

Анализ тональности высказываний в области репродуктивного поведения методами естественной обработки языка при помощи данных социальной сети ВКонтакте

Колотуша А.В.

Введение

В демографических исследованиях давно созрела потребность в использовании новых источников данных, которые превосходили бы данные регистров, переписей и выборочных обследований как по ширине охвата населения, так и по глубине и частоте измерения. За последние десятилетия демографы возлагают всё большие надежды на направление, связанное с использованием больших данных, для развития которого появляется всё больше и больше возможностей.

В настоящее время всё больше работ, посвящённых извлечению авторских мнений по демографическим вопросам, не обходится без использования больших данных. Из недавних работ можно отметить исследования, посвящённые анализу отношений людей к абортам [Hasan, Ng 2013; Sharma et al 2017; LaRoche et al 2021; Roldán-Robles et al. 2019; Ntontis, Hopkins 2018], обсуждению разных аспектов родительства [Irazú et al. 2014], анализу влияния различных факторов на демографические процессы вплоть до природных катастроф [Mandel et al 2012] и пандемии COVID-19 (включая вакцинацию для детей) [Miao et al. 2020; Glandt et al 2021; Liu, Liu 2021; Thorpe Huerta et al 2021; Abosedra et al. 2021]. Встречаются и исследования, где методы работы с большими данными выступают инструментальными средствами для получения материалов для качественных исследований в демографии: доводов за и против абортов [Hasan, Ng 2013], за и против вакцинации [Vychezhzhanin, Kotelnikov 2019].

Помимо введения в оборот и апробации новых данных в демографии, проведение подобных исследований могло бы представляться важным и для демографической политики. С одной стороны, мы не представляем возможным проведение разумной демографической политики без учёта мнений и настроений её прямых адресатов. С другой стороны, мы надеемся, что сбор больших данных по демографическим вопросам из социальных сетей или новостных ресурсов при должной отработке в долгосрочной перспективе окажется менее дорогостоящим по сравнению с существующими проектами выборочных обследований, регистров и переписей населения.

Направление анализа, которым плотно занимаемся наша исследовательская команда, также выполняется в русле взаимодействия с большими данными в демографии и связано с измерением отношения людей к процессу деторождения. Насколько нам известно, это – первое масштабное исследование на эту тему, по крайней мере, на российском пространстве.

Обзор предшествующих исследований рабочей команды

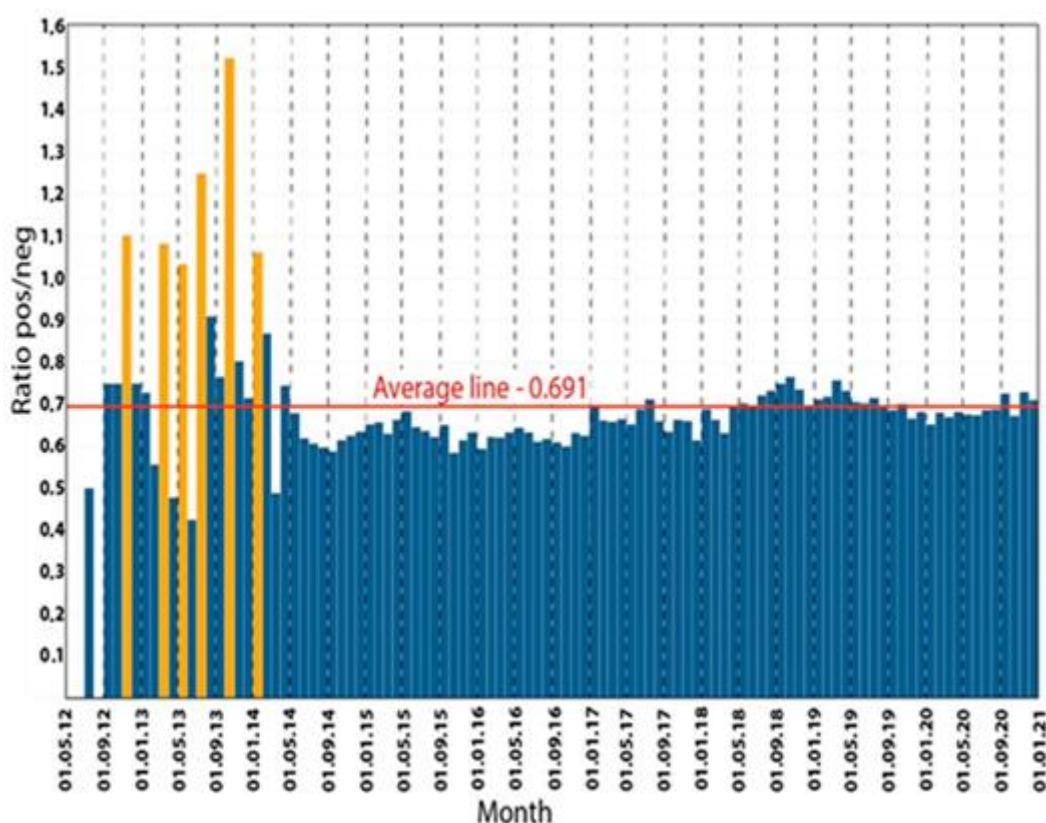
В более ранних наших исследованиях мы провели первую попытку построения индикатора тональности комментариев и назвали его демографической температурой [Kalabikhina et al 2021]. При помощи библиотек TensorFlow и tflearn в программном пакете Python и данных Твиттера по аналогичной выборке мы обучили трёхслойную нейронную сеть (количество нейронов в слоях¹: 125-25-2, функции активации: ReLU-ReLU-Softmax), которую

¹ Последний слой из двух нейронов – распределение вероятностей того, что данный комментарий является позитивным (p) или негативным (q) ($p, q > 0, p+q=1$). Если $p \geq 0,5$, пост считался позитивно окрашенным, если $p < 0,5$ – негативно.

впоследствии применили к комментариям из тематических групп про- и антинаталистов² из социальной сети ВКонтакте. Мы брали именно данные комментариев, поскольку в них, как мы заметили, в большей степени была представлена эмоциональная, нежели содержательная позиция. Построенная сеть позволила нам определить окраску предварительно обработанных (стеммизированных, лемматизированных и очищенных от пунктуации) комментариев, что в дальнейшем дало нам возможность получить общие понедельные оценки «демографической температуры» в стране за определённый период. Демографическую температуру мы рассчитываем как отношение позитивных комментариев к негативным за определённый период и называем её позитивной, если количество комментариев, определённых как позитивные, преобладает над негативно определёнными, и негативной – иначе.

Мы постарались, чтобы наша терминология была интуитивно понятной, поэтому мы окрасили позитивную демографическую температуру в жёлтый цвет, а негативную – в синий, что символизирует в первом случае в среднем тёплое, дружественное расположение людей к теме деторождений, а во втором – достаточно холодное, подчас враждебное отношение к этой теме.

Мы провели отдельные замеры для групп про- и антинаталистов за период с 2012 по 2020 годы³ (рис. 1, 2):



² Людей, относящихся к процессу деторождения позитивно и негативно - соответственно

³ Исходя из доступности данных

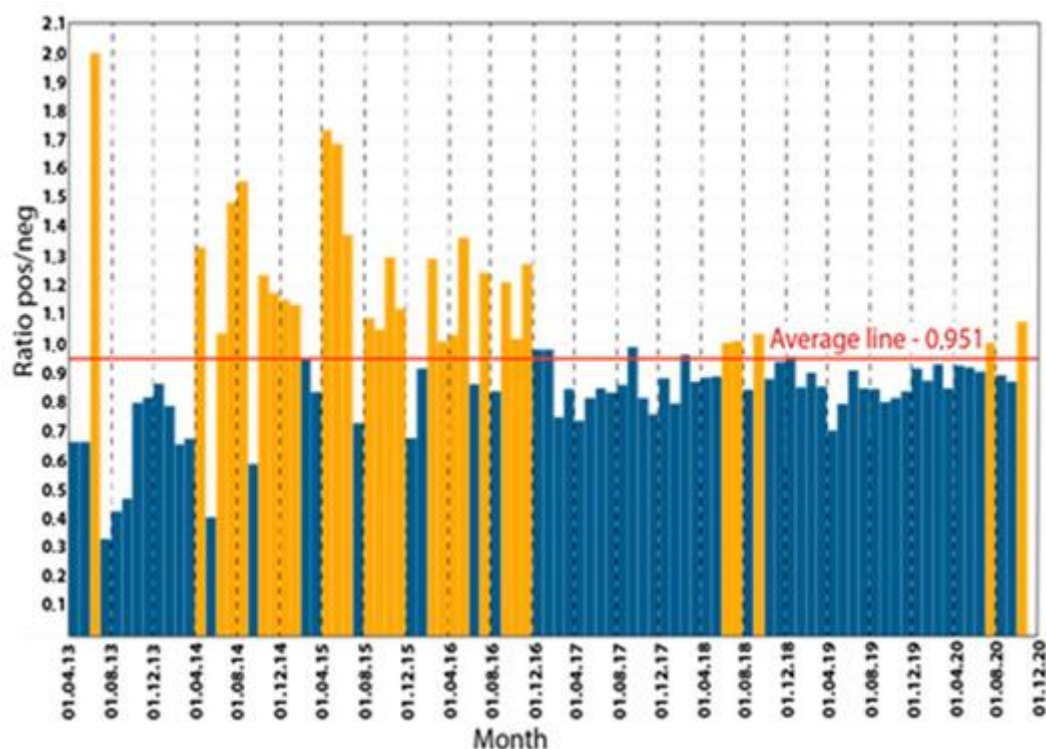


Рисунок 1-2 – Соотношение негативных и позитивных комментариев в корпусе антинаталистов (1) и пронаталистов (2) по месяцам.

Источник: Построено авторами в Tableau 19.3 на основе собранных данных из социальной сети «ВКонтакте»

Замеры указывают, во-первых, на то, что в группах антинаталистов периоды «потепления» демографической температуры заметно более редки, чем в группах пронаталистов. Во-вторых, периоды потепления температуры в группах пронаталистов очень хорошо рифмуются с датами наступления значимых событий в российской политике в области материнского капитала

Полученный результат с методической точки зрения может говорить о том, что у этих данных есть потенциал, чтобы в будущем войти в активный оборот в кругу исследователей-демографов.

Вместе с тем, проделанная работа была не лишена ряда недостатков. Во-первых, мы обучали сеть на данных Твиттера вместо фрагмента исходного датасета. Это было сделано по причине отсутствия собственной базы ручных разметок из собранного датасета на предыдущем этапе исследования из-за трудоёмкости задачи и ограничений предыдущего этапа исследования по времени, поэтому мы опирались на те размеченные данные, которые имелись в нашем распоряжении на тот момент. Безусловно, для большей надёжности результатов было бы предпочтительнее обучать сеть на основе словаря из массива комментариев той же природы, что и исходный массив данных. Во-вторых, за счёт того, что предобработка данных предполагала удаление пунктуации, были автоматически удалены из комментариев и символы, предназначенные для выражения эмоций. Это – серьёзный недостаток, поскольку в пунктуации может содержаться значимая информация для определения эмоционального тона поста. Наконец, новейшие исследования указывают на то, что BERT даёт лучшие результаты анализа авторских позиций, нежели

использованный нами простой тензор, по таким мерам, как правильность классификации (Accuracy) и среднее гармоническое из точности и полноты (F-мера) [Devlin et al., 2018].

Всё это позволяет нам учесть ошибки предыдущего этапа при дальнейшем продвижении исследования.

Описание задачи возрастной дезагрегации

Исследование на предыдущем этапе носило достаточно пионерный характер и представляет собой открытое поле для дальнейшего развития на базе существующих результатов. Один из вопросов, который напрашивается для развития исследований, касается изучения того, есть ли возрастные различия в отношении людей к деторождениям. В свете второго демографического перехода⁴, концепция которого была выдвинута Р. Лестегом и Д. Ван де Каа ещё в середине 80-х годов 20-го века, но который как никогда актуален в России в настоящее время, мы можем ожидать того, что широкое распространение индивидуалистически ориентированной системы ценностей должно найти отражение в меньшем стремлении заводить собственных детей и, как следствие, менее позитивной тональности комментариев по вопросам деторождений со стороны представителей более молодых репродуктивных возрастных групп по сравнению с более возрастными из числа репродуктивных.

Проверке вышеуказанной гипотезы при помощи сформированного ранее массива данных и немного модифицированной методологии анализа и будет посвящена дальнейшая работа.

Так выглядит видение глобальной цели проекта. В рамках данной работы мы хотим научить нейронную сеть распознавать отношение людей к деторождениям, чтобы в дальнейшем иметь возможность прикладывать рабочую технологию к объёмным массивам пользовательских комментариев. Это могло бы дать нам возможность проверить исследовательскую гипотезу надёжным образом на больших данных из социальных сетей.

Данные

Мы опираемся на данные, подвергнутые процедуре разметки силами нашего рабочего коллектива и внешних исполнителей (преимущественно демографами и лингвистами) [Калабихина и др., 2021]. Для обеспечения надёжности разметок каждый комментарий из специально отобранного подмассива⁵ размечался тремя аннотаторами. В рамках разметки каждый аннотатор высказывал экспертное заключение по тональности текста по 6 темам (бездетность, индивидуализм, многодетность, аборты, выплаты и декретные отпуска). Это даёт нам возможность в дальнейших исследованиях измерить демографическую температуру в многообразии её аспектов. Мы использовали шесть типов меток. Первая из них («нерелевантно») позволяла отбросить из рассмотрения текст, не соответствующий определённой тематике. Остальные пять меток включали в себя классические «положительно» и «отрицательно», а также метки, которые мы относили к классу «прочее»: «нейтрально» (текст, не имеющий определённой тональности), «и положительно, и отрицательно» (текст, в котором отражались одновременно и позитивные, и негативные

⁴ Качественное изменение в отношении людей к организации жизни семьи от традиционных взглядов в сторону большей открытости к новым моделям поведения (в том числе, репродуктивного). См, например, работы Ван де Каа [van de Kaa, 1987]

⁵ Отбирались из общего пула только те, которые соответствовали про- или антинаталистской направленности

аспекты)⁶ и «невозможно определить» (текст, по которому неясно из контекста, как его маркировать).

Из 3 разметок по каждому тексту окончательное решение о тональности комментария мы принимали на основе большинства голосов. Мы исключали из выборки тексты, в отношении которых аннотаторы сильно расходились в оценках⁷. До окончательного датасета дошло 4907 комментариев. Данные были соответствующим образом предобработаны (очищены от пунктуации и заглавных букв).

Приведём ниже примеры разметок на примере темы с абортom. Пример позитивного высказывания по абортom: «если бы в моём живoте завелось подобное, я бы побежала на аборт теряя тапки». Пример негативного высказывания по абортom: «не делайте аборт, какими не были бы обстоятельства». Пример неясного высказывания: «-мама, а что такое аборт? -сынoк, это» — вроде бы упомянут аборт, но отношение к нему непонятно.

Распределение данных по классам

В исходном распределении данных по классам наблюдается определённый дисбаланс (рис. 3):

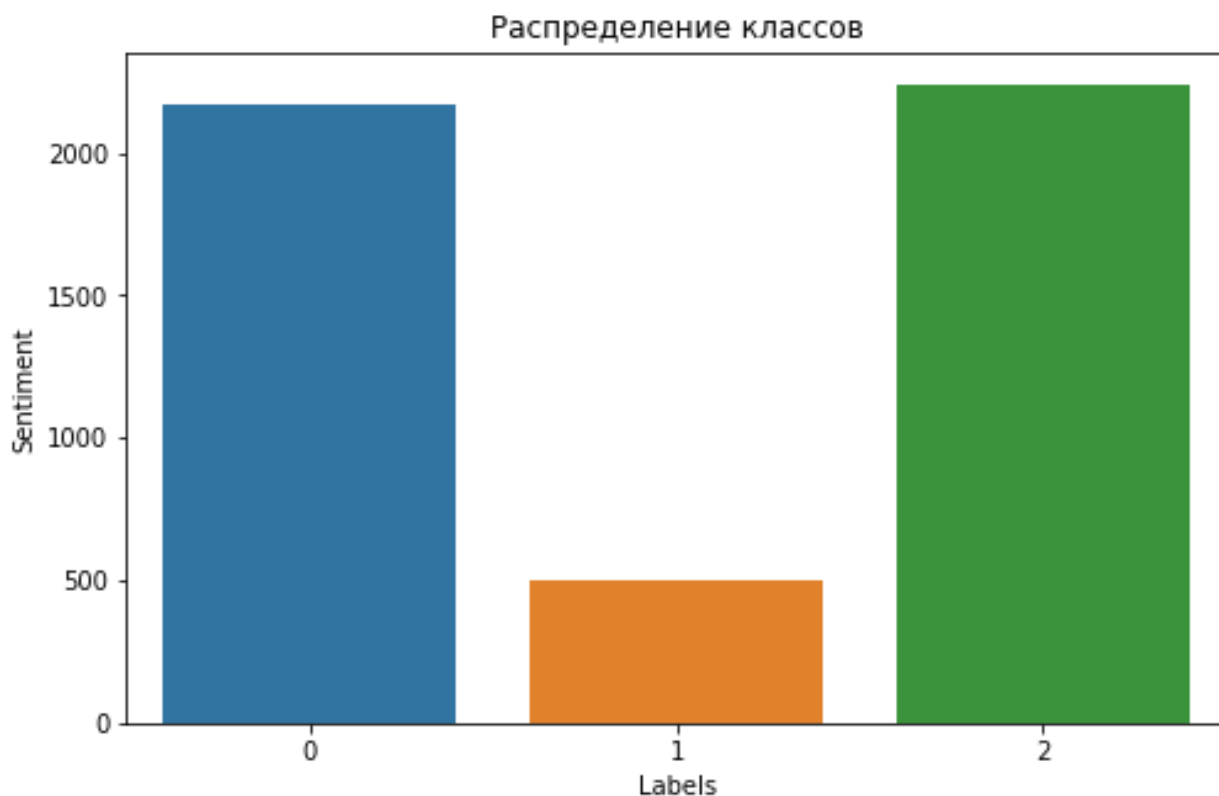


Рисунок 3 – Распределение разметок по классам – негативные («0»), «прочие» («1») и позитивные («2»)

Источник: здесь и далее - построено автором на данных из социальной сети «ВКонтакте»

⁶ Например, «хорошо, что материнский капитал проиндексировали, но плохо, что по недостаточно высокой ставке»

⁷ Например, если первый аннотатор оценивал текст как положительный, второй – как отрицательный, а третий выражал оценку, которая переводила комментарий в категорию «прочее»

Из исходных 4907 разметок 2168 отмаркированы как негативные («0»), 2239 как позитивные («2»), 500 разметок относятся к категории «прочие» («1»): либо нейтральные, либо тональность высказываний смешанная, либо тональность невозможно однозначно уловить.

В связи с этим, мы поменяли метки классов, предварительно удалив разметки из категории «прочие» и устранив слишком короткие наблюдения.

В результате итоговый датасет составил 4406 наблюдений. В окончательном датасете распределение данных по классам выглядит более сбалансированным (рис. 4).

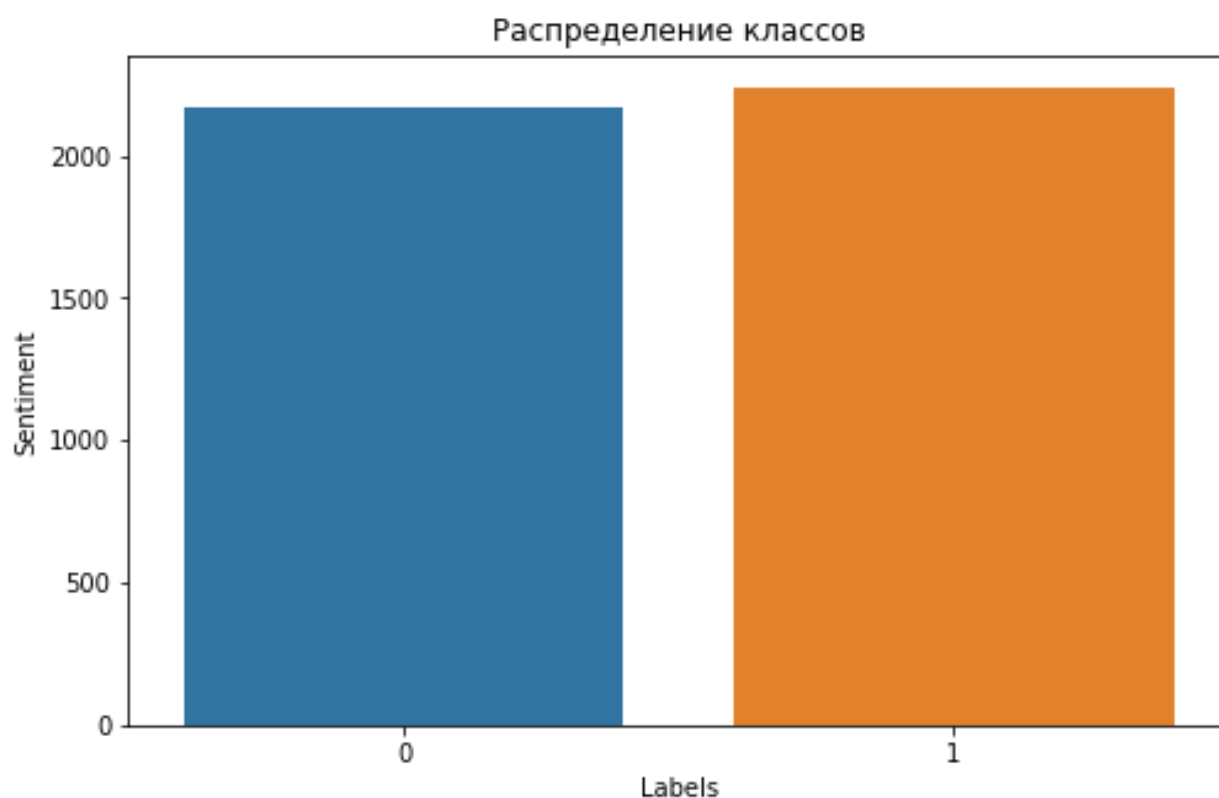


Рисунок 4 – Распределение разметок по классам – негативные («0») и позитивные («1»)

Итоговый датасет содержит 2167 негативных разметок («0») и 2239 позитивных («1»).

Деление train-valid-test

По источникам получения комментариев в окончательном датасете у нас также наблюдается дисбаланс классов: из итоговых 4406 комментариев только 348 текстов относились к группам пронаталистов. В связи с этим, было решено не разделять выборку на подвыборки про- и антинаталистов.

Исходная выборка из 4406 размеченных наблюдений может дать результаты пускай на грани, но репрезентирующие генеральную совокупность пользователей социальной сети ВКонтакте по показателю возраста. Вместе с тем, если отводить 1, 5 или 10% исходной выборки на валидационное и тестовое множества, количество наблюдений представляется нам недостаточным, чтобы результатам подбора гиперпараметров сети⁸ можно было бы надёжно доверять, а результаты тестового множества можно было бы признать

⁸ В частности, количества эпох на основе F-меры

репрезентативными. В связи с этим, для повышения надёжности результатов, мы берём валидационное и тестовое множество в объёме 20% от всего объёма выборки. Таким образом, в итоговое обучающее множество (train) мы отводим 2644 наблюдения (60% выборки), а в валидационное (valid) и тестовое (test) – по 881 (по 20% выборки).

Обоснование выбора модели

В качестве опорной модели мы используем RuBERT (подвид нейросетевой модели BERT, который был дообучен на русскоязычных материалах диалогов и текстов из социальных сетей), как ранее успешно зарекомендовавший себя по формальным метрикам качества (Accuracy, F-мера) [Devlin et al., 2018; Калабихина и др., 2021].

Поскольку RuBERT согласно предыдущим исследованиям показал себя как наиболее успешный подход, для проверки робастности результатов не имеет особого смысла использовать альтернативные модели, такие как модели на основе NLI (natural language inference – вывод по тексту), QA (question-answering – вопросно-ответный подход), или классических моделей, таких как модель на основе метода случайного леса (Random Forest), модель градиентного бустинга (GB), или модели на основе наивного байесовского классификатора – биномиального (BNB) и мультиномиального (MNB). Однако возникает вопрос, какая из модификаций RuBERT наилучшим образом подходит для семантического анализа в плоскости репродуктивного демографического поведения.

Обзор модификаций RuBERT

Мы перепробовали 3 модели: RuBERT-Tiny, RuBERT-Tiny2-Sentence-Compression и RuBERT-Tiny-Toxicity, как компактные по объёму⁹, адаптированные к русскому языку и одни из самых общеиспользуемых в практике обработки естественного языка разновидностей моделей на базе трансформера. Другие возможные трансформеры – в планах дальнейших исследований.

Модель RuBERT-Tiny¹⁰ представляет собой довольно компактную версию модели bert-base-multilingual-case, адаптированную для русского и английского языков (45 МБ, 12М параметров). Эта модель обычно удобна, когда речь идёт об относительно простой задаче вроде классификации настроений, и когда исследователя больше заботят скорость и размер, чем точность. Данная разновидность модели примерно в 10 раз меньше и быстрее, чем BERT базового размера. Она обучена на корпусах текстов из Yandex Translate, OPUS-100 и Tatoeba.

Модель RuBERT-Tiny2-Sentence-Compression¹¹ – это улучшенная версия cointegrated/rubert-tiny. Отличия от предыдущей версии заключаются в следующем:

- увеличенный словарный запас, содержащий 83828 токенов вместо 29564;
- более крупные поддерживаемые последовательности слов (supported sequences): 2048 вместо 512;
- значимые вложения сегментов (настроенные на задачу автоматического определения логической связи между текстами (Natural Language Inference, NLI))

⁹ На что указывает приставка «Tiny»

¹⁰ Источник: [cointegrated/rubert-tiny · Hugging Face](#)

¹¹ Источник: [cointegrated/rubert-tiny2 · Hugging Face](#)

- модель ориентирована только на русский язык.

Модель разработана для создания вложений предложений (например, для классификации коротких текстов) или для точной настройки для последующей задачи.

Наконец, модель RuBERT-Tiny-Toxicity¹² представляет собой модификацию исходной модели RuBERT-Tiny, адаптированную для классификации токсичности и пристойности коротких неформальных текстов на русском языке, таких как комментарии в социальных сетях. Обычно задача формулируется как многоуровневая классификация текстов, в ходе которой текст проверяется на токсичность (на то, содержит ли он оскорбления, нецензурную брань и угрозы) и пристойность (на то, может ли он нанести ущерб репутации говорящего). Если текст не является одновременно токсичным и непристойным, он признаётся безопасным.

План сравнения моделей

Мы предполагаем проводить сравнение вышеуказанных модификаций модели RuBERT на основе формальных метрик качества – точности (Accuracy), полноты (Recall) и их среднего гармонического (F1-Score).

Примерный pipeline каждой из трёх моделей выглядит следующим образом (рис. 5):



Рисунок 5 – Pipeline работы с модификациями моделей RuBERT

Источник: построено автором

На вход (рис. 5) мы предполагаем подать данные комментариев из исходного датасета. Далее на этих данных при помощи исходного размеченного датасета мы предполагаем обучить каждую из моделей. Модели строятся на базе трансформера, который автоматически подбирает подходящую архитектуру модели. В нашей работе мы опираемся на следующий набор заданных констант: `first_n_words = 200`, `learning rate = 10-5`, `batch size = 200`, количество эпох - 10. Мы посмотрели, что длина каждого комментария составляет в среднем не более 200 слов (98%), единичные тексты, которые были больше по объёму, были обрезаны.

На выходе для каждой из модели планируется получение вероятностного распределения, которое каждому комментарию будет ставить в соответствие вероятностную оценку того, насколько по мнению сети данный комментарий может быть классифицирован как положительный или как отрицательный (пороговое значение для позитивной тональности – не менее 50%).

¹² Источник: [cointegrated/rubert-tiny-toxicity](https://github.com/cointegrated/rubert-tiny-toxicity) · Hugging Face

Результаты обучения моделей

Вышеуказанные процедуры мы применили по отношению к каждой из трёх моделей.

Ниже приведём сводные графики результатов обучения каждой из трёх моделей для 10 эпох (рис. 6-8). Результаты обучения моделей в разрезе каждой эпохи вынесены в приложение (табл. A1-A3).

Модель RuBert-tiny (рис. 6):

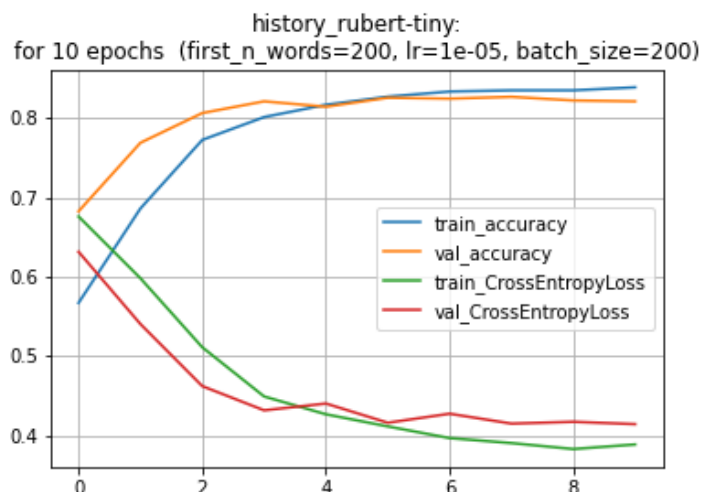


Рисунок 6 – Результаты качества обучения модели RuBERT-Tiny

На графиках (рис. 6) мы видим, что по истечению 10-ой эпохи точность обучения модели на тренировочном и валидационном массивах немногим превысила 80%. Такой результат, безусловно, ещё требует улучшения на более широком датасете, однако для текущего этапа развития работы с имеющимися в нашем распоряжении 4 с лишним тысячами разметок значения выглядят приемлемо. Значения функции потерь (CrossEntropyLoss) в пределах 40% не выглядят угрожающими, тем не менее, для целей нашей работы выглядят приемлемо.

Модель RuBert-Tiny2-Sentence-Compression (рис. 7):

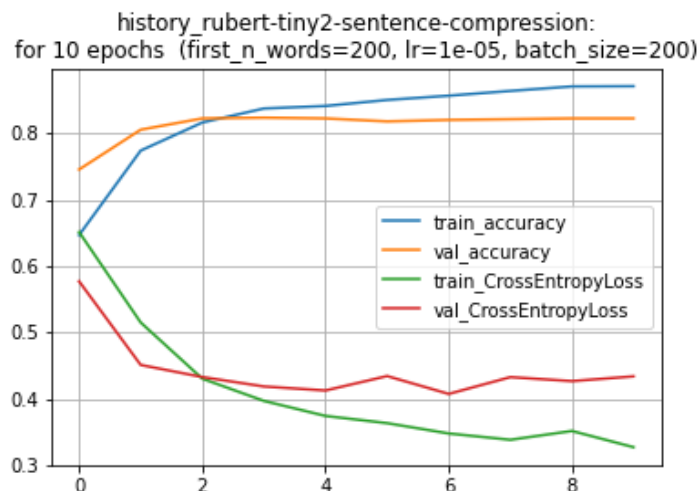


Рисунок 7 – Результаты качества обучения модели RuBERT-Tiny2-Sentence-Compression

На графиках с результатами обучения модели RuBERT-Tiny2-Sentence-Compression (рис. 7) заметно, что точность обучения модели на тренировочном датасете чуточку повыше, чем в предыдущей версии модели RuBERT (точная цифра – 87%, см. табл. A2). Потери модели на тренировочном датасете также чуть ниже (32%). Это говорит в пользу того, что RuBERT-Tiny2-Sentence-Compression – более вероятный кандидат для последующей работы с более объёмным датасетом высказываний в области анализа репродуктивного поведения.

Модель RuBert-Tiny2-Toxicity (рис. 8):

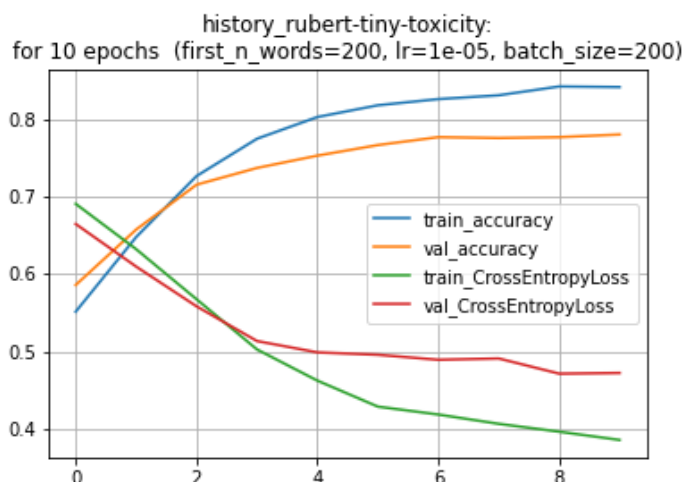


Рисунок 8 – Результаты качества обучения модели RuBERT-Tiny-Toxicity

Показатели обучения модели RuBERT-Tiny-Toxicity близки к качеству обучения исходной версии RuBERT-Tiny.

Результаты сравнения моделей на тестовой выборке

Для сравнения результатов работы каждой из разновидностей модели RuBERT на наших данных тестовой выборки мы свели в таблицу значения оценок их качества (табл. 1):

Таблица 1 – Сравнение метрик качества разновидностей моделей RuBERT

Название модели	Precision	Arr_recall	F1-score
RuBERT-tiny-toxicity	0.757929	0.757803	0.757842
RuBERT-tiny2-sentence-compression	0.832908	0.832930	0.832918
RuBERT-tiny	0.802353	0.802078	0.802149

Мы видим, что наилучшим качеством обладает модель «RuBERT-tiny2-sentence-compression»: значения всех трёх показателей у неё оказались максимальными из всех трёх моделей. Это может говорить в пользу того, что из данной линейки моделей для семантического анализа текста на основе трансформеров данная модификация RuBERT наилучшим образом будет подходить нам для дальнейшей работы на более высоких объёмах данных. Модель на основе «RuBERT-tiny2-sentence-compression» корректно предсказывает класс чуть более, чем 83% из нашего датасета. С учётом того, что исходное количество наблюдений исчисляется четырьмя тысячами, такое значение кажется нам на

данном этапе исследования более чем приемлемым. При расширении датасета хотя бы до 50 тысяч разметок мы надеемся, что появится возможность откалибровать модель таким образом, чтобы увеличить показатель точности до 95%.

Перспективы дальнейшего развития темы

В дальнейшем нам хотелось бы при помощи отработанной технологии выполнить задачу возрастной дезагрегации, осуществив разбивку исходного датасета по топикам, чтобы более точно исследовать, есть ли возрастные сдвиги в отношении людей к теме деторождений, и если да, то по каким сюжетам они могут происходить. К сожалению, наш исходный датасет недостаточно объёмен (на одну из 6 тем приходится около 800 разметок), поэтому на вход в нашей модели подаётся только исходный комментарий без информации о том, к какому сюжету он принадлежит (связан ли он с темой абортов, многодетности или декретных отпусков).

Ведётся работа по расширению выборки до хотя бы 50000 размеченных комментариев, чтобы в дальнейшем иметь возможность подключать сюжеты в анализ. Один из возможных вариантов включения сюжетов в анализ – дезагрегация выборки по сюжетам и калибровка модели внутри каждой подвыборки отдельно. Другой вариант – построение двух сетей: первая будет предсказывать вероятность отнесения комментария к тому или иному классу-сюжету, а вторая на следующем этапе будет оценивать тональность текста. Тем не менее, более детальное рассмотрение сюжетов деторождений на данном этапе остаётся для будущих исследований. Также мы в дальнейшем планируем использовать в предобработке стеммизацию, лемматизацию, очистку от стоп-слов и кодировку «эмодзи», так как она несёт важную смысловую информацию.

Заключение

В работе рассмотрены модели в применении к задаче семантического анализа в области анализа репродуктивного поведения россиян (на данных социальной сети ВКонтакте). Показано, что наилучшие показатели качества (точность, полнота, их среднее гармоническое) достигаются при использовании разновидности модели «RuBERT-Tiny2». Полученный результат может выступить важной ступенькой на пути к дальнейшей обработке более высокого объёма текстов в области репродуктивного поведения людей в России.

Мы обсудили возможные журналы для отправки рукописи с рабочим коллективом. Наиболее вероятные варианты:

- Программные системы: теория и приложения,

- Прикладная информатика

Они не входят в базу Scopus, находятся только в базе RSCI, однако в этих журналах мы уже публиковались ранее. В настоящий момент это лучшие из возможных вариантов.

Источники

- Abosedra S. et al. Dynamics and asymmetries between consumer sentiment and consumption in pre-and during-COVID-19 time: Evidence from the US. *The Journal of Economic Asymmetries*. 2021. Volume 24, November 2021, e00227.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Glandt K. et al. Stance Detection in COVID-19 Tweets. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2021. P. 1596–1611.
- Hasan K.S., V. Ng. Stance Classification of Ideological Debates: Data , Models , Features , and Constraints. *Proc. Sixth Int. Jt. Conf. Nat. Lang. Process.* 2013. October. Pp. 1348–1356.
- Irazú D. et al. Happy Parents' Tweet ? An exploration of 3 Milion Italian Twitter Data. *iussp.confex.com*. 2014. P. 4–7.
- Kalabikhina I.E., E.P. Banin. Database “Childfree (antinatalist) communities in the social network VKontakte”. *Population and Economics*. 2021. 5(2): 92–96. URL: <https://doi.org/10.5281/zenodo.4612131>
- Kalabikhina I.E., E.P. Banin. Database “Pro-family (pronatalist) communities in the social network VKontakte”. *Population and Economics*. 2020. 4(3): 98–130. URL: <https://doi.org/10.3897/popecon.4.e60915>
- Kalabikhina IE, Banin EP, Abduselimova IA, Klimenko GA, Kolotusha AV. The Measurement of Demographic Temperature Using the Sentiment Analysis of Data from the Social Network VKontakte. *Mathematics*. 2021; 9(9):987.
- LaRoche K.J. et al. Attitudes of US adults toward using telemedicine to prescribe medication abortion during COVID-19: A mixed methods study. *Contraception*. 2021. Vol. 104, № 1. P. 104–110.
- Liu S., J. Liu. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine*. 2021. Vol. 39, № 39. Pp. 5499–5505.
- Mandel B. et al. A demographic analysis of online sentiment during Hurricane Irene. *Proceedings of the 2012 Workshop on Language in Social Media*. 2012. Pp. 27–36.
- Miao L., M. Last, M. Litvak. Twitter Data Augmentation for Monitoring Public Opinion on COVID-19 Intervention Measures. 2020.
- Ntontis E., N. Hopkins. Framing a “social problem”: Emotion in anti-abortion activists’ depiction of the abortion debate. *Br. J. Soc. Psychol. John Wiley and Sons Ltd.*, 2018. Vol. 57, № 3. P. 666–683.
- Roldán-Robles P.R. et al. A conceptual architecture for content analysis about abortion using the twitter platform. *RISTI - Rev. Iber. Sist. e Tecnol. Inf.* 2019. Vol. 2019, № E22. Pp. 363– 374.

- Sharma E. et al. Analyzing ideological discourse on social media: A case study of the abortion debate. ACM International Conference Proceeding Series. Association for Computing Machinery, 2017.
- Thorpe D. Huerta et al. Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: A Twitter analysis. SSM- Popul. Heal. 2021. Vol. 15.
- Van de Kaa, D. J. (1987). Europe's second demographic transition. *Population bulletin*, 42(1), 1-59.
- Vychezhzhanin S. V., E. V. Kotelnikov. Stance Detection Based on Ensembles of Classifiers. Program. Comput. Softw. Pleiades Publishing, 2019. Vol. 45, № 5. P. 228–240.
- Калабихина, И. Е., Лукашевич, Н. В., Банин, Е. П., Алибаева, К. В., & Ребрей, С. М. (2021). Автоматическое извлечение мнений пользователей социальных сетей по вопросам репродуктивного поведения. *Программные системы: теория и приложения*, 12(4), 33-63.

Приложение

Таблицы с метриками обучения моделей в разрезе 10 эпох

Таблица A1 – Модель RuBERT-Tiny

Номер эпохи	train_accuracy	val_accuracy	train_CrossEntropyLoss	val_CrossEntropyLoss
1	0.567322239	0.682539683	0.675983612	0.631615543
2	0.686081694	0.768707483	0.598305685	0.540704298
3	0.772314675	0.806122449	0.511512054	0.462607884
4	0.801059002	0.820861678	0.449596635	0.432160527
5	0.816565809	0.814058957	0.427169987	0.440707475
6	0.82677761	0.825396825	0.411917572	0.416522413
7	0.833207262	0.824263039	0.39720127	0.427893662
8	0.834720121	0.826530612	0.390946005	0.415456396
9	0.834720121	0.821995465	0.383463453	0.417740935
10	0.838502269	0.820861678	0.389222024	0.41478591

Таблица A2 – Модель RuBert-Tiny2-Sentence-Compression

Номер эпохи	train_accuracy	val_accuracy	train_CrossEntropyLoss	val_CrossEntropyLoss
1	0.647126	0.746032	0.651673	0.577323
2	0.774584	0.806122	0.51519	0.451132
3	0.816944	0.823129	0.430446	0.432863
4	0.838124	0.824263	0.396862	0.418642
5	0.841906	0.823129	0.374177	0.412613
6	0.850983	0.818594	0.363304	0.434348
7	0.857413	0.820862	0.347718	0.407411
8	0.864599	0.821995	0.338186	0.432757
9	0.871407	0.823129	0.351511	0.426759
10	0.871785	0.823129	0.327023	0.43386

Таблица A3 – Модель RuBERT-Tiny-Toxicity

Номер эпохи	train_accuracy	val_accuracy	train_CrossEntropyLoss	val_CrossEntropyLoss
1	0.551815	0.586168	0.690757	0.664807
2	0.647882	0.657596	0.632381	0.609947
3	0.726551	0.71542	0.567858	0.55873
4	0.774584	0.736961	0.503218	0.51402
5	0.802572	0.752834	0.46301	0.499431
6	0.8177	0.76644	0.429553	0.496322
7	0.825643	0.776644	0.419388	0.489901
8	0.83056	0.77551	0.407409	0.491573
9	0.841906	0.776644	0.397272	0.472023
10	0.84115	0.780045	0.386645	0.4729

