

Научная статья

УДК 316.473+303.686

doi: 10.17223/1998863X/85/14

БИОГРАФИЧЕСКОЕ ИССЛЕДОВАНИЕ СУБЪЕКТИВНОГО БЛАГОПОЛУЧИЯ С ПОМОЩЬЮ МЕТОДОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Константин Сергеевич Дивисенко

*Социологический институт РАН Федерального научно-исследовательского
социологического центра Российской академии наук, Санкт-Петербург, Россия*

k.divisenko@socinst.ru

Аннотация. В статье рассматривается возможность изучения субъективного благополучия на основе анализа качественных данных – биографических текстов старшеклассников. Представлены результаты реконструкции основных компонентов благополучия в соответствии с концепцией К. Рифф. Продемонстрирован потенциал методов обработки естественного языка и машинного обучения для создания модели классификации и кодирования текстовых фрагментов, репрезентирующих основные компоненты субъективного благополучия.

Ключевые слова: субъективное благополучие, биографическое исследование, машинное обучение, старшеклассники, кодирование текстовых данных

Для цитирования: Дивисенко К.С. Биографическое исследование субъективного благополучия с помощью методов обработки естественного языка // Вестник Томского государственного университета. Философия. Социология. Политология. 2025. № 85. С. 164–176. doi: 10.17223/1998863X/85/14

Original article

A BIOGRAPHICAL STUDY OF SUBJECTIVE WELL-BEING USING NATURAL LANGUAGE PROCESSING METHODS

Konstantin S. Divisenko

*Sociological Institute of the RAS – Branch of the Federal Center of Theoretical
and Applied Sociology of the Russian Academy of Sciences,
Saint Petersburg, Russian Federation, k.divisenko@socinst.ru*

Abstract. Text data became more frequent in subjective well-being studies due to development of computer-assisted methods for qualitative research. The article examines the possibility of studying subjective well-being based on autobiographical data by means of natural language processing and machine learning. The six-factor model of well-being developed by Carol Ryff was used for its reconstruction in current study. Open coding of the autobiographical texts corpus written by high school students ($n = 197$) was carried out in accordance with this six-factor model of subjective well-being: self-acceptance, positive relationships with others, autonomy, environmental mastery, purpose in life, personal growth. Fragments describing purpose in life and positive relationships with others are the most frequent in high school students' autobiographical texts. The labeled data were used to build a baseline machine learning model build upon count and TF-IDF vectorisation as well as logistic regression and random decision forests algorithms. Semantic vectorisation of the text with ruBERT (Bidirectional Encoder Representations from Transformers) increased the classification accuracy. The weighted average F_1 value in the case of binary classification for

“personal growth”, “goals in life”, “positive relationships with others” was 0.92, 0.85 and 0.89, respectively. The results of the study are entirely consistent with the previously described changes in the high school students’ lifeworld and indicate the gradual development of a realistic type of a biographical project. It seems important to conduct experiments with an expanded dataset, as well as testing other language models. Probably, the classification accuracy can be increased by adding part of speech tagging. The trained models can be used to analyze similar autobiographical texts and as a screening test of subjective well-being.

Keywords: subjective well-being, biographical research, machine learning, high school students, text data coding

For citation: Divisenko, K.S. (2025) A biographical study of subjective well-being using natural language processing methods. *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science*. 85. pp. 164–176. (In Russian). doi: 10.17223/1998863X/85/14

Введение. Постановка проблемы

Анализу субъективного благополучия посвящено достаточно большое число социологических, психологических, междисциплинарных исследований. С одной стороны, интерес исследователей объясняется тем, что субъективное благополучие как компонент качества жизни остается ключевым показателем общественного развития [1], с другой – отсутствие консенсуса в теоретико-методологических вопросах [2] и ясности в определении частных и обобщенных показателей субъективного благополучия, их полноты и информативности [3] порождает новые дискуссии и поиски релевантных методических решений. Кроме того, с развитием информационных технологий становится возможным реконструкция субъективного благополучия на данных, изначально не предназначенных для этого – текстах социальных сетей и мессенджеров, цифровых следов, оставляемых пользователями Всемирной паутины [4].

Проблема измерения субъективного благополучия как оценки индивидом качества собственной жизни заключается не только в выборе наиболее релевантного из существующих методических решений, которых на сегодняшний день достаточно много ввиду разнообразных теоретических подходов к понятию «благополучие» и вариантов его операционализации [5]. Для получения сопоставимых результатов в сравнительных исследованиях или при вторичном анализе актуальной оказывается проблема реконструкции субъективного благополучия на исторических данных, когда невозможно использовать опросные методы. В этом случае одним из вариантов решения проблемы может стать анализ субъективного благополучия на основе биографических текстов [6]. Кроме того, современные информационные технологии позволяют если не автоматизировать, то по крайней мере значительно облегчить работу исследователя при работе с текстовыми данными. Так, в ряде исследований тестируются и успешно используются разнообразные способы векторизации (преобразования текста в числовые векторы) для классификации или кластеризации документов [7, 8].

Реконструкция субъективного благополучия на основе неструктурированных текстов, полученных от респондентов/информантов в ходе социологических, психологических исследований, осуществляется как в сугубо качественной парадигме [9, 10], так и с использованием количественного анализа

[11, 12] на основе специально собранных для этих целей данных. Вместе с тем в последних исследованиях все чаще для анализа качества жизни используются цифровые следы: в частности, тексты пользователей социальной сети «ВКонтакте» позволили разработать индекс (не)благополучия регионов РФ [13. С. 176–177; 14].

По сути задача реконструкции субъективного благополучия на основе текстовых данных близка одной из основных аналитических задач обработки естественного языка – анализу тональности высказываний. Если для анализа тональности предметом являются настроение, чувства, эмоции, зафиксированные в тексте, то для благополучия – субъективное восприятие качества собственной жизни. В целом и анализ тональности текста, с определенными оговорками, можно рассматривать как реконструкцию субъективного благополучия, но только на уровне эмоциональной компоненты. Для реконструкции когнитивного и поведенческого компонента нужна более сложная модель, которая классифицирует фрагменты текста в соответствии с заданной моделью благополучия.

Следует еще заметить, что формализованные методики могут показать влияние только отдельных компонентов благополучия, возможно, их синергетический эффект, но не сам феномен так, как он переживается человеком и репрезентируется в коммуникации. В этом плане преимущество биографий как эмпирических данных заключается в возможности изучения самого феномена субъективного благополучия в его нарративной репрезентации.

Результаты проведенного нами ранее исследования позволили выявить гендерно-маркированные модели представлений о будущей семейной жизни и профессиональной траектории, связанные с родительской семьей и субъективным благополучием [15]. Количественные данные анкетного опроса позволили только обозначить влияние субъективного благополучия на биографическое проектирование, но остались вопросы о самом феномене благополучия, его компонентах – что именно приносит радость, удовлетворение, позволяет себя комфортно чувствовать и планировать свое будущее.

В настоящей статье представлены результаты реконструкции субъективного благополучия на основе автобиографических текстов старшеклассников. Вначале представлены результаты выполненного вручную кодирования текстовых данных – соотнесения фрагментов биографических текстов с основными компонентами субъективного благополучия. Во второй части продемонстрирована возможность машинного обучения для создания модели для классификации и автоматизации кодирования, позволяющей осуществить по крайней мере предварительную разметку данных.

Методы и данные

Для изучения субъективного благополучия на основе биографических текстов была использована концепция психологического благополучия К. Рифф. Выбор именно этой концепции сделан по нескольким причинам. В первую очередь, в ней объединены эвдемонистический и гедонистический подходы и субъективное благополучие рассматривается как многомерный феномен, включающий эмоциональную, когнитивную и поведенческую компоненты. Очевидная релевантность содержанию биографических текстов школьников также стала причиной этого выбора.

В методику К. Рифф [16, 17] включены шесть измерений, имеющих эмпирически доказанную положительную связь с позитивным восприятием индивидом собственной жизни: 1) самопринятие (позитивная оценка себя и собственной жизни); 2) положительные отношения с другими (наличие теплых и доверительных отношений с людьми); 3) автономия (независимость от социального окружения); 4) управление ситуацией (способность изменить ситуацию или отношение к ней, умение преодолевать трудности для реализации своих целей); 5) цели в жизни (осмысленность жизни, наличие целей); 6) личностный рост (стремление к развитию, реализации потенциала). В отечественной науке существует две русскоязычные версии опросника: Т.Д. Шевеленковой и Т.П. Фесенко [18] и Н.Н. Лепешинского [19].

Эмпирические данные – корпус биографических текстов (автобиографий и биографических эссе «Я через N лет») петербургских старшеклассников 1990–1993 и 2002 гг., хранящийся в Биографическом фонде Социологического института ФНИСЦ РАН, а также аналогичные данные, полученные по аналогичной методике в 2010 и 2018–2019 гг. Ручное кодирование фрагментов биографических текстов школьников по шести обозначенным выше категориям осуществлено на выборке текстов 2002 и 2010 гг. ($n = 197$) с целью использовать размеченные данные для дальнейшего полуавтоматического кодирования оставшихся текстов.

Кодирование биографических текстов (соотнесение фрагмента текста с той или иной категорией) представляет собой задачу классификации, которая может быть решена с помощью методов обработки естественного языка, что предполагает создание векторного представления текста, выбора классификатора и подбора его параметров. В нашем случае для семантической векторизации текста использована языковая модель BERT (нейронная сеть, основанная на архитектуре трансформер и представляющая собой композицию кодировщиков) [20], а в качестве классификаторов протестированы логистическая регрессия и метод случайного леса.

Репрезентация субъективного благополучия в биографическом тексте

На первом этапе были вручную закодированы¹ фрагменты автобиографий – отдельные предложения отнесены к одной из шести категорий, соответствующих тому или иному компоненту (аспекту) благополучия. Только в редких случаях одно предложение или фрагмент текста относился к двум категориям одновременно. Кроме того, отмечены отдельные контекстуальные фрагменты, важные для понимания особенностей субъективного благополучия в конкретном случае, которые, однако, не могли быть отнесены к одной из шести категорий.

Рассмотрим каждую категорию более подробно с небольшими иллюстрациями из биографических текстов.

Самопринятие. Во фрагментах, которые были отнесены к самопринятию, как правило, описывались личностные характеристики, взгляд на себя со стороны других. В биографиях, как правило, эти описания носят положи-

¹ Здесь под «кодированием» и «кодом» понимается процесс анализа данных в рамках качественной методологии – присвоение фрагменту текста ярлыка, который соотносит его с той или иной анализируемой категорией.

тельный тон («Я была отличницей, и можно сказать „лидером“ в классе»). Отмечается также удовлетворенность жизнью, нынешним состоянием («Я обожаю свою жизнь, обожаю свою мамку, а еще я знаю, что такую бурную, хотя и маленькую жизнь не проживала ни одна девчонка, какой бы классной она не была»). Удовлетворенность собственной жизнью и собственными характеристиками проецируется в биографических эссе и на будущее («У нас проходят конкурсы, у меня не плохо получается, но я до сих пор часто смотрю на все медали и грамоты и вспоминаю те прекрасные годы») и часто носят характер благопожеланий себе: «хочется видеть себя счастливой, процветающей и удовлетворенной своей жизнью». В описаниях самих себя авторы используют весьма разнообразную лексику, но семантически эти фрагменты достаточно близки.

Положительные отношения с другими, теплые и доверительные отношения с людьми – одна из наиболее распространенных категорий в биографических текстах школьников. Положительный опыт транслируется и в будущее – в желании поддерживать имеющийся круг знакомств, сохранить друзей. К этой категории также относятся и положительные впечатления от учителей и одноклассников, других людей, встретившихся на жизненном пути («У нас был очень дружный и шумный класс»). Сюда же отнесены и значимость общения, описания влюбленности, положительные отношения внутри родительской или своей собственной будущей семьи.

Автономия. Этот аспект субъективного благополучия репрезентируется во фрагментах, где подчеркивается независимость от социального окружения, общественного мнения. Казалось бы, для подростков эта категория должна быть одной из ключевых, но в биографических текстах она встречается относительно редко («В этот же год я начала бороться со своими родственниками за свободу права и слова, в общем за полную независимость. И в частности мне это удалось»; «В музыкальной школе мне настойчиво рекомендовали идти на скрипку, но я отказалась, о чем до сих пор не жалею»). В биографических текстах автономия – это своего рода баланс между независимостью и зависимостью. К этой категории также отнесены фрагменты, где подчеркивается значимость собственного мнения, следования своим правилам.

Управление ситуацией – один из наиболее редких аспектов благополучия в автобиографиях школьников. Он часто оказывается переплетен с другими аспектами – автономией, самопринятием. К управлению ситуацией отнесено не только умение совладать с трудными жизненными обстоятельствами, но и умение «просто жить» – устроить на собственный вкус свою повседневную жизнь («Я училась в классе, который мне нравился, я проводила свое свободное время в очень приятной обстановке»). Встречается эта категория в подавляющем большинстве случаев в эссе, т.е. связана с будущей жизнью, повседневностью, работой, семьей, домом («У меня <...> не очень-то и большая, но уютная квартира, в которой я буду себя хорошо чувствовать, причем этот уют будет создан мной (дизайн в духе меня)»).

Цели в жизни. Описание целей в жизни задается самим жанром биографического эссе, поэтому достаточно часто встречается в текстах школьников. Основные маркеры фрагментов, репрезентирующих этот аспект благополучия, – «глаголы желания»: надеюсь, хочу, буду. Жизненные цели, как правило, связаны с планами на будущее, касающимися будущей семьи («Вероятно

у меня уже будет семья, любящая и любимая»), образовательной, профессиональной траектории («Я хочу поступить в академию МВД, закончить ее хорошо»), материальным благополучием («К этому времени (к 33-ем годам) мне бы хотелось какой-нибудь начальный бюджет, так скажем, материальную основу для дальнейшей жизни») с личным развитием («К этому времени я хочу „найти себя“ в этой жизни, быть полезным и нужным обществу»).

Личностный рост. К этой категории может быть отнесено достаточно большое число фрагментов из биографий школьников, в которых фиксируются собственные достижения и успехи в занятиях спортом, общеобразовательных, музыкальных или художественных школах («У меня до сих пор дома есть несколько медалей»; «первое выступление (более крупного масштаба, чем музыкальная школа) в детской филармонии – играла на фортепиано»). Вместе с тем конституируют этот аспект благополучия осознанный жизненный опыт и рефлексия над собственной жизнью («стала относиться к жизни более по-взрослому»; «теперь я считаю, что мой переход в этот лицей, стал очень важным для меня»). В биографических эссе также подчеркивается успех, но в более широком круге жизненных областей: работе, карьере, саморазвитии, воспитании детей и т.п.

Всего в 197 биографических текстах выделено 565 фрагментов. Из них к компоненту «самопринятие» отнесено 66, «положительные отношения с другими» – 138, «автономия» – 10, «управление ситуацией» – 40, «цели в жизни» – 216, «личностный рост» – 95.

Лексико-семантический анализ и классификация фрагментов

Для прояснения вопроса о возможности классификации фрагментов текста, относящихся к тому или иному аспекту благополучия, необходимо проведение лексического анализа, который может показать наличие или отсутствие значимых различий между исследуемыми категориями на уровне слов. Кодирование текстов в контексте настоящего исследования представляет собой задачу их классификации, т.е. отнесение отдельного фрагмента к одному из заранее заданных классов, соответствующих анализируемым аспектам субъективного благополучия. Для ее решения необходимы, с одной стороны, нахождение релевантного способа векторизации текстов (преобразования текстов в числовые векторы, матрицы), а с другой – выбор классификатора, способного с наибольшей вероятностью прогнозировать принадлежность текста той или иной категории.

Одним из самых простых способов векторизации текста является частотный – нахождение для каждой категории абсолютного или относительного числа слов (словосочетаний) в соответствующих текстах. Для этого все фрагменты, относящиеся к тому или иному аспекту благополучия, были объединены. Корпус фрагментов, таким образом, сжался до шести документов, объединяющих тексты разных авторов. Поскольку частотный анализ и определение веса слов весьма чувствительны к исходным данным и требуют тщательной их предобработки, в полученном корпусе осуществлена лемматизация (приведение слов к словарной форме) и удалены шумовые слова. Множество шумовых слов было заимствовано из библиотеки nltk, кроме этого, в него добавлены «я», «год», «еще», «все», «весь», «который», «это», «г».

Из текстов также были удалены цифры, отдельные слова на латинице, слова, состоящие из одной буквы и встречающиеся во всем корпусе закодированных фрагментов только один раз.

Ограничение частотного анализа заключается в том, что с его помощью выделяются слова, характерные для отдельного документа, но безотносительно их распределения в других документах корпуса. Это ограничение частично снимает мера TF-IDF, которая показывает вес *n*-грамм (слов и словосочетаний) для документа из корпуса текстов, представляющий собой произведение частоты *n*-грамм в документе и обратной частоты употребления *n*-грамм в документах корпуса. Таким образом, мера TF-IDF показывает значимость слов для отдельного документа из всего корпуса.

Вместе с частотой отдельных слов (униграмм) проанализирована частота словосочетаний из двух слов (биграмм). Результаты частотного анализа и определения веса *n*-грамм на основе TF-IDF оказались довольно близкими (табл. 1).

Таблица 1. Распределение *n*-грамм по компонентам субъективного благополучия

Компонент благополучия	Наиболее частотные <i>n</i> -граммы
Самопринятие	жизнь, человек, очень, свой, свой, жизнь , знать, ребёнок, общительный, счастливый, довольный, жить, каждый, достаточно, хороший, активный
Положительные отношения	друг, хороший , свой, очень, класс, подруга , человек, семья, отношение , общаться, сей, познакомиться, поддерживать, любить, хороший подруга
Автономия	свой, мочь, совет, мнение , человек, проблема, решение, свой_проблема , начало, родственник, свобода, право, слово, общий, полный
Управление ситуацией	свой, уютный , очень, дом, работа , жизнь, жить, время, мочь, хороший, квартира, хотеть, обстановка , сидеть, нравиться
Цели в жизни	хотеть, работа , свой, работать, поступить , семья, жизнь, закончить , институт, ребёнок, хороший, жить, школа, стать, человек
Личностный рост	класс, очень, заниматься , жизнь, школа, стать, закончить , свой, пора, работать, хороший, сей, первый, время, сей, пора

Примечание. Для каждого аспекта благополучия приводятся 15 наиболее частотных *n*-грамм, жирным начертанием выделены пять *n*-грамм, имеющих наибольший вес.

Как частотное распределение *n*-грамм, так и их вес показывают, что фрагменты текста, репрезентирующие разные аспекты благополучия, различаются на лексическом уровне. Так, каждый аспект представлен только отчасти пересекающимися множествами слов и словосочетаний, к тому же большую дифференциацию добавляет их вес. Например, «человек» и «жизнь» попадают в множество наиболее частотных лексем для трех аспектов благополучия, но больший вес имеют только в *самопринятии*. При всей априорно предполагаемой схожести категорий *цели в жизни* и *личностный рост* эти аспекты различаются на лексическом уровне, что также подтверждается и многомерным шкалированием на основе матрицы TF-IDF: как видно на диаграмме рассеяния, точки соответствующих компонентов благополучия не накладываются друг на друга (рис. 1).

Как можно было заметить выше, фрагменты биографических текстов неравномерно распределены по категориям (аспектам субъективного благополучия): к двум категориям (*положительные отношения с другими* и *цели в жизни*) относятся более 60% фрагментов, а к *автономии* – только 2%. По-

скольку методы машинного обучения для создания модели классификации предполагают достаточную по объему выборку и к тому же разделение этой выборки на две части, в представленном ниже анализе использованы только три наиболее представленные в имеющемся корпусе категории (*цели в жизни, положительные отношения с другими и личностный рост*). Корпус фрагментов поделен на две подвыборки: для обучения модели и тестирования в соотношении 4 : 1.

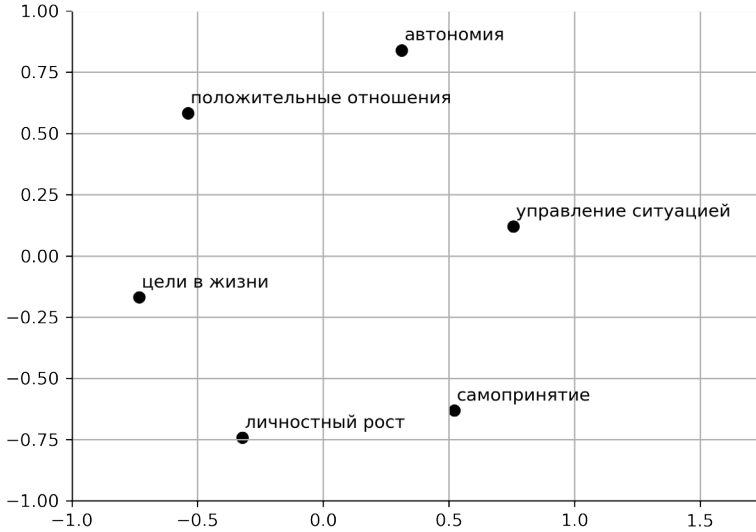


Рис. 1. Результаты многомерного шкалирования на основе матрицы TF-IDF

Для получения базовых метрик классификации использована частотная векторизация текста, а в качестве классификатора – логистическая регрессия. При многоклассовом варианте классификации средневзвешенное значение F_1 составляет 0,622, что нельзя признать вполне удовлетворительным результатом. На матрице TF-IDF это значение увеличивается, но незначительно – лишь до 0,639. При бинарном варианте классификации, когда один целевой класс противопоставляется всем другим, качество классификации улучшается (табл. 2).

Таблица 2. Средневзвешенное значение F_1 для бинарной классификации

Класс (компонент благополучия)	Частотная матрица	Матрица TF-IDF
Личностный рост	0,873	0,874
Цели в жизни	0,771	0,784
Положительные отношения	0,857	0,889

Тестирование семантической векторизации текста, осуществленной с помощью языковой модели *ruBert*, и использование аналогичного классификатора позволило увеличить точность классификации. При многоклассовом варианте средневзвешенное значение F_1 достигло 0,795. Следует отметить, что добавление переменных из матрицы TF-IDF не увеличило, а даже немного снизило точность до 0,787.

Метрики классификации для трех анализируемых классов представлены в табл. 3. Средневзвешенное значение F_1 для бинарной классификации для

компонентов «личностный рост», «цели в жизни», «положительные отношения» составило 0,92; 0,85 и 0,89 соответственно.

Таблица 3. Результаты классификации на основе семантической векторизации текста

Класс (компонент благополучия)	Основные метрики для целевого класса			Средневзвешенное значение F_1 для бинарной классификации
	Точность	Полнота	F_1	
Личностный рост	0,68	0,88	0,77	0,92
Цели в жизни	0,84	0,82	0,83	0,85
Положительные отношения	0,68	0,88	0,76	0,89

Обсуждение результатов и заключение

Результаты исследования вполне соотносятся с описанными ранее изменениями в биографическом проекте старшеклассников и также свидетельствуют о постепенном становлении реалистического типа биографического проектирования [15], когда жизненные притязания и планы адекватно соотносятся с готовностью и средствами их реализации.

Реконструкция субъективного благополучия на основе биографических данных позволила зафиксировать в нем структурные изменения между волнами исследования. Несмотря на то, что для учащихся независимо от волны исследования на первом месте в структуре их благополучия оказывается наличие *целей в жизни*, в 2010 г. замечен существенный рост веса этого аспекта благополучия в общей структуре с 34,6 до 41,2% (доля от общего числа фрагментов соответствующего года). Аналогично изменилось и положение *личностного роста*: этот компонент благополучия в 2010 г. относительно структуры 2002 г. переместился с четвертого на третье место. Таким образом, осмысленность жизни (постановка целей) и удовлетворение от уже достигнутого в большей мере оказываются значимыми для старшеклассников 2010 г. Заметное снижение веса *самопринятия* (с 15,6% в 2002 г. до 8,4% в 2010 г.) в структуре благополучия требует отдельного изучения.

С точки зрения методических результатов проведенное исследование подтвердило принципиальную возможность реконструкции субъективного благополучия на основе текстовых данных с использованием семантической векторизации и методов машинного обучения. Особое значение такого рода реконструкция имеет для тех случаев, когда невозможно по различным причинам проведение опроса с использованием той или иной методики для анализа субъективного благополучия, но имеются релевантные тексты от представителей целевой группы.

Из приведенных выше результатов может показаться, что биографии школьников изобилуют благополучием. Отнюдь. Определенная доля фрагментов с негативной тональностью присутствует. При ручном кодировании отмечались фрагменты, описывающие неблагополучие. Важный методический вопрос – как рассматривать и учитывать негативные формулировки, сопряженные с анализируемыми категориями, например: «Я не люблю своих сверстников и общение с ними». В корпусе такого рода фрагменты являются скорее исключением и не учитывались при дальнейшем анализе. В перспективе может быть использована отдельная модель для определения тональности фрагментов, а сами фрагменты с негативной тональностью позволят определить, с чем связано неблагополучие.

Из-за относительно небольшого объема размеченных вручную данных предложены дихотомические варианты классификации, где целевым классом является фрагмент биографического текста, репрезентирующий отдельный аспект субъективного благополучия, а вторым – любой другой фрагмент текста. Это позволило достичь большей точности классификации. В дальнейшем видится важным создание одной мультиклассовой модели, определяющей принадлежность к той или иной исследуемой категории и маркирующей фрагменты, не связанные ни с одним из аспектов благополучия.

В качестве продолжения работы видится важным проведение экспериментов с расширением набора данных, а также с тестированием других больших языковых моделей. Вероятно, увеличению точности классификации может способствовать добавление результатов морфологического анализа: признаков, характеризующих части речи используемых во фрагментах слов. Расширение набора данных для обучения может быть сделано с помощью представленных выше моделей, но настроенных в большей мере на полноту, чем на точность. Это позволит автоматически проработать большой объем текстовых данных с выделением фрагментов, потенциально соотносящихся с репрезентацией субъективного благополучия.

Полученные модели могут использоваться для анализа по крайней мере аналогичных биографических текстов как скрининговая диагностика субъективного благополучия. Их применение на других текстовых данных, вероятно, потребует дообучения модели.

В целом же использование языковых моделей для семантической векторизации текста в социологическом исследовании имеет большой потенциал: методы машинного обучения могут быть использованы для решения разнообразных задач предварительного кодирования данных. При этом могут использоваться не только алгоритмы классификации, но и выявления семантически близких фрагментов, кластеризации текстов. Безусловно, автоматическое кодирование не заменит работу исследователя, но способно облегчить ее, взяв на себя рутинные операции, предварительную обработку и классификацию данных.

Список источников

1. Широканова А.А. Тренды субъективного благополучия в России: 1998–2018 // Вестник Санкт-Петербургского университета. Социология. 2020. Т. 13, № 1. С. 4–24. doi: 10.21638/spbu12.2020.101
2. Шамионов Р.М., Бескова Т.В. Методика диагностики субъективного благополучия личности // Психологические исследования. 2018. Т. 11, № 60. С. 8. doi: 10.54359/ps.v11i60.277
3. Кученкова А.В., Татарова Г.Г. Субъективное благополучие: проблема анализа качественной (не)однородности населения (часть 1) // Социологические исследования. 2024. № 4. С. 14–25. doi: 10.31857/S0132162524040029
4. Кученкова А.В. Измерение субъективного благополучия на основе текстов социальных медиа: обзор современных практик // Вестник РГГУ. Серия: Философия. Социология. Искусствование. 2020. № 4 (23). С. 92–101. doi: 10.28995/2073-6401-2020-4-92-101
5. Осин Е.Н., Леонтьев Д.А. Краткие русскоязычные шкалы диагностики субъективного благополучия: психометрические характеристики и сравнительный анализ // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 1 (155). С. 117–142. doi: 10.14515/monitoring.2020.1.06
6. Дивисенко К.С. Субъективное благополучие: возможности биографического исследования // Петербургская социология сегодня. 2018. № 9. С. 47–61.

7. Кравченко Ю.А., Мансур А.М., Мохаммад Ж.Х. Векторизация текста с использованием методов интеллектуального анализа данных // Известия Южного федерального университета. Технические науки. 2021. № 2 (219). С. 154–167. doi: 10.31857/S0132162524040029
8. Артемова Е.Л., Максименко А.А., Охрименко Д.А. Применение методов машинного обучения для классификации контента коррупционной тематики в русскоязычных и англоязычных интернет-СМИ // Социология: методология, методы, математическое моделирование (Социология: 4М). 2021. № 52. С. 131–157. doi: 10.19181/4m.2021.52.5
9. Qualitative Studies in Quality of Life: Methodology and Practice / ed. by G. Tonon. Heidelberg : Springer International Publishing, 2015. doi: 10.1007/978-3-319-13779-7
10. Костина Е.Ю., Орлова Н.А., Панфилова А.О. Образ благополучия в нарративах жителей Дальнего Востока: результаты исследовательского проекта // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 1 (155). С. 38–50. doi: 10.14515/monitoring.2020.1.03
11. Carrillo A., Martínez-Sanchis M., Etchemendy E., Baños R.M. Qualitative analysis of the Best Possible Self intervention: Underlying mechanisms that influence its efficacy // PLOS ONE. 2019. May. Vol. 14, № 5. P. 1–15. doi: 10.1371/journal.pone.0216896
12. Loveday P.M., Lovell G.P., Jones C.M. The importance of leisure and the psychological mechanisms involved in living a good life: A content analysis of best-possible-selves texts // The Journal of Positive Psychology. 2018. Vol. 13, № 1. P. 18–28. doi: 10.1080/17439760.2017.1374441
13. Щекотин Е. Цифровые следы как новый источник данных о качестве жизни и благополучии: обзор современных тенденций // Вестник Томского государственного университета. 2021. № 467. С. 170–181. doi: 10.17223/15617793/467/21
14. Щекотин Е.В., Гойко В.Л., Басина П.А., Бакулин В.В. Использование машинного обучения для изучения качества жизни населения: методологические аспекты // Цифровая социология. 2022. Т. 5, № 1. С. 87–97. doi: 10.26425/2658-347X-2022-5-1-87-97
15. Дивисенко К.С. Модели будущей жизни в биографическом проекте старшеклассников // Вестник Пермского университета. Философия. Психология. Социология. 2023. № 4. С. 570–578. doi: 10.17072/2078-7898/2023-4-570-578
16. Ryff C.D. Happiness is everything, or is it? Explorations on the meaning of psychological well-being // Journal of Personality and Social Psychology. 1989. Dec. Vol. 57, № 6. P. 1069–1081. doi: 10.1037/0022-3514.57.6.1069
17. Ryff C.D., Keyes C.L.M. The structure of psychological well-being revisited // Journal of Personality and Social Psychology. 1995. Vol. 69, № 4. P. 719–727. doi: 10.1037/0022-3514.69.4.719
18. Шевеленкова Т.Д., Фесенко Т.П. Психологическое благополучие личности // Психологическая диагностика. 2005. № 3. С. 95–121.
19. Лепешинский Н.Н. Адаптация опросника «Шкала психологического благополучия» К. Рифф // Психологический журнал. 2007. № 3. С. 24–37.
20. Zmitrovich D., Abramov A., Kalmykov A., Tikhonova M., Taktasheva E., Astafurov D., Baushenko M., Snegirev A., Shavrina T., Markov S., Mikhailov V., Fenogenova A. A Family of Pre-trained Transformer Language Models for Russian // Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) / ed. by N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue. Torino, Italia : ELRA, ICCL, 05/2024. P. 507–524. doi: 10.48550/arXiv.2309.10931

References

1. Shirokanova, A.A. (2020) Trendy' sub"ektivnogo blagopoluchiya v Rossii: 1998–2018 [Trends of subjective well-being in Russia: 1998–2018]. *Vestnik Sankt-Peterburgskogo universiteta. Sotsiologiya*. 13(1). pp. 4–24. DOI: 10.21638/spbu12.2020.101
2. Shamionov, R.M. & Beskova, T.V. (2018) Metodika diagnostiki sub"ektivnogo blagopoluchiya lichnosti [Methods of diagnostics of subjective well-being of the person]. *Psikhologicheskie issledovaniya*. 11(60). pp. 8. DOI: 10.54359/ps.v11i60.277
3. Kuchenkova, A.V. & Tatarova, G.G. (2024) Sub"ektivnoe blagopoluchie: problema analiza kachestvennoy (ne)odnorodnosti naseleniya (chast' 1) [Subjective well-being: The problem of analyzing population qualitative heterogeneity (Part 1)]. *Sotsiologicheskie issledovaniya*. 4. pp. 14–25. DOI: 10.31857/S0132162524040029
4. Kuchenkova, A.V. (2020) Izmerenie sub"ektivnogo blagopoluchiya na osnove tekstov sotsial'nykh media: obzor sovremennykh praktik [Measuring subjective well-being based on social media texts. Overview of modern practices]. *Vestnik RGGU. Seriya: Filosofiya. Sotsiologiya. Iskuststvedenie*. 4(23). pp. 92–101. DOI: 10.28995/2073-6401-2020-4-92-101

5. Osin, E.N. & Leontiev, D.A. (2020) Kratkie russkoyazychnye shkaly diagnostiki sub"ektivnogo blagopoluchiya: psikhometricheskie kharakteristiki i sravnitel'nyy analiz [Brief Russian-language instruments to measure subjective wellbeing: psychometric properties and comparative analysis]. *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny – Monitoring of Public Opinion: Economic and Social Changes*. 1(155). pp. 117–142. DOI: 10.14515/monitoring.2020.1.06
6. Divisenko, K.S. (2018) Sub"ektivnoe blagopoluchie: vozmozhnosti biograficheskogo issledovaniya [Subjective Well-Being: the Potentialities of Biographical Research]. *Peterburgskaya sotsiologiya segodnya*. 9. pp. 47–61.
7. Kravchenko, Yu.A., Mansur, A.M. & Mokhammad, J.H. (2021) Vektorizatsiya teksta s ispol'zovaniem metodov intellektual'nogo analiza dannykh [Text vectorization using data mining methods]. *Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki*. 2(219). pp. 154–167. DOI: 10.31857/S0132162524040029
8. Artemova, E.L., Maksimenko, A.A. & Okhrimenko, D.A. (2021) Primenenie metodov mashinnogo obucheniya dlya klassifikatsii kontenta korruptsionnoy tematiki v russkoyazychnykh i angloyazychnykh internet-SMI [Application of machine learning methods in the classification of corruption related content in Russian-speaking and English-speaking Internet media]. *Sotsiologiya: metodologiya, metody, matematicheskoe modelirovanie (Sotsiologiya: 4M)*. 52. pp. 131–157. DOI: 10.19181/4m.2021.52.5
9. Tonon, G. (ed.) (2015) *Qualitative Studies in Quality of Life: Methodology and Practice*. Heidelberg: Springer International Publishing. DOI: 10.1007/978-3-319-13779-7
10. Kostina, E.Yu., Orlova, N.A. & Panfilova, A.O. (2020) Obraz blagopoluchiya v narrativakh zhiteley Dal'nego Vostoka: rezul'taty issledovatel'skogo proekta [The image of well-being in narratives of far-easterners: results of a research project]. *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny – Monitoring of Public Opinion: Economic and Social Changes*. 1(155). pp. 38–50. DOI: 10.14515/monitoring.2020.1.03
11. Carrillo, A., Martínez-Sanchis, M., Etchemendy, E. & Baños, R.M. (2019) Qualitative analysis of the Best Possible Self intervention: Underlying mechanisms that influence its efficacy. *PLOS ONE*. 14(5). pp. 1–15. DOI: 10.1371/journal.pone.0216896
12. Loveday, P.M., Lovell, G.P. & Jones, C.M. (2018) The importance of leisure and the psychological mechanisms involved in living a good life: A content analysis of best-possible-selves texts. *The Journal of Positive Psychology*. 13(1). pp. 18–28. DOI: 10.1080/17439760.2017.1374441
13. Shchekotin, E.V. (2021) Tsifrovye sledy kak novyy istochnik dannykh o kachestve zhizni i blagopoluchii: obzor sovremennykh tendentsiy [Digital footprints as a new source of data on quality of life and well-being: an overview of current trends]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 467. pp. 170–181. DOI: 10.17223/15617793/467/21
14. Shchekotin, E.V., Goiko, V.L., Basina, P.A. & Bakulin, V.V. (2022) Ispol'zovanie mashinnogo obucheniya dlya izucheniya kachestva zhizni naseleniya: metodologicheskie aspekty [Using machine learning to study the population life quality: methodological aspects]. *Tsifrovaya sotsiologiya*. 5(1). pp. 87–97. DOI: 10.26425/2658-347X-2022-5-1-87-97
15. Divisenko, K.S. (2023) Modeli budushchey zhizni v biograficheskom proekte starsheklassnikov [Models of the future life in the biographical project of high school students]. *Vestnik Permskogo universiteta. Filosofiya. Psikhologiya. Sotsiologiya*. 4. pp. 570–578. DOI: 10.17072/2078-7898/2023-4-570-578
16. Ryff, C.D. (1989) Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*. 57(6). pp. 1069–1081. DOI: 10.1037/0022-3514.57.6.1069
17. Ryff, C.D. & Keyes, C.L.M. (1995) The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*. 69(4). pp. 719–727. DOI: 10.1037/0022-3514.69.4.719
18. Shevelenkova, T.D. & Fesenko, T.P. (2005) Psikhologicheskoe blagopoluchie lichnosti [Psychological well-being of the individual]. *Psikhologicheskaya diagnostika*. 3. pp. 95–121.
19. Lepeshinskiy, N.N. (2007) Adaptatsiya oprosnika "Shkala psikhologicheskogo blagopoluchiya" K. Riff [Adaptation of the questionnaire "Scale of psychological well-being" by C. Ryff]. *Psikhologicheskii zhurnal*. 3. pp. 24–37.
20. Zmitrovich, D., Abramov, A., Kalmykov, A., Tikhonova, M., Taktasheva, E., Astafurov, D., Baushenko, M., Snegirev, A., Shavrina, T., Markov, S., Mikhailov, V. & Fenogenova, A. (2024) A Family of Pretrained Transformer Language Models for Russian. In: Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S. & Xue, N. (eds) *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA, ICCL. pp. 507–524. DOI: 10.48550/arXiv.2309.10931

Сведения об авторе:

Дивисенко К.С. – кандидат социологических наук, старший научный сотрудник, сектор социологии семьи, гендерных и сексуальных отношений Социологического института РАН филиала Федерального научно-исследовательского социологического центра Российской академии наук (СИ РАН – филиал ФНИСЦ РАН) (Санкт-Петербург, Россия). E-mail: k.divisenko@socinst.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Divisenko K.S. – Cand. Sci. (Sociology), senior researcher, Department of Sociology of the Family, Gender and Sexual Relationships, Sociological Institute of the RAS – Branch of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences (Saint Petersburg, Russian Federation). E-mail: k.divisenko@socinst.ru

The author declares no conflicts of interests.

*Статья поступила в редакцию 16.10.2024;
одобрена после рецензирования 28.05.2025; принята к публикации 30.06.2025
The article was submitted 16.10.2024;
approved after reviewing 28.05.2025; accepted for publication 30.06.2025*