ПРОГРАММИРОВАНИЕ, 2019, № 5, с. 10–24

АНАЛИЗ ДАННЫХ

УДК 004.89

ОПРЕДЕЛЕНИЕ ТОЧКИ ЗРЕНИЯ АВТОРА ТЕКСТА НА ОСНОВЕ АНСАМБЛЕЙ КЛАССИФИКАТОРОВ

© 2019 г. С. В. Вычегжанин[a],*, Е. В. Котельников[a],**

[a] *Вятский государственный университет 610000 Киров, ул. Московская, д. 36, Россия*

**E-mail: vychegzhaninsv@gmail.com*

***E-mail: kotelnikov.ev@gmail.com*

Поступила в редакцию 16.07.2018 г.
После доработки 09.10.2018 г.
Принята к публикации 28.11.2018 г.

В статье предложен метод решения задачи определения точки зрения автора текстового документа, основанный на машинном обучении ансамблей классификаторов. Известно, что ансамбли имеют ряд преимуществ перед отдельно взятыми классификаторами, благодаря чему часто повышается качество классификации. При этом важным вопросом является определение того, какие классификаторы должны входить в ансамбль. Представленный в статье метод построения ансамблей, основанный на процедуре перекрестной проверки, позволяет оптимизировать параметры базовых классификаторов, оценить эффективность каждого сочетания классификаторов из предложенного набора и выбрать наиболее оптимальное их сочетание. Для тестирования предлагаемого метода сформированы корпуса русскоязычных сообщений пользователей интернет-форумов и социальной сети "ВКонтакте" по трем социально значимым вопросам: прививки детям, ЕГЭ в школе и клонирование человека. Экспериментальное исследование показывает преимущество предложенного метода по сравнению с другими классификаторами.

DOI: 10.1134/S0132347419050078

1. ВВЕДЕНИЕ

Регулярно проводимые исследования показывают, что социальные медиа, такие как социальные сети, блоги и микроблоги, форумы и сайты отзывов, имеют огромную популярность в среде пользователей сети Интернет [1]. По данным компании Brand Analytics[1] в России за май 2017 года в социальных медиа зафиксировано 38 млн. авторов и 670 млн. сгенерированных ими сообщений. Анализ текстовых данных, содержащихся в таком объеме сообщений, позволяет извлекать информацию, которая полезна частным лицам, коммерческим организациям или государственным структурам. Постоянно растущий объем контента требует разработки автоматических средств анализа [2].

В настоящей работе решается задача определения точки зрения автора текстового документа (англ. *stance detection*), состоящая в выявлении позиции, которой придерживается автор текста, по отношению к объекту (или объектам) обсуждения [3]. Актуальность анализа точек зрения авторов обусловлена его применением в информационном поиске, аннотировании, рекомендательных системах, рекламе, обзорах продуктов и проверке фактов [3–6].

Выделяют следующие основные классы позиций [6, 7]:

1. *За* (for) – по тексту можно определить, что автор высказывается в поддержку целевого объекта, например[2]:

Целевой объект: *ЕГЭ в школе.*

Текст: *ЕГЭ в России однозначно принесло только пользу и более высокий уровень образования!*

2. *Против* (against) – по тексту можно определить, что автор высказывается против целевого объекта, например:

Целевой объект: *Клонирование человека.*

Текст: *Любое клонирование – это большой риск, потому лучше вообще этим не пользоваться.*

3. *Нейтрально* (neutral) – из текста можно сделать вывод о наличии нейтральной позиции, при этом она может быть указана явно ("*Я нейтрально отношусь к ...*"), либо автор приводит соображе-

[1] http://blog.br-analytics.ru/

[2] Здесь и далее примеры сообщений взяты с различных интернет-форумов и из социальной сети "ВКонтакте" и приводятся с сохранением авторской орфографии и пунктуации.

ния в пользу *за* и *против*, не склоняясь ни к одной из позиций, например:

Целевой объект: *Прививки детям*.

Текст: *А я не знаю! Вся в сомнениях... Есть и свои плюсы и свои минусы во всем этом...*

4. *Невозможность определения точки зрения* (neither) – из текста невозможно определить позицию автора, например, в том случае, когда не приводятся аргументы ни *за*, ни *против*; этот класс часто объединяется с классом *нейтрально*. Например:

Целевой объект: *ЕГЭ в школе*.

Текст: "*За или против"... Все равно от него никуда не денешься*.

5. *Согласие с предыдущей точкой зрения* (observing) – в тексте повторяется предыдущее мнение, например:

Целевой объект: *Прививки детям*.

Текст: *Во всем согласна с Вами! Тоже очень жалею, что кое-что успела поставить. Последствия АКДС до сих пор пожинаем*.

Множество целевых объектов, относительно которых определяется точка зрения, может состоять из одного объекта, например, "легализация абортов" или "феминистское движение" [7], двух объектов, например, "iPhone против Blackberry" [8], или большего количества объектов, например, "левые, правые и другие политические ориентации" [9].

Целевые объекты могут относиться к различным областям знаний и сферам деятельности, например, к политике ("коммунизм против капитализма", "Дональд Трамп"), религии ("существование Бога", "православие и католицизм"), социально значимым вопросам ("изменение климата", "смертная казнь"), продуктам ("Firefox против Internet Explorer", "Windows против Mac") и развлечениям ("Супермен против Бэтмена", "Звездные войны против Властелина колец").

В качестве источников мнений кроме социальных медиа, таких как социальные сети [10, 11] и форумы [8, 12], могут выступать дебаты в законодательных собраниях [13, 14], онлайн-новости [6] и комментарии к новостным статьям [15].

При разработке систем автоматического анализа позиции авторов сталкиваются со следующими трудностями:

– выражение автором в одном и том же тексте противоположных точек зрения по отношению к объекту или приведение автором доводов в пользу своей позиции и против другой позиции, в результате чего оказывается сложно выявить позицию автора [8];

– смена автором точки зрения на протяжении дискуссии [3];

– отсутствие упоминания целевого объекта в тексте [4];

– использование авторами, придерживающимися разных позиций, одинаковой или близкой лексики [16];

– использование иронии, сарказма и других риторических приемов [9].

Задача определения точки зрения автора текста тесно связана с проблемой анализа тональности (sentiment analysis), которая заключается в определении эмоциональной оценки, выраженной в тексте (позитивной, негативной или нейтральной), но имеет существенные отличия [7]:

1. Тональности может не быть, а позиция выражена, например:

Целевой объект: *Прививки детям*.

Текст: *Я делаю дочке все плановые прививки. А вообще, это индивидуальное решение каждой мамы*.

2. Тональность в тексте может выражаться по отношению к объекту, отличному от целевого, например:

Целевой объект: *ЕГЭ в школе*.

Текст: *Давно пора возвращать советскую систему образования, не зря ведь ее лучшей считали*.

3. Тональность и позиция могут не совпадать, например:

Целевой объект: *Клонирование человека*.

Текст: *Я считаю, что это мерзость, у клона нет души, хотя, думаю, эта идея имеет смысл и дальнейшее процветание, это своеобразный шаг к вечной жизни*.

Также имеется связь рассматриваемой задачи с аспектно-ориентированным анализом тональности (aspect-based sentiment analysis) [17]. Автор может выражать свою позицию не к объекту в целом, а к отдельным его аспектам [8].

В настоящей статье исследуются мнения пользователей интернет-форумов и социальной сети "ВКонтакте" по социально значимым вопросам. Для анализа взяты сообщения, в которых выражена позиция *за* или *против* относительно обсуждаемых объектов. Каждый пост рассматривается отдельно, поэтому считается, что позиция автора не меняется на протяжении дискуссии.

Работа имеет следующую структуру. Во втором разделе представлен обзор работ, посвященных задаче определения точки зрения автора текстового документа. В третьем разделе приведен обзор известных подходов к разработке ансамблей классификаторов. В четвертом разделе предлагается метод формирования ансамблей классификаторов на основе перекрестной проверки. В пятом разделе приводится описание текстовых корпусов, используемых при тестировании предлагаемого

метода. В шестом разделе указаны условия реализации эксперимента. Седьмой раздел посвящен анализу результатов эксперимента. В заключении подводятся итоги работы и намечаются перспективы дальнейших исследований.

2. ОБЗОР ПРЕДЫДУЩИХ РАБОТ

В зависимости от учета связей между сообщениями можно выделить два основных подхода к решению задачи определения позиции автора текста:

1) каждое сообщение рассматривается с учетом дискурса и связей с другими сообщениями, как правило при помощи графов [9, 12, 13, 16, 18, 19];

2) сообщение изучается изолированно от других сообщений [7, 8, 15].

В статье [16], которая является одной из первых работ, посвященных задаче определения позиции, рассматриваются обсуждения в группах новостей (newsgroups) и осуществляется деление авторов на противоположные лагеря на основе анализа связей сети взаимодействий и нахождения максимального разреза графа.

В работе [9] исследуются неформальные политические дискуссии. Для определения позиции используется поточечная взаимная информация и наивный байесовский классификатор. Также применяется подход на основе графа совместного цитирования, для которого выполняется сингулярное разложение с последующей иерархической кластеризацией. Метки кластеров определяются с использованием наивного байесовского классификатора. Такой подход более точен, чем простой наивный байесовский классификатор.

Авторы работы [12] используют алгоритм нахождения максимального разреза графа, представляющего диалогические отношения между говорящими. В сравнении с методом опорных векторов, наивным байесовским классификатором и классификатором, основанным на правилах, предложенный алгоритм показывает небольшое превосходство.

В статье [8] используется обучение без учителя. Из веб-страниц извлекаются оценочные слова и для них вычисляются условные вероятности по отношению к целевым объектам. Выраженная в сообщении позиция определяется методом целочисленного линейного программирования.

В [15] авторы осуществляют неотрицательную матричную факторизацию для кластеризации текстов, затем вручную размечают аргументы в пользу различных точек зрения тэгами на основе ключевых слов. Полученные тэги аргументов используют в методе опорных векторов для определения позиции.

В 2016 году в рамках Международного семинара по семантической оценке (International Workshop on Semantic Evaluation, SemEval) проходило соревнование систем автоматического определения точки зрения автора текста [7]. Исследовались англоязычные твиты, относящиеся к темам “атеизм”, “изменение климата – реальная проблема”, “феминистское движение”, “Хиллари Клинтон”, “легализация абортов” и “Дональд Трамп”. Победителем стала система организаторов на основе метода опорных векторов, а также словесных и символьных n-грамм ($F_1 = 68.98\%$). Первое место среди участников заняла система MITRE ($F_1 = 67.82\%$), основанная на рекуррентных нейронных сетях.

Для русского языка задача определения точки зрения автора решалась в [20]. При этом использовались традиционные методы машинного обучения: метод опорных векторов, метод k-ближайших соседей, наивный байесовский классификатор, деревья решений и алгоритм AdaBoost. Также применялась процедура рекурсивного сокращения признаков (Recursive Feature Elimination), позволившая отобрать наиболее значимые для данной задачи признаки, что привело к значительному улучшению качества классификации.

В настоящей статье решение задачи определения позиции основано на ансамбле классификаторов. Известно, что ансамбли позволяют повысить качество решения задачи классификации по сравнению с применением отдельно взятых классификаторов, составляющих ансамбль. По сравнению со статьей [20] данная работа имеет следующие отличия:

– в [20] целью являлось исследование различных способов представления признаков, целью настоящей статьи является исследование методов классификации;

– в данной работе предложен подход к решению задачи классификации текстов на основе ансамблей классификаторов вместо подхода, использующего отбор признаков в [20];

– при проведении экспериментов в настоящей статье используются три текстовых корпуса вместо одного.

3. АНСАМБЛИ КЛАССИФИКАТОРОВ

Ансамбли имеют следующие преимущества перед отдельными классификаторами [21]:

– уменьшают влияние случайностей, усредняя ошибки базовых классификаторов;

– уменьшают дисперсию результатов, так как несколько разных моделей, которые исходят из разных гипотез, с большей вероятностью придут к правильному результату, чем одна базовая модель.

Ключевой особенностью метода ансамблей является наличие разнообразия [22]. Это может быть разнообразие данных, разнообразие параметров или структурное разнообразие. Ниже перечислены основные подходы для создания разнообразия и описаны применяемые в рамках этих подходов техники.

1. Использование одного метода машинного обучения и разных подмножеств обучающих данных.

Наиболее известными техниками при таком подходе являются бэггинг (bagging) и бустинг (boosting). При бэггинге из исходных данных случайным образом отбираются несколько подмножеств, содержащих столько же примеров, сколько в исходном множестве [23]. Для каждого подмножества строится отдельный базовый классификатор. Результат получается на основе объединения результатов базовых классификаторов. При бустинге происходит последовательное обучение классификаторов [24]. На каждой последующей итерации обучающий набор формируется из данных, неправильно классифицированных предыдущим базовым классификатором. Такой подход применяется в алгоритме AdaBoost.

2. Использование одного метода машинного обучения и разных обучающих параметров.

Эффективной техникой составления ансамблей является отбор признаков. В методе случайных подпространств (random subspace) [25] осуществляется обучение базовых алгоритмов по случайным подмножествам признаков, а в бэггинге атрибутов (attribute bagging) [26] дополнительно определяется подходящий размер таких подмножеств признаков.

3. Использование разных методов машинного обучения.

В работах [27–30] рассматриваются ансамбли из следующих методов машинного обучения: метод опорных векторов (SVM), случайный лес (RF), наивный байесовский классификатор (NB), линейная регрессия (LR), градиентный бустинг (GB). Анализируя представленные в данных работах результаты можно сделать вывод, что для разных задач и текстовых корпусов нельзя однозначно утверждать о преимуществе ансамблей перед отдельными классификаторами.

4. Кодирование меток классов.

Для решения задачи многоклассовой классификации может быть использовано кодирование меток классов [31, 32]. Каждый класс описывается уникальным бинарным вектором. Ансамбль составляется таким образом, что каждый базовый классификатор предсказывает значение конкретной компоненты этого вектора.

Одним из важных аспектов обучения на основе ансамблей является выбор способа определения результирующего предсказания ансамбля. Для этого применяют метод взвешивания или мета-обучения [33]. При использовании метода взвешивания каждому классификатору присваивается определенный вес. Результирующее предсказание ансамбля получает вес, пропорциональный весам предсказаний базовых классификаторов. При мета-обучении применяется мета-алгоритм, для обучения которого в качестве признаков используются не исходные признаки, а предсказания базовых классификаторов. Известными техниками мета-обучения являются стекинг (stacking) [34] и деревья арбитров (arbiter trees) [35].

Еще одним важным вопросом при использовании ансамблей является определение того, какие классификаторы должны входить в ансамбль. В статье [36] выделяются методы отбора ансамблей, позволяющие определять лучший классификатор среди множества базовых классификаторов или лучшее подмножество таких классификаторов. Существует несколько стратегий отбора классификаторов в ансамбль:

1. Методология “тестируй и выбирай” основана на жадном подходе, при котором новый классификатор добавляется в ансамбль, если при этом уменьшается средняя квадратичная ошибка [37]. Для выбора нового классификатора может использоваться любая техника оптимизации, в том числе генетические алгоритмы [38].

2. Стратегия каскадных классификаторов заключается в последовательном применении базовых классификаторов. Если уверенность первого классификатора высокая, то берутся его предсказания, иначе рекурсивно запрашиваются предсказания следующего классификатора и т. д. [39, 40]

3. Динамический отбор классификаторов основан на оценке компетентности каждого классификатора в локальной области пространства признаков вблизи конкретного тестового примера. При таком подходе выбирается классификатор, который оказывается наиболее правильным в локальной области [41].

4. Область компетенции классификаторов может быть определена с помощью алгоритмов кластеризации, которые используются для выделения групп базовых классификаторов, дающих похожие предсказания. После выполнения процедуры кластеризации из каждого кластера выбираются базовые классификаторы для составления ансамбля и увеличения в нем степени разнообразия [42–44].

5. Методология жадного поиска, основанная на ранжировании, дает предпочтение классификаторам, которые способны исправлять неправильные предсказания ансамбля, как в случае ансамблей

направленного упорядочивания (Orientation Ordering Ensembles), таким образом гарантируя выбор базовых классификаторов, способных улучшать предсказания ансамбля [45].

6. Ансамбли классификаторов могут быть также составлены с помощью статистических процедур, что позволяет отбирать классификаторы с лучшей производительностью [46–48].

7. Существуют алгоритмы построения ансамблей прямым выбором (forward selection) [49] и обратным сокращением (backward elimination) [50], перенесенные из литературы по отбору признаков, которые соответственно добавляют или удаляют базовые классификаторы согласно минимизации целевой функции.

Для выбора лучшего классификатора среди множества базовых классификаторов часто используется процедура перекрестной проверки [29, 51]. Несмотря на простоту данного метода, он оказывается достаточно эффективным и сопоставим с другими более сложными методами [52]. В настоящей работе для оптимизации параметров классификаторов и оптимизации ансамбля в целом используется одна и та же процедура перекрестной проверки, что позволяет сопоставить между собой по эффективности все варианты сочетаний классификаторов и выбрать наилучший. Отличием предлагаемого в настоящей статье метода составления ансамблей от известных методов, использующих процедуру перекрестной проверки, является двухэтапный процесс отбора классификаторов. На первом этапе определяются оптимальные параметры для отдельных методов машинного обучения, т. е. производится отбор лучшего базового классификатора в рамках одного метода машинного обучения. На втором этапе отбирается лучший ансамбль из разных методов машинного обучения. За счет предлагаемого двухэтапного подхода сокращается количество вариантов перебора базовых классификаторов.

Известны работы, в которых исследуются ансамбли классификаторов для решения задачи определения точки зрения автора текста. В [53] используется ансамбль классификаторов SVM, NB и RF, а для выбора признаков применяется словарь тонально-ориентированных слов и критерий χ^2. В [29] авторы используют ансамбль классификаторов RF, GB, LR и SVM, который оптимизируют с помощью генетического алгоритма.

В статье [28] исследуется ансамбль из SVM, RF, GB и два дополнительных признака, характеризующих позитивную или негативную тональность слов при определении позиции автора. Данный ансамбль тестируется на пяти коллекциях сообщений пользователей сети Twitter. При этом улучшение F_1-меры по сравнению с применением одиночных базовых классификаторов наблюдается для двух коллекций.

В [30] исследуется ансамбль классификаторов RF, AdaBoost, SVM с линейным и радиально-базисным ядрами. Для удаления неинформативных признаков применяется критерий χ^2, а учет семантических зависимостей осуществляется с помощью методов параграфных векторов (Paragraph Vector), латентно-семантического анализа (Latent Semantic Analysis), латентного размещения Дирихле (Latent Dirichlet Allocation), лапласианских собственных карт (Laplacian Eigenmaps) и индексирования с сохранением локализации (Locality Preserving Indexing). Ансамбль также тестируется на пяти коллекциях текстов. При этом повышение качества в сравнении с одиночными базовыми классификаторами наблюдается для трех коллекций.

В отличие от приведенных выше работ в настоящей статье используется подход, основанный на процедуре перекрестной проверки, который позволяет выбрать базовые классификаторы в ансамбль, наиболее эффективный для решения задачи определения точки зрения автора текста. В качестве признаков используются слова без приведения к нормальной форме, нормализованные слова или слова определенных частей речи.

4. МЕТОД ОПРЕДЕЛЕНИЯ ПОЗИЦИИ АВТОРА ТЕКСТА

Для автоматического определения позиции автора текста предлагается использовать ансамбль классификаторов, оптимальное сочетание и параметры которых определяются на основе процедуры q-кратной перекрестной проверки (q-fold cross-validation) [54].

В основе процедуры лежит деление текстового корпуса на q равных по размеру блоков случайным образом. Каждый из этих блоков поочередно объявляется контрольным, а оставшиеся $q-1$ блоков – обучающими. Производится обучение классификаторов на объединении обучающих блоков, а затем выполняется тестирование обученных классификаторов на контрольном блоке с вычислением значений метрик качества. Итоговую оценку качества классификации вычисляют как среднее арифметическое значений метрик на всех контрольных блоках. Текстовый корпус должен быть размечен, т.е. каждый текстовый документ в корпусе имеет метку авторской позиции (*за*, *против* и т.п.).

Процесс построения ансамбля классификаторов состоит из двух основных этапов:

1) определяются оптимальные параметры классификаторов, образующих ансамбль (рис. 1);

Алгоритм 1: Процедура выбора оптимальных параметров классификаторов

Вход: $C = \{c_1, \ldots, c_n\}$ – множество классификаторов,
$P = \{P_1, \ldots, P_i, \ldots, P_n\}$ – множество параметров классификаторов,
где $P_i = \{p_1^i, \ldots, p_j^i, \ldots, p_{m_i}^i\}$ – множество наборов значений параметров классификатора c_i, T – размеченный текстовый корпус,
q – количество блоков в процедуре перекрестной проверки

Выход: $P_{opt} = \{p_{opt}^1, \ldots, p_{opt}^n\}$ – множество оптимальных значений параметров для всех классификаторов

1 Разделить корпус T на q равных блоков случайным образом:

$$T = \{T_1, \ldots, T_q\}, \quad \bigcap_k T_k = \varnothing$$

2 **цикл** *для каждого классификатора* $c_i \in C$ **выполнять**

3 **цикл** *для каждого набора значений параметров* $p_j^i \in P_i$ **выполнять**

4 **цикл** *для каждого блока* $T_k \in T$ **выполнять**

5 Обучить классификатор c_i с параметрами p_j^i на множестве

$$T_{train} = \bigcup_{l \in [1,q], l \neq k} T_l$$

6 Протестировать классификатор c_i на множестве $T_{test} = T_k$ и получить оценки качества Q_{ij}^k

конец

7 Вычислить среднее арифметическое оценок качества:

$$\overline{Q}_{ij} = \sum_k Q_{ij}^k / q$$

конец

8 Выбрать в качестве оптимального для классификатора c_i набор параметров с максимальной средней оценкой качества:

$$j_{opt} = \arg\max_j \overline{Q}_{ij}, \quad p_{opt}^i = p_{j_{opt}}^i$$

конец

Рис. 1. Процедура выбора оптимальных параметров классификаторов.

2) определяется оптимальное сочетание классификаторов в ансамбле (рис. 2).

В работе для формирования ансамбля исследовались шесть классификаторов, обеспечивающих разнообразие за счет разных подходов, лежащих в их основе:

– линейный классификатор: метод опорных векторов (Support Vector Machine, SVM) [55];

– вероятностный классификатор: наивный байесовский классификатор (Naïve Bayes, NB) [56];

– метрический классификатор: метод k-ближайших соседей (k-Nearest Neighbors, kNN) [57];

– логический классификатор: дерево решений (Decision Tree, DT) [58];

– ансамблевый классификатор: адаптивный бустинг (Adaptive Boosting, AB) [59];

– нейросетевой классификатор: быстрый текстовый классификатор (Fasttext, FT) [60].

Итоговое предсказание ансамбля определяется методом голосующего большинства, согласно которому анализируемый текстовый документ d относят к классу, выбранному большинством классификаторов [33]:

$$class(d) = \arg\max_{s \in S} \left(\sum_k g(y_k(d), s) \right), \qquad (1)$$

Алгоритм 2: Процедура выбора оптимального сочетания классификаторов

Вход: $C = \{c_1, \ldots, c_n\}$ – множество классификаторов, $P_{opt} = \{p^1_{opt}, \ldots, p^n_{opt}\}$ – множество оптимальных значений параметров классификаторов, T – размеченный текстовый корпус, q – количество блоков в процедуре перекрестной проверки

Выход: E_{opt} - оптимальное сочетание классификаторов

1 Разделить корпус T на q равных блоков случайным образом:

$$T = \{T_1, \ldots, T_q\}, \quad \bigcap_k T_k = \varnothing$$

2 Сгенерировать множество E ансамблей – сочетаний классификаторов, количество которых равно сумме биномиальных коэффициентов:

$$E = \{E_1, \ldots, E_m\}, \quad m = \sum_{k=1}^{n} \binom{n}{k} = 2^n - 1$$

3 **цикл** *для каждого ансамбля* $E_i \in E$, $E_i = \{c^1_i, \ldots, c^i_{m_i}\}$, $c^i_j \in C$ **выполнять**

4 **цикл** *для каждого блока* $T_k \in T$ **выполнять**

5 Обучить классификаторы c^i_j с оптимальными параметрами $p^i_j \in P_{opt}$ на множестве

$$T_{train} = \bigcup_{l \in [1,q], l \neq k} T_l$$

6 Протестировать ансамбль E_i на множестве $T_{test} = T_k$ и получить оценки качества Q^k_i

конец

7 Вычислить среднее арифметическое оценок качества:

$$\overline{Q}_i = \sum_k Q^k_i / q$$

8 Выбрать в качестве оптимального ансамбль с максимальной средней оценкой качества:

$$i_{opt} = \arg\max_i \overline{Q}_i, \quad E_{opt} = E_{i_{opt}}$$

конец

Рис. 2. Процедура выбора оптимального сочетания классификаторов.

где $y_k(d)$ – предсказание k-го классификатора для документа d; S – множество классов; $g(y, s)$ – функция, определяемая следующим образом:

$$g(y,s) = \begin{cases} 1, & y = s, \\ 0, & y \neq s. \end{cases} \qquad (2)$$

В случае четного числа классификаторов в ансамбле и равного количества голосов в каждом классе документу присваивается метка *за*.

5. ТЕКСТОВЫЕ КОРПУСА

Тестирование предложенного метода построения ансамблей классификаторов осуществлялось с использованием трех текстовых корпусов, составленных из русскоязычных сообщений пользователей интернет-форумов[3] и социальной сети “ВКонтакте”[4]. В процессе работы над статьей со-

[3] http://www.woman.ru/forum, http://www.yaplakal.com, http://www.rusforum.com, http://www.kid.ru и др.

[4] https://vk.com.

здавались текстовые корпуса со следующими свойствами, аналогичными свойствам корпусов из статьи [7, p. 33]:

1. Тексты и целевые объекты понятны большей части русскоязычного населения России.

2. В корпусе должно быть значительное количество текстов по трем классам точек зрения авторов: “за”, “против” и “невозможность определения точки зрения”.

– К классу “невозможность определения точки зрения” (neither) в [7] относятся как нейтральные тексты, так и тексты, для которых невозможно определить позицию.

– Количество текстов в [7] составляет по разным темам от 564 до 984, в нашем корпусе – от 1500 до 1900. При этом в [7] используется жесткое разделение текстов на обучающую и тестовую выборки, в то время как у нас применяется процедура перекрестной проверки, обеспечивающая бóльшую объективность результатов исследования.

3. Наряду с текстами, в которых явно выражен целевой объект, корпус должен содержать тексты, в которых выражается мнение по отношению к целевому объекту без явного указания на него. Корпус должен содержать относительно сложные случаи для определения точки зрения, когда целевой объект выражен через местоимения, эпитеты, взаимосвязи.

4. Наряду с текстами, в которых выражается мнение по отношению к целевому объекту, корпус должен содержать тексты, в которых объект, относительно которого выражается мнение, отличается от интересуемого целевого объекта.

Для создания корпусов с отмеченными свойствами были выбраны три социально острых вопроса, в отношении которых высказываются противоположные точки зрения: прививки детям, ЕГЭ в школе, клонирование человека. Всего размечено 5000 текстов. В процессе разметки каждый текст был отнесен к одному из трех классов: *за* (1550 текстов), *против* (1950 текстов), *невозможность определения точки зрения* (1500 текстов).

Корпус, используемый в экспериментах, составлен из текстов, принадлежащих двум классам: *за* и *против* (всего 3500 текстов[5]). Разметка осуществлялась с привлечением трех аннотаторов. Для оценки степени согласованности между аннотаторами использовалась статистическая мера каппа Флейса [61], которая оказалась равна для корпуса *Прививки детям* – 0.87, *ЕГЭ в школе* – 0.89, *Клонирование человека* – 0.86. Согласно [62] значение каппы Флейса большее 0.8 указывает на почти полное согласие аннотаторов.

В таблице 1 представлены количественные характеристики корпуса. В столбце “Общее количество слов” указано количество словоформ с повторами для каждого класса, в столбце “Количество слов в нормальной форме” – размер словаря, содержащего слова, приведенные к нормальной форме. Морфологический анализ текстов выполнялся с помощью программы *MyStem* от компании Яндекс [63]. В таблице 2 приведены примеры сообщений пользователей.

[5] https://tinyurl.com/y3n67c24

6. УСЛОВИЯ РЕАЛИЗАЦИИ ЭКСПЕРИМЕНТОВ

В качестве отдельных классификаторов, а также при построении ансамблей, использовались следующие классификаторы и параметры:

– SVM с линейным, радиально-базисным и полиномиальным ядрами, коэффициентом регуляризации $C = 10^p$, $p = [-1, 0, ..., 6]$, коэффициентом ядра $\gamma = 10^q$, $q = [-6, -5, ..., -1]$ и $\gamma = 1/N_f$, где N_f – количество признаков;

– kNN с количеством соседей $k = [1, 2, ..., 20]$;

– NB с полиномиальным распределением;

– AB с количеством решающих деревьев $n = 50$;

– DT с наибольшей глубиной дерева $max\_depth = [1, 2, ..., 20]$;

– FT с размерностью векторов слов $dim = [100, 150, 200]$, количеством итераций $epoch = [10, 20, ..., 50]$, размером контекстного окна $ws = [3, 5, 7]$, коэффициентом обучения $lr = [0.1, 0.5, 1.0]$.

При построении ансамблей применялась процедура 5-кратной перекрестной проверки ($q = 5$).

В экспериментах использовалась реализация алгоритмов машинного обучения из библиотек *scikit-learn* [64] и *fasttext* [60]. В качестве модели текстового представления использовалась векторная модель [65]. Тексты представлялись в виде n-мерного двоичного вектора, значения компонент которого характеризовали наличие или отсутствие соответствующего слова в тексте. Множество слов, учитываемых в векторной модели, образует словарь данной модели.

Для каждого текстового корпуса было проведено по четыре эксперимента, отличающихся составом словаря векторной модели:

– *Dict1* – словарь составлен из всех слов текстового корпуса без приведения их к нормальной форме;

– *Dict2* – словарь составлен из всех слов текстового корпуса, приведенных к нормальной форме;

– *Dict3* – словарь составлен из существительных, прилагательных, глаголов, наречий и междометий, приведенных к нормальной форме;

– *Dict4* – словарь содержит все слова из *Dict3* и два дополнительных слова “за” и “не” (слово “против” уже есть в словаре).

Размеры словарей представлены в таблице 3.

Таблица 1. Характеристики текстовых корпусов

Название корпуса	Метка текста	Количество текстов	Общее количество слов	Средняя длина текста, слов	Количество слов в нормальной форме
Прививки детям	за	500	35326	70	4108
	против	500	34167	68	3992
	невозможность определения точки зрения	500	22161	44	2858
ЕГЭ в школе	за	600	35410	59	4492
	против	800	40453	51	4931
	невозможность определения точки зрения	500	28038	56	3877
Клонирование человека	за	450	18860	42	3501
	против	650	27405	42	4191
	невозможность определения точки зрения	500	29238	58	4250

Таблица 2. Примеры сообщений пользователей

Название корпуса	Метка текста	Текст сообщения
Прививки детям	за	*Прививаю. Иммунолог подруга семьи, так что проблем нет.* *Опять же на мой взгляд, если придумали прививки, то не просто так.*
	против	*Мне тоже плевать на мнения врачей, что надо делать прививки.* *Надо надеяться не на врачей а верить в Бога и все будет хорошо.*
	невозможность определения точки зрения	*Прививка это дело добровольное.*
ЕГЭ в школе	за	*Боятся ЕГЭ и против него – лодыри и тупари. ЕГЭ – реальный шанс умного ребенка поступить в престижный институт.* *Ничего сложного нет.*
	против	*И еще утверждают, что ЕГЭ лучше советских экзаменов.* *Не было в СССэРе такого безобразия – НЕ БЫЛО...*
	невозможность определения точки зрения	*Мне все равно как сдавать, традиционно или ЕГЭ.*
Клонирование человека	за	*Если сам человек не против, чтобы его клонировали – почему бы и нет. Я бы разрешила.*
	против	*Я за то, что бы клонирования хоть что бы не говорили не было.*
	невозможность определения точки зрения	*Мне кажется, что это обыкновенный прогресс. И никуда от него не деться.*

В качестве простейших базовых классификаторов, результаты которых используются как точка отсчета при сравнении, реализованы два алгоритма:

– первый алгоритм (baseline 1, BL1) относит текст к классу *против*, если в этом тексте присутствует слово “против”, и к классу *за* в остальных случаях;

– второй алгоритм (baseline 2, BL2) относит текст к классу *за*, если в этом тексте присутствует слово “за”, и к классу *против* в остальных случаях.

Для оценки качества классификации использовались значения *точности* (precision, P), *полноты* (recall, R) и F_1-*меры* (F_1-measure, F_1), вычисленные по формулам [65] относительно класса s:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F1 = \frac{2 \cdot P \cdot R}{P + R}, \quad (3)$$

где TP (true positive) – количество текстов, которые классификатор правильно классифицировал как принадлежащие классу s; FP (false positive) –

Таблица 3. Размеры словарей

	Прививки детям	ЕГЭ в школе	Клонирование человека
Dict1	11584	13726	10632
Dict2	6007	6944	5807
Dict3	5716	6625	5527
Dict4	5718	6627	5529

количество текстов, которые классификатор правильно классифицировал как не принадлежащие классу s; FN (false negative) – количество текстов, которые классификатор неправильно классифицировал как не принадлежащие классу s.

Итоговые значения метрик получены способом макроусреднения по всем классам (macro-averaging) по формулам [65]:

$$P_{av} = \frac{\sum_{i=1}^{|S|} P_i}{|S|}, \quad R_{av} = \frac{\sum_{i=1}^{|S|} R_i}{|S|}, \quad F1_{av} = \frac{\sum_{i=1}^{|S|} F1_i}{|S|}, \quad (4)$$

где $|S|$ – общее количество классов.

Подбор оптимальных параметров классификаторов и итоговое тестирование отдельных классификаторов и ансамблей в настоящей работе осуществляются с использованием одних и тех же текстовых корпусов. С целью минимизации вероятности возникновения ситуации переобучения в процессе проведения экспериментов применялась процедура вложенной перекрестной проверки (nested cross-validation). Проверочное подмножество (рис. 3) использовалось как для оценки качества отдельных классификаторов с разными параметрами с целью выбора оптимальных параметров, так и для оценки качества разных ансамблей с целью выбора лучшего ансамбля. Итоговое тестирование отдельных классификаторов с оптимальными параметрами и лучшего ансамбля, выбранных с использованием проверочного подмножества, выполнялось на контрольном подмножестве. Результирующие значения F_1-меры, приведенные в таблице 4, получены для контрольного подмножества данных.

7. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Для получения объективных оценок качества в экспериментах использовалась процедура $r \times p$-кратной перекрестной проверки ($r \times p$-fold cross-validation). В данной процедуре p-кратная перекрестная проверка выполняется r раз, причем каждый раз перед ее выполнением тексты в корпусе перемешиваются случайным образом. В работе было принято $r = p = 5$. В результате для каждого классификатора было получено 25 значений F_1-меры на контрольных подмножествах данных, которые затем усреднялись для получения итоговых значений оценок качества[6], представленных в таблице 4.

Проверка статистической значимости результатов осуществлялась с использованием критерия знаковых рангов Уилкоксона (Wilcoxon signed

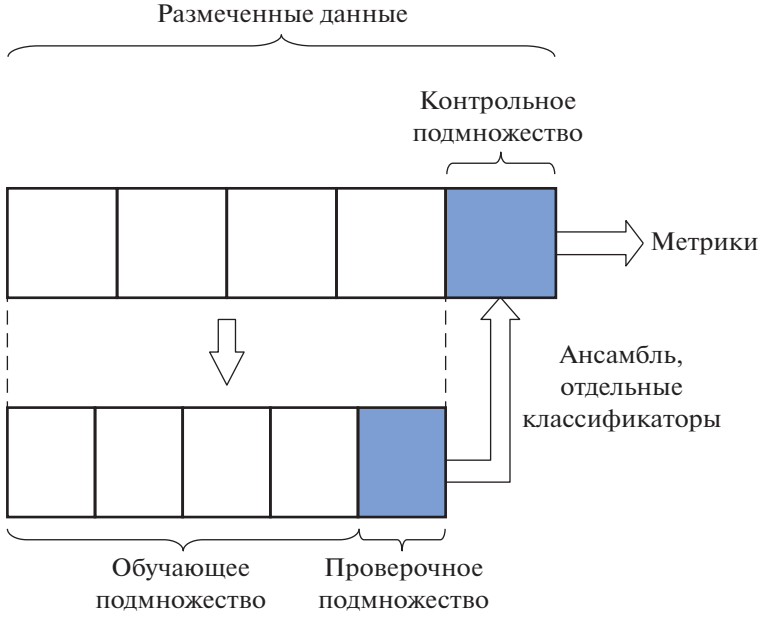

Рис. 3 – Процедура 5-кратной вложенной перекрестной проверки.

Таблица 4. Значения F_1-меры, %

Название корпуса	Словарь	BL1	BL2	SVM	NB	kNN	AB	DT	FT	Ens*
Прививки детям	Dict1	55.1	57.8	77.5	76.2	64.2	71.7	65.0	74.9	79.4
	Dict2			77.1	78.0	62.0	73.0	64.8	74.6	80.4
	Dict3			73.1	74.7	55.9	66.9	61.9	73.1	77.8
	Dict4			75.9	77.2	62.3	72.4	64.8	74.3	79.6
ЕГЭ в школе	Dict1	57.9	59.0	73.3	74.2	54.4	70.0	65.9	74.0	79.0
	Dict2			73.4	72.6	55.5	70.0	65.7	70.1	77.1
	Dict3			68.6	69.7	52.0	63.6	58.0	69.7	74.5
	Dict4			72.8	72.1	57.1	68.6	64.8	70.7	77.8
Клонирование человека	Dict1	55.9	55.7	73.3	70.4	51.2	68.1	64.1	72.4	79.3
	Dict2			72.2	72.2	54.3	68.8	63.2	72.8	78.3
	Dict3			71.5	72.3	54.7	66.9	62.6	72.0	77.4
	Dict4			72.2	72.3	54.0	68.0	64.0	71.7	77.7

* ансамбль классификаторов

rank test) [66]. В соответствии с критерием Уилкоксона выдвигаются две гипотезы: H_0 – отклонения оценок качества классификации носят случайных характер, и объединение базовых классификаторов в ансамбль не влияет на качество классификации; H_1 – отклонения оценок качества не случайны.

Расчет критерия проводился для 25 значений F_1-меры, полученных с применением процедуры 5 × 5-кратной перекрестной проверки. При расчете выполнялись следующие действия [67]:

1) вычислялись разности между значениями F_1-меры для наилучшего базового классификатора на тестовой выборке и ансамбля классификаторов на этой выборке;

2) разности ранжировались по абсолютному значению от наименьшего (ранг 1) до наибольшего (ранг 25);

3) вычислялись суммы рангов для положительных (T^+) и отрицательных (T^-) разностей (таблица 5);

4) вычислялось эмпирическое значение статистики $T_{эмп} = \min(T^+, T^-)$ и сравнивалось с критическим значением $T_{кр} = 68$, соответствующим уровню значимости $\alpha = 0.01$ и количеству испытаний $N = 25$.

На основании данных из таблицы 5 можно сделать вывод, что во всех экспериментах $T_{эмп} < T_{кр}$. Следовательно, гипотеза H_0 отвергается, т.е. разность между показателями качества для лучшего базового классификатора и ансамбля классификаторов значима на уровне $\alpha = 0.01$. Таким образом, объединение базовых классификаторов в ансамбль достоверно улучшает качество классификации.

В таблице 6 представлены наилучшие сочетания классификаторов по результатам процедуры 5 × 5-кратной перекрестной проверки. В скобках указано количество случаев получения соответствующего сочетания, выраженное в процентах к

Таблица 5. Значения сумм рангов для критерия Уилкоксона

Название корпуса	Словарь	T^+	T^-
Прививки детям	Dict1	64	261
	Dict2	36	289
	Dict3	21	304
	Dict4	16	309
ЕГЭ в школе	Dict1	3	325
	Dict2	1	324
	Dict3	2	323
	Dict4	1	324
Клонирование человека	Dict1	0	325
	Dict2	16	309
	Dict3	8	317
	Dict4	13	312

[6] Следует отметить, что указанная процедура $r \times p$-кратной перекрестной проверки использовалась только для оценки классификаторов и никак не связана с процедурой q-кратной перекрестной проверки, применяемой для построения ансамблей (см рис. 1 и 2).

Таблица 6. Наилучшие сочетания классификаторов и процент случаев их получения

	Прививки детям	ЕГЭ в школе	Клонирование человека
Dict1	SVM, FT (64%)	SVM, FT (80%)	SVM, FT (80%)
Dict2	SVM, NB, FT (44%)	SVM, FT (60%)	SVM, FT (76%)
Dict3	NB, FT (76%)	SVM, FT (72%)	SVM, FT (56%)
Dict4	SVM, FT (44%)	SVM, FT (88%)	SVM, FT (56%)

общему числу оптимальных сочетаний ($N = 25$) при однократном запуске процедуры. На основании полученных результатов можно заключить, что наиболее часто лучшее качество классификации достигалось применением ансамблей, состоящих из комбинаций методов машинного обучения SVM, NB и FT. Анализируя данные из таблицы 4, можно заметить, что именно эти методы по отдельности дают лучшие оценки F_1-меры в решении рассматриваемой задачи. Однако следует отметить, что в отдельных случаях в наилучшие сочетания классификаторов входят AB, KNN и DT, поэтому отказ от их использования привел бы к некоторому ухудшению средних оценок качества.

Применение предложенного метода построения ансамблей классификаторов позволило повысить оценку F_1-меры в зависимости от вида словаря для корпусов *Прививки детям* от 1.9% до 3,1%, *ЕГЭ в школе* от 3.7% до 5.0%, *Клонирование человека* от 5.1% до 6.0%. При этом нельзя выделить универсальный вид словаря для любого текстового корпуса.

Анализ таблицы 4 не позволяет сделать однозначный вывод о необходимости нормализации слов в задаче определения позиции автора текста: для корпусов *ЕГЭ в школе* и *Клонирование человека* словарь без нормализации *Dict1* дает наилучшие результаты, а для корпуса *Прививки детям* лидером оказывается словарь с нормализованными словами *Dict2*.

Словарь *Dict3* (нормальные формы существительных, прилагательных, глаголов, наречий и междометий) дает худшие результаты среди остальных словарей. Интересно, что добавление к этому словарю только двух слов “за” и “не” (словарь *Dict4*) позволяет повысить качество классификации (от 0.3% до 3.3%).

С помощью метода SVM были получены оценки значимости признаков для данной задачи. Вес признака (слова) определялся значением соответствующего коэффициента в линейной модели для словаря *Dict2*. В таблице 7 для каждого класса позиции приведены примеры слов, входящих в первую десятку признаков с наибольшим весом.

8. ЗАКЛЮЧЕНИЕ

Проведенное исследование продемонстрировало превосходство ансамблей по сравнению с отдельно взятыми классификаторами в решении задачи определения точки зрения автора текстового документа. С помощью статистического критерия знаковых рангов Уилкоксона обоснована достоверность наблюдаемого превосходства на уровне значимости $\alpha = 0{,}01$.

Вклад настоящей работы состоит в следующем: 1) впервые созданы три русскоязычных текстовых корпуса, снабженные разметкой для задачи определения точки зрения автора; 2) разработан метод построения ансамблей классификаторов на основе процедуры перекрестной проверки. Отличие от предыдущих работ, посвященных ансамблям для решения задачи определения позиции, заключается в способе выбора отдельных алгоритмов машинного обучения и их параметров для создания

Таблица 7. Примеры слов, имеющих наибольший вес

Название корпуса	Метка текста	Слова с наибольшим весом
Прививки детям	за	*переносить, за, честно, ставить, прививаться*
	против	*зло, отказ, против, отказник, категорически*
ЕГЭ в школе	за	*дорабатывать, шаг, честный, тяжелый, высоко*
	против	*советский, издевательство, против, придираться, неправильно*
Клонирование человека	за	*разрешать, положительно, довод, огромный, иметь*
	против	*естественный, бред, отрицательно, нельзя, индивидуальность*

ансамбля. Преимуществом предложенного метода является обоснованный выбор сочетания классификаторов среди всех возможных вариантов объединения классификаторов из заданного набора; 3) исследовано влияние нормализации слов на качество определения авторской позиции.

В дальнейших исследованиях для решения задачи определения точки зрения автора текстового документа предполагается применение методов отбора признаков, а также создание словарей, содержащих значимые слова для идентификации позиции автора текста.

9. БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Министерства образования и науки РФ, государственное задание ВятГУ № 34.2092.2017/4.6.

СПИСОК ЛИТЕРАТУРЫ

1. *Obar J.A., Wildman S.* Social media definition and the governance challenge: An introduction to the special issue // Telecommunications policy. 2015. V. 39. № 9. P. 745–750.
2. *Zafarani R., Abbasi M.A., Liu H.* Social Media Mining: An Introduction // Cambridge University Press, 2014. 320 p.
3. *Sridhar D., Foulds J., Huang B., Getoor L., Walker M.* Joint Models of Disagreement and Stance in Online Debate // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015. P. 116–125.
4. *Mohammad S.M.* Sentiment analysis: detecting valence, emotions, and other affectual states from text // Emotion Measurement. 2015.
5. *Elfardy H., Diab M., Callison-Burch C.* Ideological Perspective Detection Using Semantic Features // Fourth Joint Conference on Lexical and Computational Semantics (*SEM 2015). 2015. P. 137–146.
6. *Ferreira W., Vlachos A.* Emergent: a novel data-set for stance classification // Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2016. P. 1163–1168.
7. *Mohammad S.M., Kiritchenko S., Sobhani P., Zhu X., Cherry C.* SemEval-2016 Task 6: Detecting Stance in Tweets // Proceedings of Semantic Evaluation-2016. 2016. P. 31–41.
8. *Somasundaran S., Wiebe J.* Recognizing Stances in Online Debates // Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. 2009. P. 226–234.
9. *Malouf R., Mullen T.* Taking sides: User classification for informal online political discourse // Internet Research. 2008. V. 18. P. 177–190.
10. *Rajadesingan A., Liu H.* Identifying Users with Opposing Opinions in Twitter Debates. In: *Kennedy W.G., Agarwal N., Yang S.J.* (eds) Social Computing, Behavioral-Cultural Modeling and Prediction. SBP 2014. Lecture Notes in Computer Science, vol 8393. Springer, Cham. P. 153–160.
11. *Sobhani P., Mohammad S.M., Kiritchenko S.* Detecting Stance in Tweets and Analyzing its Interaction with Sentiment // Fifth Joint Conference on Lexical and Computational Semantics (*SEM 2016). 2016. P. 159–169.
12. *Walker M.A., Anand P., Abbott R., Grant R.* Stance Classification using Dialogic Properties of Persuasion // Conference of the North American Chapter of the ACL: Human Language Technologies. 2012. P. 592–596.
13. *Thomas M., Pang B., Lee L.* Get out the vote: Determining support or opposition from Congressional floor-debate transcripts // Conference on Empirical Methods in Natural Language Processing. 2006. P. 327–335.
14. *Burfoot C., Bird S., Baldwin T.* Collective Classification of Congressional Floor-Debate Transcripts // 49th Annual Meeting of the Association for Computational Linguistics. 2011. P. 1506–1515.
15. *Sobhani P., Inkpen D., Matwin S.* From Argumentation Mining to Stance Classification // 2nd Workshop on Argumentation Mining. 2015. P. 67–77.
16. *Agrawal R., Rajagopalan S., Srikant R., Xu Y.* Mining Newsgroups Using Networks Arising from Social Behavior // 12th International Conference on World Wide Web (WWW 2003). 2003. P. 529–535.
17. *Liu B.* Sentiment analysis and opinion mining // Synthesis lectures on human language technologies. 2012. V. 5. № 1. P. 1–167.
18. *Anand P., Walker M., Abbott R., Fox Tree J.E., Bowmani R., Minor M.* Cats Rule and Dogs Drool!: Classifying Stance in Online Debate // 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. 2011. P. 1–9.
19. *Hasan K.S., Ng V.* Stance Classification of Ideological Debates: Data, Models, Features, and Constraints // International Joint Conference on Natural Language Processing. 2013. P. 1348–1356.
20. *Vychegzhanin S., Kotelnikov E.* Stance Detection in Russian: a Feature Selection and Machine Learning Based Approach // Supplementary Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017) July 2017, Moscow, Russia. 2017. V. 1975. P. 166–179.
21. *Dietterich T.G.* Ensemble methods in machine learning // Multiple Classifier Systems. 2001. V. 1857. P. 1–15.
22. *Ren Y., Zhang L., Suganthan P.N.* Ensemble Classification and Regression – Recent Developments, Applications and Future Directions // IEEE Computational Intelligence Magazine. 2016. V. 11. № 1. P. 41–53.
23. *Breiman L.* Bagging predictors // Machine Learning. 1996. V. 24. № 2. P. 123–140.
24. *Freund Y., Schapire R.E.* Experiments with a new boosting algorithm // Machine learning: proceedings of the thirteenth international conference. 1996. P. 325–332.
25. *Ho T.K.* The random subspace method for constructing decision forests // IEEE Trans Pattern Anal Mach Intell. 1998. V. 20. № 8. P. 832–844.
26. *Bryll R., Gutierrez-Osuna R., Quek F.* Bagging: improving accuracy of classifier ensembles by using random

feature subsets // Pattern Recognition. 2003. V. 36. № 6. P. 1291–1302.

27. *Silva N.F., Hruschka E.R., Hruschka Jr E.R.* Tweet Sentiment Analysis with Classifier Ensembles // Decision Support Systems. 2014. V. 66. P. 170–179.
28. *Liu C., Li W., Demarest B., Chen Y., Couture S., Dakota D., Haduong N., Kaufman N., Lamont A., Pancholi M., Steimel K., Kubler S.* IUCL at SemEval-2016 task 6: An Ensemble Model for Stance Detection in Twitter. 2016. P. 406–412.
29. *Tutek M., Sekulic I., Gombar P., Paljak I., Culinovic F., Boltuzic F., Karan M., Alagic D., Snajder J.* TakeLab at SemEval-2016 Task 6: Stance Classification in Tweets Using a Genetic Algorithm Based Ensemble // Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 2016. P. 476–480.
30. *Xu J., Zheng S., Shi J., Yao Y., Xu B.* Ensemble of Feature Sets and Classification Methods for Stance Detection // Natural Language Understanding and Intelligent Applications. Lecture Notes in Computer Science. 2016. P. 679-688.
31. *Dietterich T.G., Bakiri G.* Solving multiclass learning problems via error-correcting output codes // Journal of Artificial Intelligence Research. 1995. V. 2. P. 263–286.
32. *Furnkranz J.* Round robin classification // Journal of Machine Learning Research. 2002. V. 2. P. 721–747.
33. *Rokach L.* Ensemble-based classifiers // Artificial Intelligence Review. 2010. V. 33. № 1. P. 1–39.
34. *Wolpert D.H.* Stacked generalization // Neural Networks. 1992. V. 5. P. 241–259.
35. *Chan P.K., Stolfo S.J.* Toward parallel and distributed learning by meta-learning // AAAI Workshop in Knowledge Discovery in Databases. 1993. P. 227–240.
36. *Re M., Valentini G.* Ensemble methods: A review // Advances in Machine Learning and Data Mining for Astronomy. Publisher: Chapman & Hall. 2012. P. 563–594.
37. *Perrone M.P., Cooper L.N.* When networks disagree: ensemble methods for hybrid neural networks // Artificial Neural Networks for Speech and Vision. Publisher: Chapman & Hall. London. 1993. pp. 126–142.
38. *Langdon W.B., Buxton B.F.* Genetic programming for improved receiver operating characteristics // Second International Conference on Multiple Classifier System. 2001. V. 2096. P. 68–77.
39. *Alpaydin E., Kaynak C.* Cascading classifiers // Kybernetika. 1998. V. 34. № 4. P. 369–374.
40. *Gamma J., Brazdil P.* Cascade generalization // Machine Learning. 2000. V. 41. № 3. P. 315–343.
41. *Cruz R.M.O., Sabourin R., Cavalcanti G.D.C.* Dynamic classifier selection // Recent advances and perspectives. Information Fusion. 2018. V. 41. P. 195–216.
42. *Giacinto G., Roli F., Fumera G.* Design of effective multiple classifier systems by clustering of classifiers // 15th International Conference on Pattern Recognition ICPR 2000. 2000. P. 160–163.
43. *Giacinto G., Roli F.* An approach to the automatic design of multiple classifier systems // Pattern Recognition Letters. 2001. V. 22. № 1. P. 25–33.
44. *Lazarevic A., Obradovic Z.* Effective pruning of neural network classifiers // Proceedings of the IEEE International Joint Conference on Neural Networks. 2001. P. 796–801.
45. *Martinez-Muniz G., Suarez A.* Pruning in ordered bagging ensembles // 23th International Conference on Machine Learning, ICML 2006. 2006. P. 609–616.
46. *Tsoumakas G., Katakis I., Vlahavas I.* Effective voting of heterogeneous classifiers // Proceedings of the 15th European Conference on Machine Learning, ECML 2004. 2004. P. 465–476.
47. *Tsoumakas G., Angelis L., Vlahavas I.* Selective fusion of heterogeneous classifiers // Intelligent Data Analysis. 2005. V. 9. № 6. P. 511–525.
48. *Yang L.* Classifiers selection for ensemble learning based on accuracy and diversity // Procedia Engineering. 2011. V. 15. P. 4266–4270.
49. *Caruana R., Niculescu-Mizil A., Crew G., Ksikes A.* Ensemble selection from libraries of models // 21th International Conference on Machine Learning, ICML 2004. 2004. P. 18.
50. *Banfield R.E., Hall L.O., Bowyer K.W., Kegelmeyer P.* Ensemble diversity measure and their application to thinning // Information Fusion. 2005. V. 6. № 1. P. 49–62.
51. *Benkeser D., Lendle S.D., Cheng J., van der Laan M.J.* Online cross-validation-based ensemble learning // Stat. Med. 2017. V. 37. № 2. P. 249–260.
52. *Dzeroski S., Zenko B.* Is combining classifiers with stacking better than selecting the best one? // Machine Learning. 2004. V. 54. № 3. P. 255–273.
53. *Liu L., Feng S., Wang D., Zhang Y.* An Empirical Study on Chinese Microblog Stance Detection Using Supervised and Semi-supervised Machine Learning Methods // Natural Language Understanding and Intelligent Applications. Lecture Notes in Computer Science. 2016.
54. *Refaeilzadeh P., Tang L., Liu H.* Cross-Validation // Liu L., Özsu M.T. (eds.). Encyclopedia of Database Systems. Springer, Boston, MA. 2009.
55. *Vapnik V.N.* The Nature of Statistical Learning Theory. NY: Springer, 2000. 314 p.
56. *McCallum A.K.* Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. 1996. http://www.cs.cmu.edu/~mccallum/bow.
57. *Cover T.M., Hart P.E.* Nearest Neighbor Pattern Classification // IEEE Transactions on Information Theory. 1967. V. 13. № 1. P. 21–27.
58. *Hunt E.B., Marin J., Stone P.J.* Experiments in induction. England: Academic Press, 1966.
59. *Freund Y., Schapire R.E.* Experiments with a new boosting algorithm // Machine learning: proceedings of the thirteenth international conference. 1996. pp. 325–332.
60. *Joulin A., Grave E., Bojanowski P., Mikolov T.* Bag of tricks for efficient text classification // Technical report, arXiv:1607.01759. 2016.
61. *Fleiss J.L.* Measuring nominal scale agreement among many raters // Psychological Bulletin. 1971. V. 76. № 5. P. 378–382.

62. *Artstein R., Poesio M.* Inter-coder agreement for computational linguistics // Journal of Computational Linguistics. 2008. V. 34. №4. P. 555–596.

63. *Segalovich I.* A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine // International Conference on Machine Learning; Models, Technologies and Applications . 2003.

64. *Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E.* Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. 2011. V. 12. P. 2825–2830.

65. *Manning C.D., Raghavan P., Schütze H.* Introduction to Information Retrieval. NY: Cambridge University Press, 2008. 482 p.

66. *Wilcoxon F.* Individual comparisons by ranking methods // Biometrics Bulletin. 1945. V. 1. № 6. P. 80–83.

67. *Flach P.A.* Machine Learning: The Art and Science of Algorithms That Make Sense of Data. Cambridge University Press, 2012. 409 p.