём было взято сто разных выборок при помощи параметра seed, варьируемого в диапазоне от 0 до 99. Для каждой из выборок предложений Горбачевского была обучена каждая из четырёх выбранных моделей. При обучении 80 % предложений отводилось под тренировочную выборку, а 20 % – под тестовую. Для каждой модели были подобраны наилучшие гиперпараметры методом Grid Search.

Результаты

Нами было получено четыре блока по 100 моделей. После обучения каждая модель использовалась для предсказания автора каждого предложения из «Записок» по отдельности. Доли предложений, отнесённых моделями разных блоков к Горбачевскому, представлены на Рисунках 3-6. Также на этих графиках малиновой пунктирной линией отмечены средние доли предложений, отнесённых к Горбачевскому. Среднее взято по всем моделям в рамках одного блока. Малиновой сплошной линией отмечены максимальные и минимальные доли. Например, как видно из Рисунка 4, модель Count Vectorizer + LR, обученная на одной из ста выборок классифицировала больше 80 % предложений «Записок», как принадлежащие авторству Горбачевского. В то же время, минимальная доля у этой модели превосходит 57 %. У моделей Count Vectorizer + SVC и BERT + SVC максимальная доля чуть больше 79 %. Минимальные доли у этих моделей превосходят 51 % и 48 % соответственно. В Таблице 1 приведены примеры матриц ошибок четырёх моделей на тестовых выборках – по одной модели из каждого блока. Видно, что точность моделей, использующих BERT выше, чем остальных. У всех моделей во всех блоках точность на обучающей и тестовой выборках отличалась на 1-3 %. Также

видно, что модели, не использующие BERT имеют более высокую точность на предложениях Горбачевского, а модели с BERT имеют примерно одинаковую точность на обоих авторах. В Таблице 2 приведены средние доли по выборкам для всех моделей, а также средние значения точности на обучении в рамках каждого блока.

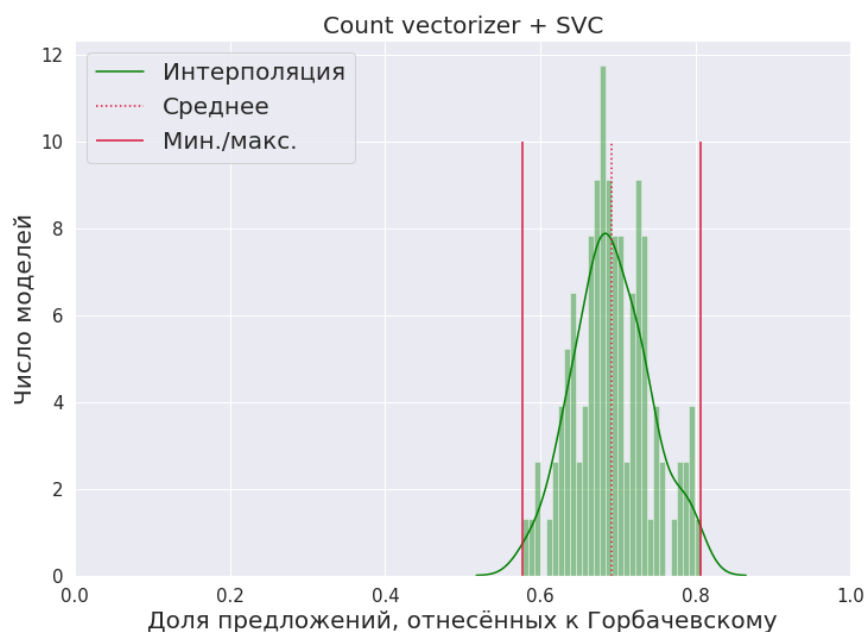


Рисунок 3.

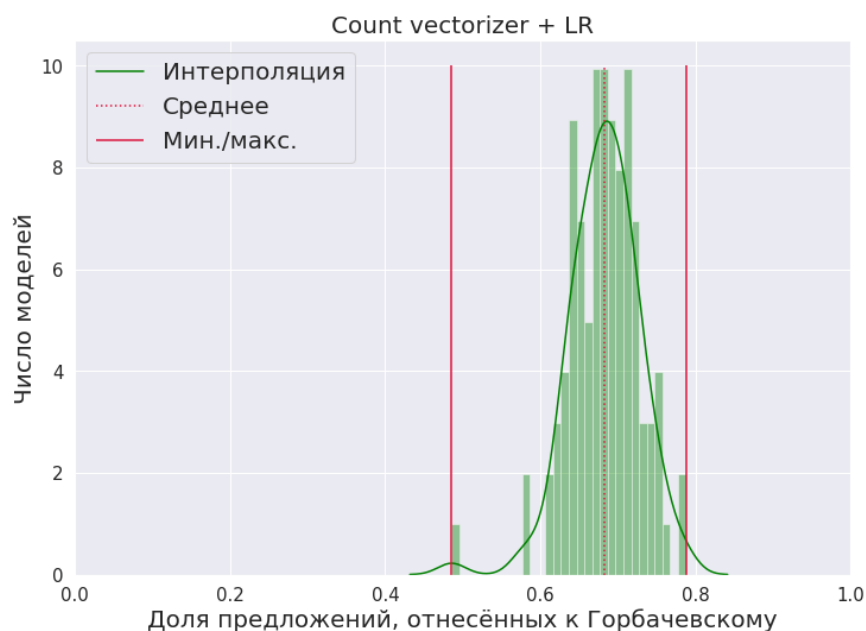


Рисунок 4.

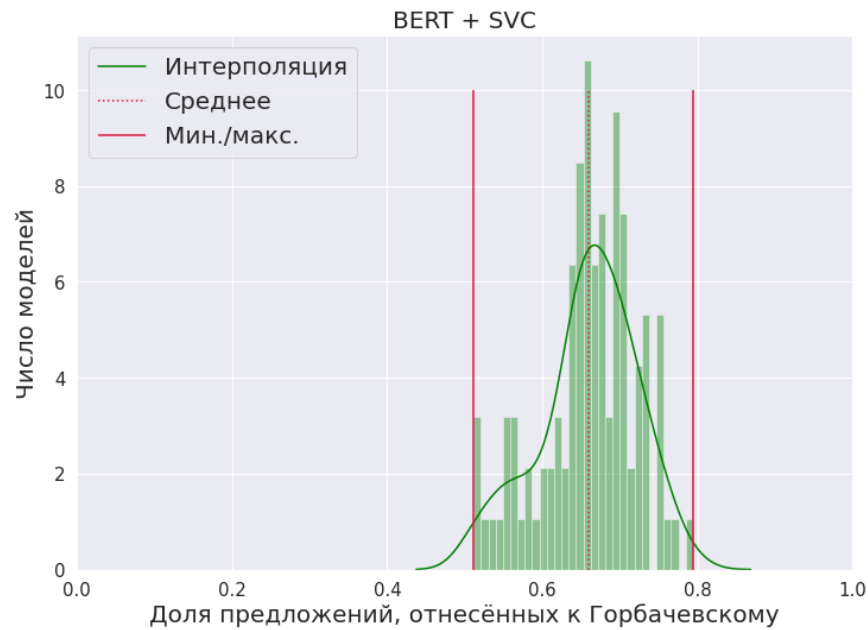


Рисунок 5.

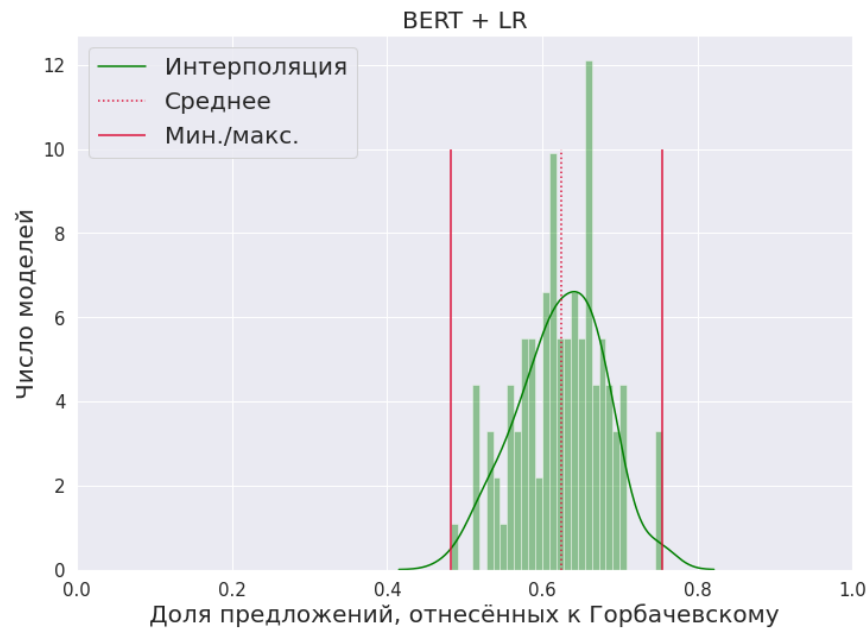
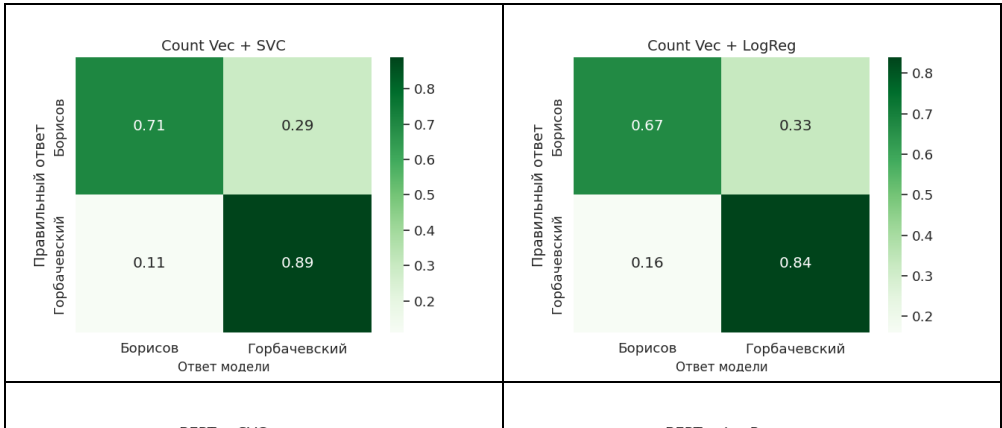


Рисунок 6.



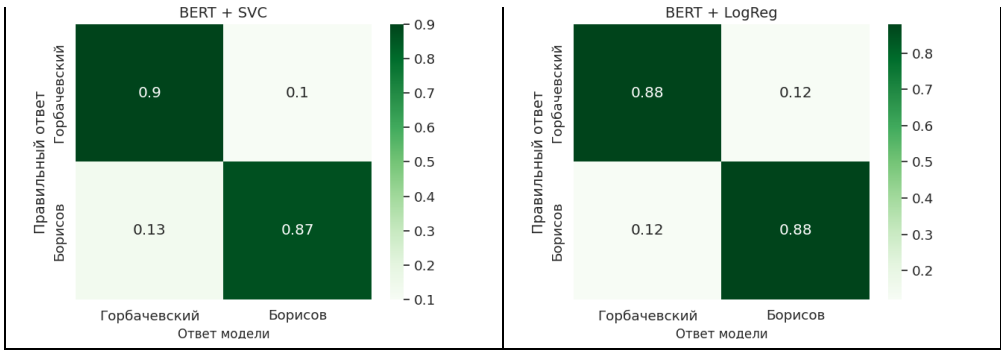


Таблица 1. Примеры матриц ошибок моделей на тестовых выборках.

Модель	Средняя точность на обучении	Средний процент предложений	
		Борисов	Горбачевский
Count Vectorizer + SVC	0.80	30.8 %	69.2 %
Count Vectorizer + LR	0.81	31.8 %	68.2 %
BERT + SVC	0.89	34.0 %	66.0 %
BERT + LR	0.88	37.6 %	62.4 %

Таблица 2. Точность моделей и предсказания авторства предложений из «Записок».

Большинство обученных моделей классифицируют около 70 % предложений «Записок», как написанные Горбачевским. Более точные модели (использующие BERT) относят к Горбачевскому чуть меньше предложений «Записок», однако это всё равно в среднем более 64 %. Разброс в результатах классификации, который виден на рисунках 3-6, объясняется тем, что для обучения ста моделей в каждом из блоков использовались сто разных выборок. Тем не менее этот разброс не влияет на интерпретацию результата, поскольку на графиках виден значительный перевес в пользу Горбачевского в каждом из четырех блоков.

Выводы

В работе рассмотрена задача определения авторства «Записок Горбачевского», с предположением, что автором мог быть либо сам И.И. Горбачевский, либо П.И. Борисов. Используемые методы машинного обучения в последние годы зарекомендовали себя, как наиболее точные в задачах атрибуции авторства, и в нашей работе на обучении показали точность более 80 % (BERT + SVC и BERT + LR – около 90 %). Почти все обученные модели классифицировали около 70 % предложений «Записок», как написанные Горбачевским. Таким образом, можно сделать вывод, что «Записки Горбачевского» следует атрибутировать как принадлежащие перу реального И.И. Горбачевского, а не П.И. Борисова, если в качестве вариантов выбора рассматривать дихотомию из этих двух авторов.

Библиография

1. Нечкина М.В. Движение декабристов. [В 2 т.] М.: Изд-во Акад. наук СССР, 1955. Т. 2.
2. Шатрова Г.П. Декабристы и Сибирь. Томск: Издательство Томского университета, 1962.

3. Шатрова Г.П. Декабрист И.И. Горбачевский. Красноярск: КГПИ, 1973.
4. Матханова Н.П. Сибирская мемуаристика XIX века. Новосибирск: Издательство СО РАН, 2010.
5. Сыроечковский Б.Е., Сокольский Л.А., Порох И.В. Декабрист Горбачевский и его «Записки» // И.И. Горбачевский. Записки; Письма / Изд. подгот. Б.Е. Сыроечковский, Л.А. Сокольский, И.В. Порох. М: Изд-во АН СССР, 1963. С. 257-305.
6. Мироненко М.П. Мемуарное наследие декабристов в журнале «Русский архив» // Археографический ежегодник за 1975 год. М., 1976. С. 112-114.
7. Злобин Е.В. К вопросу об авторстве "Записок" декабриста И.И. Горбачевского // История СССР, 1990. № 2. С. 140-155.
8. Горбачевский Иван Иванович // Большая российская энциклопедия: [в 35 т.] / гл. ред. Ю.С. Осипов. М.: Большая российская энциклопедия, 2004–2017. Т. 7. М., 2007. С. 423-424.
9. Туманик Е.Н. Роль мемуарного наследия декабристов в научной концепции Г.П. Шатовой // Гуманитарные науки в Сибири, 2020. № 27(4). С. 50-57.
10. Мартыненко Г. Я., Гребенников А. О. Основы стилистики: учеб.-метод. пособие. СПб.: Изд-во С.-Петербург. ун-та, 2018.
11. Морозов Н.А. Лингвистические спектры: средство для отличения плагиатов от истинных произведений того или иного известного автора. Стилистический этюд // Изв. отд. русского языка и словесности Имп. акад. наук. 1915. № 20(4). С. 93-134.
12. Марков А.А. Об одном применении статистического метода // Изв. Имп. акад. наук. Сер. 6. 1916. № 4. С. 239-242.
13. Марков А.А. Пример статистического исследования над текстом "Евгения Онегина", иллюстрирующий связь испытаний в цепь // Изв. Имп. акад. наук. Сер. 6. 1913. № 3. С. 153-162.
14. Хмелёв Д.В. Распознавание автора текста с использованием цепей А.А. Маркова // Вести. МГУ. Сер. 9. Филология. 2000. № 2. С. 115-126.
15. О. В. Кукушкина, А. А. Поликарпов, Д. В. Хмелёв, Определение авторства текста с использованием буквенной и грамматической информации // Пробл. передачи информ. 2001. № 37(2), 96-109.
16. Stamatatos P. D. et al. On the robustness of authorship attribution based on character n-gram features // Journal of Law and Policy. 2013. № 21(2). P. 7.
17. Burrows S., Tahaghoghi S. M. M. Source code authorship attribution using n-grams // Proceedings of the twelfth Australasian document computing symposium, Melbourne, Australia, RMIT University, 2007. P. 32-39.
18. Sapkota U. et al. Not all character n-grams are created equal: A study in authorship attribution // Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies. 2015. P. 93-102.
19. Peng F. et al. Language independent authorship attribution with character level n-grams // 10th Conference of the European Chapter of the Association for Computational Linguistics. 2003.
20. Burrows J. 'Delta': a measure of stylistic difference and a guide to likely authorship // Literary and linguistic computing. Oxford University Press. 2002. № 17(3). P. 267-287.
21. Hoover D. Testing Burrows' Delta. // Literary and Linguistic Computing. 2004. № 19(4). P. 453-475.
22. Evert. S., Proisl T., Jannidis F., Reger. I., Pielström S., Schöch C., Vitt T. Understanding and explaining Delta measures for authorship attribution. // Digital Scholarship in the Humanities. 2017. № 32(2). P. 4-16.
23. Jannidis F. et al. Improving Burrows' Delta. An empirical evaluation of text distance measures // Digital Humanities Conference. 2015. № 11. P. 10.

24. Ковалев Б.В. Рождение третьего автора: стилеметрический анализ рассказов Онорио Бустоса Домека // Литература двух Америк. 2024. № 16. С. 120-146.
25. Бородкин Л.И., Милов Л.В., Морозова Л.Е. К вопросу о формальном анализе авторских особенностей стиля в произведениях Древней Руси // Математические методы в историко-экономических и историко-культурных исследованиях / Отв. ред. И.Д. Ковальченко. М.: Наука. 1977. С. 298-326.
26. Borodkin L., Milov L. Some Aspects of the Application of Quantitative Methods and Computers in the Analysis of Narrative Texts // Soviet Quantitative History / Ed. by D.K. Rowny. Sage Publications. Beverly Hills/London/New Delhi, 1984.
27. Милов Л.В., Бородкин Л.И., Иванова Т.В. и др. От Нестора до Фонвизина: Новые методы определения авторства / под ред. Л. В. Милова. М., Прогресс, 1994.
28. Jockers M. L., Witten D. M. A comparative study of machine learning methods for authorship attribution // Literary and Linguistic Computing. 2010. № 25(2). С. 215-223.
29. Fix E., Hodges J. L. Discriminatory analysis, nonparametric discrimination // International Statistical Review / Revue Internationale de Statistique. 1989. No. 57(3), P. 233-238
30. Vapnik V. Support-vector networks // Machine learning. 1995. № 20. P. 273-297.
31. Bumber D., Zhang Y., Mukherjee A. Experiments with convolutional neural networks for multi-label authorship attribution // Proceedings of the Eleventh International Conference on Language Resources and Evaluation, 2018.
32. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain // Psychological review. 1958. № 65(6). P. 386.
33. Fedotova A. et al. Authorship attribution of social media and literary Russian-language texts using machine learning methods and feature selection // Future Internet. 2021. № 14(1). P. 4.
34. Hochreiter S. Long Short-term Memory // Neural Computation MIT-Press, 1997.
35. Kenton J. D. M. W. C., Toutanova L. K. Bert: Pre-training of deep bidirectional transformers for language understanding // Proceedings of naacL-HLT. 2019. № 1. P. 2.
36. Hosmer D. W. Lemeshow S. Applied Logistic Regression // John Wiley & Sons, 2013.

Результаты процедуры рецензирования статьи

Рецензия скрыта по просьбе автора