

Эмоциональный анализ постов ВКонтакте: классификатор или регрессор?

Колмогорова А. В.
Сибирский федеральный
университет /Красноярск, Россия
nastiakol@mail.ru

Калинин А. А.
Сибирский федеральный
университет /Красноярск, Россия
verbalab@yandex.ru

Аннотация

В статье обсуждаются результаты решения двух задач машинного обучения: задачи классификации текстов социальных сетей на русском языке по критерию доминирующей эмоции и задачи регрессии, в рамках которой эмоции в тех же текстах социальных сетей предсказываются. В основе экспериментов – сформированный авторами датасет из 3879 текстов из пабликов ВКонтакте, размеченный 2000 ассессорами на краудсорсинговой платформе Толока. Аннотирование проводилось с использованием разработанного интерфейса для недискретной эмоциональной разметки текстов.

Ключевые слова: эмоциональный анализ текстов; классификатор; регрессор; тексты социальных сетей
DOI: 10.28995/2075-7182-2022-21-311-322

Emotional Analysis of posts in VKontakte: classification or regression?

Kolmogorova A. V.
Siberian Federal University /
Krasnoyarsk, Russia
nastiakol@mail.ru

Kalinin A. A.
Siberian Federal University /
Krasnoyarsk, Russia
verbalab@yandex.ru

Abstract

The article summarizes the results of two tasks in machine learning paradigm: the task of classification according to the criterion of dominating emotion on the data of social networks posts in Russian and the regression task using the same data. The experiments are conducted on the data set collected from VKontakte social network and consisted of 3879 posts assessed by 2000 informants on Toloka crowd sourcing platform. The annotation procedure was based on the original interface for non-discrete emotion assessment elaborated by researchers.

Keywords: emotional text analysis; classifier; regressor; texts of social networks

1 Введение

Эмоциональный анализ текстов – интенсивно развивающееся направление, продолжающее традиции уже сложившейся методологии сентимент-анализа [1; 6]. Если для последнего характерно внимание к двум или трем тональностям (позитивная/негативная/нейтральная), то эмоциональный анализ сфокусирован на детектировании конкретных эмоций: страха, гнева, радости и т.д.

Количество выделяемых классов и характер используемых языков в подобного рода проектах варьирует: например, 4 – для испанского [8], 6 – для русского и английского языков [16; 2]. В нашей работе мы использовали в качестве основы восьмичастную модель эмоций шведского нейрофизиолога Г. Лёвхема [11] – так называемый Куб Лёвхейма.

Для большинства проектов в данной области своеобразным камнем преткновения является организация процедуры разметки (ассессмента) текстов на предмет содержащейся в них эмоции. Отобрав посты из трех эмоционально насыщенных пабликов ВКонтакте в качестве материала, мы разработали авторский интерфейс для недискретной оценки эмоций в тексте, т. е. разметки,

предусматривающей не приписывание тексту эмоционального лейбла, а помещение текста в определенную точку эмоционального континуума. Размеченная таким образом коллекция из почти 4000 текстов стала основой для дальнейшего обучения моделей.

В фокусе данной публикации – 2 эксперимента. Один связан с решением задачи классификации текстов по критерию ведущей эмоции, другой – с решением задачи построения регрессионной модели, предсказывающей многомерные эмоциональные характеристики текста.

2 Задачи классификации и регрессии

Регрессия – контролируемый алгоритм машинного обучения, используемый для прогнозирования непрерывных значений выходных данных на основе входных данных. Задача регрессии требует предсказания количества. Иначе говоря, прогнозирующее регрессионное моделирование – это задача приближения функции отображения (f) от входных переменных (X) к непрерывной выходной переменной (y). Регрессия может иметь действительные или дискретные входные переменные. Применительно к решению задачи эмоционального анализа текста, регрессионные модели предсказывают «количество» той или иной эмоции в тексте, подаваемом на вход для анализа – например, в нем представлены в таких-то количествах Грусть, Радость, Удивление и Отвращение.

Классификация представляет собой контролируемый алгоритм машинного обучения, который предсказывает определенные дискретные значения (категории или классы), к которым принадлежат входные данные. При решении задачи классификации от модели МО требуется заключение о принадлежности документа к определенному классу на основе ранее изученных примеров каждого из возможных классов. Иными словами, классификационное прогнозирующее моделирование – это задача приближения функции отображения (f) от входных переменных (X) к дискретным выходным переменным (y). В контексте нашего исследования это означает, что модель МО предсказывает эмоциональный класс, к которому принадлежит поданный на вход для анализа текст – он, например, грустный.

Хотя модели, основанные на классификации, широко распространены в практике эмоционального анализа, за последние пять лет количество работ, использующих регрессионные модели для обработки текстовых данных, содержащих эмоции, также постоянно растет. Это связано с тем, что понимание эмоции как дискретной системы накладывает определенные ограничения на точность и нюансированность анализа: естественное переживание некоторого эмоционального состояния редко бывает окрашено только одной эмоцией – даже если ведущим аффектом является, например, Радость, к ней могут примешиваться, хотя и с меньшей выраженностью, Воодушевление или Удивление, а также нотки Грусти. Переход от дискретной парадигмы, когда эмоция, вербализованная в тексте, рассматривается как класс, к использованию непрерывных ее измерений открывает более широкие перспективы для получения результатов максимально приближенных к естественным процессам интерпретации эмоции в тексте читающим его человеком.

Отметим, что в пока в исследовательской практике наблюдаются противоречивые результаты и полярные мнения относительно преимуществ регрессионных моделей над классификационными в контексте анализа эмоций в текстовых данных.

Например, в [12], взяв в качестве исходных данных короткие тексты о погоде в Твиттере, исследователи сосредоточились на проверке гипотезы о том, что регрессионный анализ в сочетании с оценкой достоверности классов работает лучше, чем традиционные методы классификации, такие как SVM и K-Nearest Neighbor, основанные на дискретных целевых переменных. Разработчики пришли к выводу, что в целом регрессионный подход не показал явных преимуществ: по видимому, он более оправдывает себя в задачах, требующих более детализированного анализа, предполагающего вычисление множества разнородных признаков.

В исследовании [4] регрессионный подход был применен для анализа текстовых данных из различных открытых датасетов, размеченных на основе многомерной модели эмоций, известной как VAD-модель [13]. Модель построена на количественном измерении трех параметров: уровня возбуждения, уровня доминирования и значений полярности эмоции (негативная / позитивная). Комбинация значений данных параметров, в конечном счете, определяет степень интенсивности шести базовых эмоций. В итоге исследователями было достигнуто приемлемое качество предсказания, а сама методология оценена ими как перспективная.

Очевидно, что не последнюю роль в достижении показателей эффективности модели играют данные, на которых она будет работать: обучаться и тестироваться. Мы предположили, что наш датасет, который был получен путем применения процедуры недискретной разметки текстов из ВКонтакте на основе многомерной модели эмоций Куб Лёвхейма, хорошо совместим с регрессионными моделями.

3 Специфика интерфейса для недискретной оценки эмоции в тексте

В эмоциональном анализе текстов принято два основных подхода к разметке: дискретный (категориальный) и многомерный [5]. Как мы указывали выше, первый хорошо сочетается с задачами по классификации, основу которой, как правило, составляет модель шести базовых эмоций П. Экмана [7]. Второй используется реже, поскольку моделей, представляющих эмоции как континуум, в рамках некоторого пространства, заданного тремя параметрами, немного. Куб Лёвхейма – одна из таких моделей.

Шведский нейрофизиолог Гуго Лёвхейм в [11] предложил оригинальную концепцию эмоций, которую визуализировал, подобно Дж. Расселу и А. Мерабиану, авторам VAD-модели, в виде куба. Суть идеи Лёвхейма сводится к тому, что характер эмоции, субъективно испытываемой человеком, определяется уровнем трех нейротрансмиттеров в амигдале головного мозга. Это норадреналин (вертикальная ось Куба Лёвхейма), серотонин (горизонтальная ось) и дофамин (глубина Куба Лёвхейма) (рис.1). В восьми вершинах Куба расположены восемь базовых эмоций. Каждая из них имеет двучленную номинацию, где первое слово обозначает самую слабую степень интенсивности эмоции, а второе – самую сильную. В дальнейшем нам пришлось упростить номинации, оставив только по одному слову. Для того, чтобы найти адекватные эквивалентные номинации на русском языке, мы предварительно провели психолингвистический эксперимент, в котором 30 информантам, носителям русского языка, было предложено соотнести эмоциональные тексты из ВКонтакте (например, под хештегами #Подслушано_успех или #Подслушано_фууу и др.) с названиями эмоций из списка, в котором фигурировали разные возможные варианты перевода английских слов, в оригинале использованных Г. Лёвхеймом. Так например, для лексемы *excitement* нами были предложены варианты: Возбуждение, Воодушевление, Энтузиазм, Оживление. Наибольший рейтинг по итогам эксперимента с текстами получила номинация Воодушевление. Для лексемы *disgust* нами были предложены варианты: Отвращение, Омерзение, Брезгливость, Раздражение. Первое место в «рейтинге» информантов получила номинация Отвращение.

Итак, по вертикальной оси, указывающей на уровень норадреналина, такие эмоции как Стыд, Страх, Радость и Отвращение располагаются внизу (что указывает на низкий уровень норадреналина), а Грусть, Гнев, Воодушевление и Удивление – наверху (указывая на высокий уровень этого нейротрансмиттера).

Следующая ось – горизонтальная, она соответствует уровню серотонина. Стыд, Страх, Грусть, Гнев находятся на «левой стороне» (низкий уровень). Отвращение, Радость, Удивление и Воодушевление находятся на «правой» стороне, что соответствует высокому уровню данного нейротрансмиттера.

И, наконец, последняя ось соответствует уровню дофамина. На ней располагаются Грусть, Стыд, Отвращение и Удивление как низко-дофаминовые эмоции, и Страх, Гнев, Воодушевление и Радость как высоко-дофаминовые эмоции.

Мы использовали данную модель в качестве концепт-идеи, поскольку данное пространство внутри Куба является своеобразным эмоциональным континуумом, в котором мы можем локализовать каждый текст, оцененный информантами при аннотировании. Проблема состояла в том, какой интерфейс можно предложить аннотаторам, чтобы они могли не просто приписать тексту определенную метку, но и показать количество той или иной эмоции, присутствующей в нем.

Ввиду трехмерности и недискретности эмоциональной модели Лёвхейма, процедура аннотирования должна учитывать следующую специфику целевых переменных:

- аннотирование должно представлять собой недискретную, в идеале – непрерывную оценку на определенном интервале, а не простую атрибуцию к определенному классу;
- аннотирование должно содержать оценки для всех трех измерений Куба Лёвхейма.

Поскольку эмоция в рамках подхода Лёвхейма представляет собой точку в трехмерном пространстве, то в идеале пользователь должен выбрать эту самую точку в пределах Куба (рис. 1). Но здесь мы сталкиваемся с проблемой удобства интерфейса. Для такого рода задачи требуется интерфейс, позволяющий использовать 3D-пространство со множеством сопутствующих операций: вращение, приближение, движение в двух плоскостях. Такой инструментарий крайне дорог в разработке, но, что самое важное – он был бы крайне неудобен в использовании, т. к. прежде чем поставить эту самую точку, соответствующую эмоциональному положению текста, ассессору необходимо было бы совершить немало предварительных шагов, предварительно ознакомившись с информацией о нейротрансмиттерах. Это непростая задача.

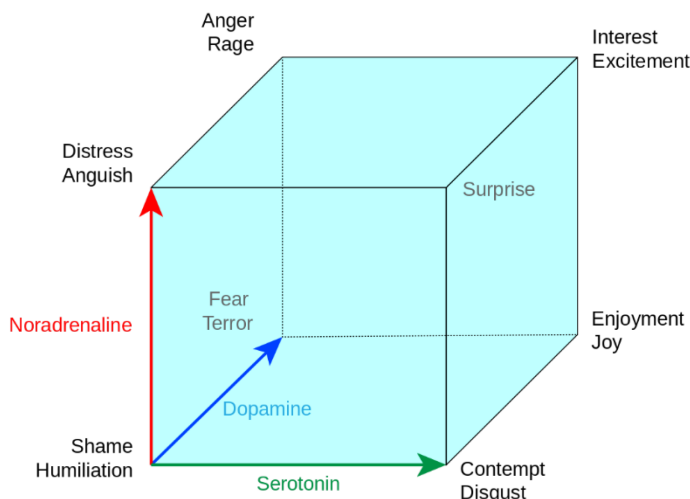


Рис. 1: Визуализация модели Куба Лёвхейма [Lövheim, 2012]

Приняв вышеперечисленные ограничения, мы пришли к выводу, что процедуру разметки необходимо упрощать. Немного потеряв в точности, мы все же сможем получить большое количество данных за счет привлечения ассессоров, которым можно будет поставить задачу, выполнимую с минимальным уровнем подготовки и с использованием знакомых и понятных UI-примитивов в интерфейсе разметки. Для реализации такого подхода мы обратили внимание на одно интересное свойство кубов, а именно – главные диагонали, соединяющие противоположные вершины Куба.

Эти вершины Куба (соответствующие восьми крайним эмоциям) образуют четыре оппозиции, которые накладываются на диагонали внутри него (см. рис. 2). Оппозиции основаны на принципе «качелей нейротрансмиттеров» (табл.1): у эмоций, образующих вершины диагоналей, противоположные значения уровня нейротрансмиттеров. Если, например, у Отвращения максимальный уровень серотонина, и минимальные – норадреналина (гормона стресса, вызывающего активацию всех физиологических проявлений возбуждения [14]) и дофамина (гормона, вызывающего прилив мотивации, обострение восприятия времени и пространства [Ibid]), то у Гнева – все с точностью наоборот, поэтому в определенном смысле он является агрессивным антонимом Отвращения.

оппозиция эмоций	серотонин	норадреналин	дофамин
Грусть	–	+	–
Радость	+	–	+
Гнев	–	+	+
Отвращение	+	–	–
Стыд	–	–	–
Воодушевление	+	+	+
Страх	–	–	+
Удивление	+	+	–

Таблица 1: Сочетания максимального / минимального уровня нейротрансмиттеров у эмоций, образующих вершины диагоналей Куба Лёвхейма

Подобная дихотомия, опирающаяся на диагонали, позволяет нам условно представить «спаренные диагональю» эмоции как антиподы, а поскольку понятие «противоположность», или «антоним», более близко специально неподготовленному информанту (ассессору), чем, например, понятия «моноамин» или «нейротрансмиттер», то мы можем заменить задачу указания конкретной моноаминовой конфигурации на задачу поиска баланса между четырьмя противоположностями, существенно упростив таким образом задачу информантам.

Указание значения баланса на уровне интерфейса, предназначенного для ассессора, может быть реализовано через ползунок (slider-bar). Перемещая ползунок по шкале, ассессор может гибко указывать на «степень выраженности» двух противоположных эмоций.

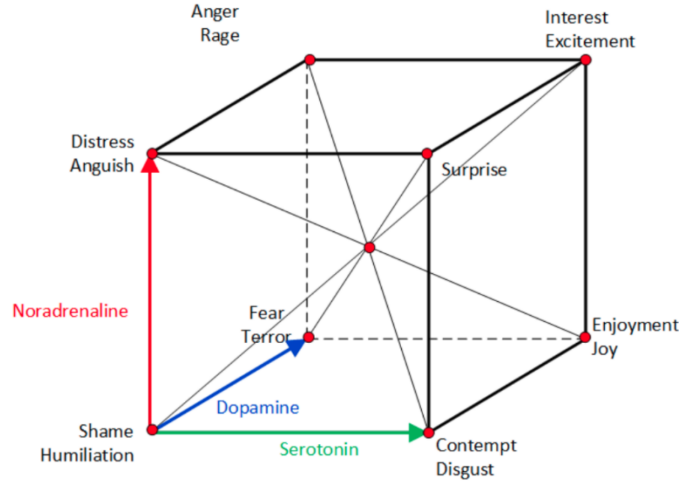


Рис. 2: Диагонали внутри Куба Лёвхейма

Так, например, с использованием такого инструмента можно разметить текст как «умеренно гневный», «немного стыдный», «не содержащий ни грусти, ни радости» и «слишком страшный», используя все четыре шкалы.

Дизайн интерфейса разметки представлен на рисунке 3. Крайние точки каждой шкалы обозначают максимально выраженную эмоцию. Точка посередине – это нейтральное значение. Чем меньше, по мнению ассессора, выражена эмоция, тем ближе к середине он перемещает слайдер. Перед оценкой текста, на экране ассессору предлагалась следующая инструкция: *Внимательно прочитайте текст. Какие эмоции автор выражает в тексте? На каждой из шкал поставьте отметку ближе к той эмоции, которая сильнее выражена в тексте. Поставьте отметку настолько близко, насколько очевидна и сильна эта эмоция в тексте. Например, 1 шаг от центра – если оттенок эмоции присутствует, но выражен слабо; 3 шага – если эмоция явно присутствует; 5 шагов – если эмоция, без сомнения, доминирует. Если в тексте нет эмоций, обозначенных на шкале, оставьте отметку в среднем положении.*

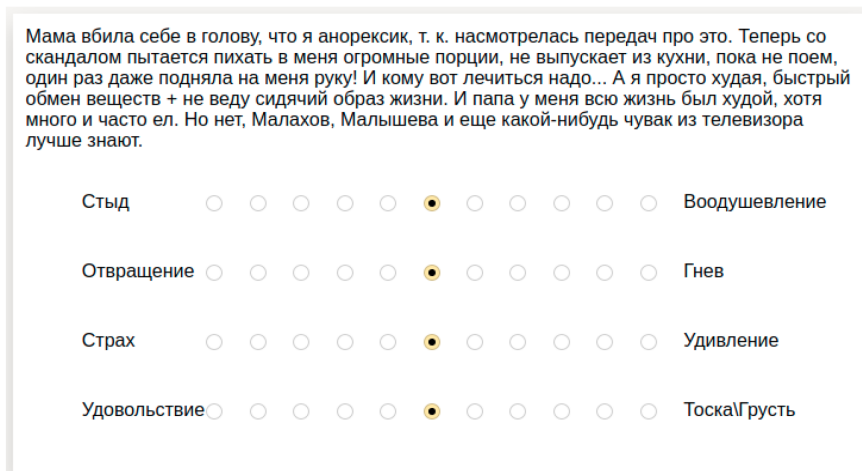


Рис. 3: Пример экрана с недискретной разметкой с опорой на диагонали Куба Лёвхейма

Основными преимуществами вышеупомянутого инструмента оценки являются:

- 1) ассессор может указать не только на первичную эмоцию, но и на вторичную, если, например, образец текста в основном касается чувства страха, но аннотатор хочет упомянуть легкое воздействие гнева;
- 2) в течение разметки ассессоры могут изменять степень выражения эмоции, перемещая ползунок относительно средней точки;
- 3) по результатам разметки мы имеем возможность рассматривать каждую оппозицию как период $(-1,1)$ и сопоставлять положение ползунка с вектором, угол которого совпадает с соответствующей диагональю, и величину этих векторов, которая равна значению, выбранному для данной эмоциональной оппозиции.

После снятия данных мы можем аппроксимировать результирующую точку в виде агрегации векторов. Разумеется, мы теряем возможность более точного указания точки, поскольку указание координат идет только на отрезках диагоналей, и финальное агрегированное значение получается приблизительным, аппроксимированным, но одновременно мы делаем интерфейс удобным, а целевые метрики – максимально приближенными, и можем соблюсти баланс точности данных и размера выборки.

С использованием данного интерфейса было размечено 3789 текстов из пабликов «Подслушано», «Палата №6» и «Карамель» ВКонтакте. В разметке приняли участие 2000 ассессоров, зарегистрированных на платформе Толока. Минимальным условием было двойное покрытие выборки, т.е. минимально каждый текст из выборки оценивался двумя информантами, но для большинства текстов этот критерий перевыполнялся.

В качестве метрики для оценки согласованности ответов аннотаторов мы использовали коэффициент внутриклассовой корреляции, предложенный в [3]. Согласно метрике, коэффициент 0 означает, что между значениями в классе нет никакой корреляции, 1 – что они идентичны. В нашей выборке коэффициент межклассовой корреляции составил 0,6325, что показывает, что степень согласованности аннотаторов в целом удовлетворительная. Явные аутлайнеры, составившие в общей сложности не более 2 % от выборки, исключались, поскольку как было показано в одном из наших экспериментов [10], маргинальные оценки чаще всего давались ассессорами с высоким уровнем индивидуальной эмпатии.

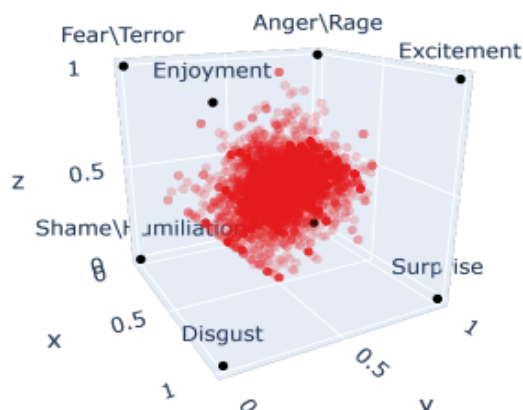


Рис. 4: Визуализация выборки оцененных текстов в трехмерном пространстве Куба

Полученная нами выборка визуализирована на рис. 4. Каждый единожды оцененный текст – это точка в пространстве Куба.

4 Решение задачи классификации

При разработке системы автоматической классификации эмоций нами были проведены эксперименты, комбинирующие ранее описанные модели векторного представления текста и алгоритмы машинного обучения. Первая задача, которая была поставлена, сводилась к апробации моделей дискретной классификации, поскольку данные модели имеют широкую практику применения.

Для решения этой задачи четыре вектора, которые получал при аннотировании каждый текст нашей выборки, агрегировались в один усредняющий вектор, а ближайшая к нему вершина принималась за метку эмоционального класса текстов: если ближайшей вершиной оказывалась Радость, то текст относился к соответствующей категории «радостных» текстов.

В каждом из экспериментов использовались идентичные обучающая и тестовая выборки. Количество фрагментов на каждый из восьми классов эмоций в обучающей выборке было следующим: Стыд – 757, Грусть – 551, Воодушевление – 397, Радость – 390, Гнев – 336, Отвращение – 307, Удивление – 248, Страх – 159.

Текстовые данные в тестовой выборке разбиты подобным образом (соотношение обучающей и тестовой – 80% / 20%): текстов, размеченных как Стыд – 189, Грусть – 138, Воодушевление – 100, Радость – 98, Гнев – 83, Отвращение – 76, Удивление – 62, Страх – 40.

Векторизация текстов была выполнена с помощью трех наиболее популярных моделей векторного представления:

- Bag-of-words + TF-IDF (Sparse vectors);
- fastText (Common Crawl RU 300);
- ELMo (Deep Pavlov, Russian Wikipedia 1024).

В качестве методов машинного обучения, проводящих непосредственную классификацию, нами использовались:

1. GaussianNB
2. KNeighbors Classifier
3. LogisticRegression
4. Linear Discriminant Analysis
5. XGBClassifier (xgboost)
6. MultiLayerPerceptron — 5 Layers + SGD + Cross-Entropy Loss
7. bidirectional LSTM
8. Conv1D

Для работы с векторами fastText и ELMo в случаях, где алгоритм ожидает одного вектора, а не цепочки (1-6), вектора токенов, входящих в тексты, были усреднены, чтобы классификатор мог работать с одним вектором, представляющим весь текст. Для «нейронных» классификаторов (6-8) применялся SGD оптимизатор и перекрестная энтропия в качестве лосс-функции.

Для оценки качества работы классификатора мы использовали комбинацию двух наиболее важных характеристик классификатора – точности и полноты.

Такая оценка позволяет учитывать количество истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных атрибуций по отношению к полноте выполнения задачи и дает сбалансированную характеристику модели. В дальнейшем под точностью/эффективностью работы классификатора мы будем понимать взвешенное среднее арифметическое значений метрики F1, рассчитанное для каждой из эмоций. Результаты значения метрики F1 при комбинации рассмотренных ранее алгоритмов и моделей векторизации представлены в Таблице 2. Алгоритм линейного дискриминантного анализа с признаками репрезентаций TF-IDF обучить не удалось ввиду недостаточно высокой вычислительной мощности. При вычислении TF-IDF весов модель обрабатывает обширные массивы данных, устанавливая большое количество линейных зависимостей на мелких списках признаков.

Модели векторизации/ алгоритмы МО	Наивный Байес	к-ближ. соседей	Логист. регрессия	ЛДА	xgboost	многос. перцеп.	LSTM	Conv1D
TF-IDF	43,26	45,40	58,64	–	47,97	54,92	55,92	55,43
fastText	40,43	41,53	44,06	60,80	52,97	52,66	53,65	53,40
ELMo	40,34	44,87	61,62	54,89	53,97	53,03	55,20	55,86

Таблица 2: Результаты работы классификатора по мере F1, %

Среди описанных выше классификаторов наилучшую точность показывает алгоритм логистической регрессии (табл.2). Лучшее качество предсказания этот алгоритм получает при использовании признаков векторного представления ELMo, достигая среднего значения метрики F1 в 61,62 %. Худшее качество классификатор, основывающийся на логистической регрессии, показывает при использовании векторных представлений fastText. Интересно, что fastText на всех остальных алгоритмах дает достаточно высокую точность.

Помимо логистической регрессии, близкое к максимальному значению точности значение показывает алгоритм линейного дискриминантного анализа, работающий в совокупности с векторным представлением fastText – 60,8 %. Для всех остальных алгоритмов и векторных представлений текста максимальная точность классификации по мере F1 варьирует от 55 до 38 %.

Худшее качество при предсказании и обучении демонстрирует наивный байесовский классификатор. Такие результаты можно объяснить тем, что данный алгоритм считает все признаки независимыми друг от друга, в то время как в векторных репрезентациях все признаки, на самом деле, зависимы.

Для более подробного анализа эффективности алгоритма логистической регрессии в совокупности с векторным представлением ELMo рассмотрим матрицу ошибок (рис. 5).

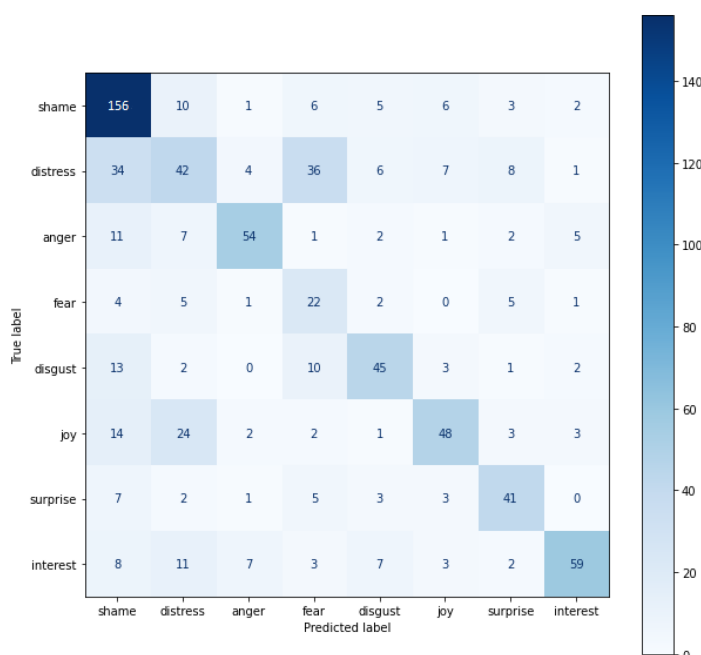


Рис. 5: Матрица ошибок алгоритма логистической регрессии с векторами ELMo

Как видно на изображении (рис. 5), большинство текстов классификатор соотносит корректно, диагональные элементы матрицы явно выражены. Каждая из эмоций в большинстве случаев определяется правильно. Эмоция Грусти иногда ошибочно распознается как Стыд или Страх, а эмоция Радости – как Стыд или как Грусть. Во всех остальных случаях ошибки незначительны.

В таблице 3 представлена числовая оценка эффективности работы классификатора логистической регрессии с векторным представлением ELMo, рассчитанная по метрикам точности, полноты и F1 для каждого из классов эмоций.

Эмоциональный класс	Точность	Полнота	F1
Стыд	82,54	63,16	71,56
Грусть	30,43	40,78	34,85
Гнев	65,06	77,14	70,59
Страх	55	25,88	35,2
Отвращение	59,21	63,38	61,22
Радость	49,48	67,60	57,14
Удивление	66,13	63,08	64,57
Воодушевление	59	80,82	68,21

Таблица 3: Эффективность алгоритма логистической регрессии с векторами ELMo по классам эмоций, %

Табличные данные, как и матрица ошибок, наглядно показывают, что Грусть является наименее распознаваемой эмоцией: к данному классу корректно относится только каждый третий текст. С наибольшей точностью система детектирует Стыд: 8 из 10 предсказаний совпадают с фактической оценкой ассессоров. Также достаточно хорошо определяются эмоции Удивление и Гнев. Отдельного внимания заслуживает класс «Страх», который определяется с относительно приемлемой точностью, но выделяется низкой полнотой. Как мы ранее упоминали, грустные тексты достаточно часто ошибочно относят к категории страшных, что и вызывает соответствующее снижение полноты классификации.

К сожалению, в силу ряда причин, в основном связанных с несовпадением технического стека, в данном эксперименте по классификации были упущены современные архитектуры – «трансформеры», которые можно использовать в качестве входных векторов с последующей тренировкой классификатора. Однако данные модели нашли свое применение во втором эксперименте с регрессией.

5 Решение задачи регрессии

Задача регрессии в контексте анализа эмоций выглядит, на наш взгляд, более интересной так как позволяет предсказывать числовую переменную, а не дискретную, как в случае с классификацией. Такой подход делает анализ более нюансированным и точным. Нами уже были предприняты некоторые эксперименты по применению регрессионных подходов в [9]. Однако в данной публикации мы хотели бы обсудить результаты последующих экспериментов.

При помощи классических алгоритмов (DecisionTreeRegressor, ExtraTreeRegressor, RandomForestRegressor, MLPRegressor – строки 1–4 в табл. 4) мы не получили результатов приемлемого качества, поэтому в данной работе мы решили обратиться к архитектурам типа «трансформер», базовая схема которых изложена в [15].

В настоящее время существует значительное количество уже подготовленных языковых моделей, таких как BERT, RoBERTa, GPT-2, XLNet и других эмбедингов на основе трансформер-архитектуры. Эти модели хорошо себя показывают для разных текстовых задач, но у них есть ряд серьезных ограничений:

- Недостаточные возможности для получения агрегированного вектора текста из векторов представлений отдельных токенов, входящих в текст. Эмбединги на базе трансформеров – это плотные векторы, сопоставленные с токенами из текста, поэтому, чтобы использовать их для описания всего текста, мы должны их агрегировать. Простое усреднение – не лучший подход, поскольку теряется много данных. Токен CLS также не является хорошим кандидатом, поскольку он был обучен для задачи классификации – мы стремимся к регрессии.
- Плохая поддержка русского языка. Несмотря на то, что BERT имеет многоязычную версию и есть специально обученный RuBERT, мы не можем их использовать, поскольку исходный набор данных для многоязычной поддержки представляет собой дампы Википедии. Таким образом, он полностью отличается от того источника данных, с которым мы работаем (личные эмоционально насыщенные тексты из социальных сетей)

- Тяжеловесность. Это главный недостаток. Даже самая младшая модель трансформеров занимает около 100 МБ. При таком «весе» ее использование неэффективно на этапе проверки концепции, когда основная цель – собрать наиболее жизнеспособное и легкое решение, чтобы при необходимости доработать его.

Несмотря на обозначенные выше ограничения, было решено опробовать данные подходы в связи с общей высокой эффективностью их архитектур.

Перед нами стоял ряд вопросов:

- Какую базовую модель выбрать?
- Какие функции для оценки ошибки выбрать?
- Как организовать обучение?

Как уже было отмечено выше, несмотря на распространенность архитектуры, все популярные и доступные модели ориентированы на английский язык. Среди инициатив по тренировке моделей можно отметить проект DeepPavlov, в рамках которого тренируются модели, адаптированные под русский язык. Среди серии таких предобученных моделей была выбрана модель “distilrubert-tiny-cased-conversational”. Она отличается относительно небольшим размером 100М параметров, обучена на смешанном датасете из русских текстов Википедии и русскоязычных новостей, а затем сжата из RuBERT методом дистилляции.

Мы реализовали следующую схему (рис.6):

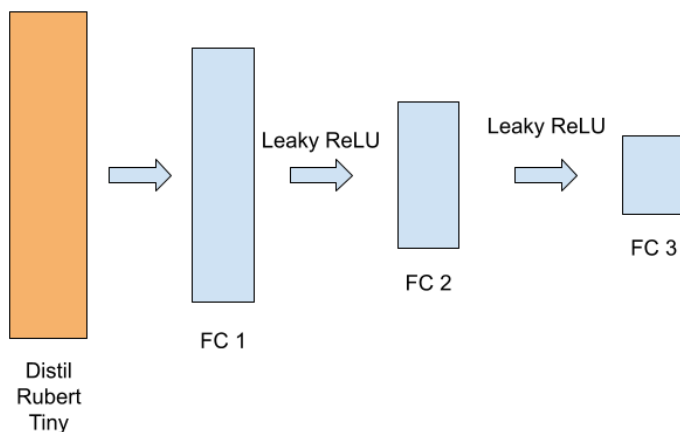


Рис. 6: Схема модели для тренировки регрессора на основе языковой модели

Получив из языковой модели эмбединги для цепочки токенов, мы усредняли их, чтобы получить вектор фиксированной размерности, описывающий весь текст. Усредненный вектор текста проходил через цепочку полносвязных слоев (Fully Connected – FC) и в результате преобразовывался в четырехмерный тензор, соответствующей целевой переменной – значениям диагоналей Куба. В качестве функции активации для перехода использовалась Leaky ReLU. В качестве функции потерь использовалась MSE, в качестве валидационной метрики – MAE.

В данном подходе мы «замораживали» веса языковой модели, и тренировали исключительно блок полносвязных слоев. Такой подход (Frozen BERT) позволяет быстрее тренировать модель за счет меньшего количества тренируемых параметров.

Удовлетворительных результатов добиться не получилось, так при использовании MSE для нескольких целевых переменных, усреднение выходных тензоров поощряло значение 0 для всех целевых переменных. После этого мы решили тренировать также и параметры языковой модели (Fine-tuned BERT). Это заняло существенно больше времени, но и результаты существенно улучшились (см. Табл. 4).

№ п/п	Модель	MSE	MAE
1	DecisionTreeRegressor	10,78	2,40
2	ExtraTreeRegressor	10,32	2,33
3	RandomForestRegressor	6,05	1,82
4	MLPRegressor	6,35	1,95
5	Frozen BERT	6,34	2,57
6	Fine-tuned BERT	3,59	1,36

Таблица 4.: Результаты тренировки на базе регрессионных моделей

Как можно видеть при использовании трансформер-архитектур наблюдается рост метрик сравнительно с использованием классических методов машинного обучения, но данный рост стал возможен только при включении всех параметров языковой модели при дообучении. Исходный код доступен в Colab ноутбуке [https://colab.research.google.com/drive/14PZHxMzAN2SXz_jHzMmW-PW6uu_vxNxj?usp=sharing].

6 Заключение

Проведенные эксперименты в целом демонстрируют, что на данном, размеченном недискретным образом, датасете задача классификации решается с «неровными» результатами – Гнев и Воодушевление показывают достаточно высокие, хотя и недостаточные, значения F1, а Страх и Грусть дают чрезвычайно низкие значения данной метрики.

В то же время регрессионная модель дает неплохие значения метрик MSE и MAE, что можно интерпретировать как признак экологичности процедуры в случае, когда данные недискретной разметки используются для предсказания эмоциональных признаков текста.

Именно регрессионная модель будет использоваться в дальнейших экспериментах и станет основой для разработки приложения для эмоционального анализа русскоязычных текстов социальных сетей

References

- [1] Alexeeva Svetlana. V., Kolcov Sergey N., Koltsova Olesia Yu. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2016” Moscow, June 1–4, 2016. pp. 277–288.
- [2] Alm Cecilia O., Roth Dan, Sproat Richard. Emotions from Text: Machine Learning for Text-based Emotion Prediction // Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Vancouver, 2005. pp. 579–586.
- [3] Bland J. Martin, Altman Douglas G. Measurement Error and Correlation Coefficients // British Medical Journal, 1996, Vol. 313 (7048), pp. 41–24.
- [4] Buechel Sven & Hahn Udo. Emotion Analysis as a Regression Problem-Dimensional Models and Their Implications on Emotion Representation and Metrical Evaluation // ECAI, 2016, pp. 1114–1122.
- [5] Calvo Rafael A. & Mac Kim Sunghwan. Emotions in text: Dimensional and categorical models // Computational Intelligence, 2013, Vol. 29(3), pp. 527–543.
- [6] Chaturvedi Deepshikha, Chopra Shalu. Customers Sentiment on Banks // International Journal of Computer Applications, 2014, Vol. 98(13), pp. 8–13.
- [7] Ekman Paul. An argument for basic emotions // Cognition & Emotion, 1992, Vol. 6(3-4), pp. 169–200.
- [8] Heras-Pedrosa De Las Carlos, Sánchez-Núñez Pablo, Peláez José. I. Sentiment Analysis and Emotion Understanding during the COVID-19 Pandemic in Spain and Its Impact on Digital Ecosystems // International Journal of Environmental Research and Public Health, 2020, Vol. 17(15), pp. 5542, July 2020.

- [9] Kolmogorova Anastasia, Kalinin Alexander, Malikova Alina. Lövheim Cube-Backed Emotion Analysis: From Classification to Regression. In: Digital Transformation and Global Society. DTGS 2021. Communications in Computer and Information Science, 2022, Vol 1503. Springer, Cham. https://doi.org/10.1007/978-3-030-93715-7_7
- [10] Kolmogorova Anastasia, Kalinin Alexander, Malikova Alina. Semiotic Function of Empathy in Text Emotion Assessment // Biosemiotics, 2021, №14 (3). Cham: Springer Nature. DOI: 10.1007/s12304-021-09434-y
- [11] Lövheim Hugo. A New Three-Dimensional Model for Emotions and Monoamine Neurotransmitters // Medical Hypotheses, 2012, Vol. 78, pp. 341–348.
- [12] Önal, I., Ertuğrul, A.M.: Effect of using regression in sentiment analysis. In: 22nd Signal Processing and Communications Applications Conference (SIU), pp. 1822–1825. IEEE, Trabzon (2014). <https://doi.org/10.1109/SIU.2014.6830606>
- [13] Russell James A. & Mehrabian Albert. Evidence for a three-factor theory of emotions // Journal of Research in Personality, 1977, Vol.11(3), pp. 273– 294.
- [14] Talanov Max & Leukhin Alexey & Lövheim Hugo, Viverdú Jordi, Toshev Alexander, Gafarov Fail. Modeling psycho-emotional states via neurosimulation of monoamine neurotransmitters. In: Viverdú, J., Müller, V. (eds.) Blended cognition. Springer Series in Cognitive and Neural Systems, 2019, Vol. 12, pp. 127–157. Springer, Cham (2019). doi: 10.1007/978-3-030-03104-6_6
- [15] Vaswani Ashish & Shazeer Noam & Parmar Niki & Uszkoreit Jakob & Jones Llion & Gomez Aidan & Kaiser Lukasz & Polosukhin Illia. Attention is all you need // Computing Research Repository. — 2017. — Vol. arXiv:1706.03762. —Access mode: <https://doi.org/10.48550/arXiv.1706.03762>.
- [16] Volkova Svetlana, Chetviorkin Ilya, Arendt Dustana, Van Durme Bedjamin. Contrasting public opinion dynamics and emotional response during crisis // Social Informatics, Cham, Switzerland:Springer, 2016. pp. 312–329.