ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

(Н И У «Б е л Г У»)

ИНСТИТУТ ИНЖЕНЕРНЫХ И
ЦИФРОВЫХ ТЕХНОЛОГИЙ
КАФЕДРА МАТЕРИАЛОВЕДЕНИЯ И
НАНОТЕХНОЛОГИЙ

ИСПОЛЬЗОВАНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ И МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРЕДСКАЗАНИЯ ПЛАСТИЧНОСТИ ВЫСОКОЭНТРОПИЙНЫХ СПЛАВОВ

Выпускная квалификационная работа
обучающегося по направлению подготовки
28.03.03 Наноматериалы
очной формы обучения,
группы 12001714
Вережак Михаила Викторовича

Научный руководитель
д.т.н., доцент Жеребцов
Сергей Валерьевич

БЕЛГОРОД 2021

Оглавлени

e

ВВЕДЕНИЕ.....3
1 Генетические алгоритмы.....5
2 Общие сведения о машинном обучение.....10
2.1 Задача линейной регрессии.....16
2.2 Метод опорных векторов.....17
2.3 Метод к-ближайших соседей.....18
2.4 Постановка задачи исследования.....19
3. Материалы и методики.....20
3.1 Выбор материала исследования.....20
3.2 Испытания на одноосное сжатие.....21
3.3 Создание обучающего набора, расчет признаков для работы генетического алгоритма.....21
3.4 Общая схема работы генетического алгоритма, выбор набора признаков и алгоритма машинного обучения, обеспечивающего максимальную точность предсказания.....26
3.5 Алгоритм машинного обучения.....28
4 Результаты.....29
4.1 Результаты выбора дескриптора материалов для различных моделей машинного обучения.....29
4.2 Результаты работы машинного обучения.....34
4.3 Получение сплавов и оценка точности модели.....38
Выводы.....39
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ……..40

ВВЕДЕНИЕ

Металлические материалы играют решающую роль в мировой истории со времён бронзового века. Современные металлические сплавы находят ряд применений в автомобильной и аэрокосмической промышленности, а также в ядерной и биомедицинской отрасли.

До недавнего времени дизайн большинства сплавов осуществлялся путём внедрения в матрицу, состоящую из одного основного элемента легирующих компонентов, в количестве значительно меньших, чем концентрация основного элемента. Однако совсем недавно появился новый метод создания перспективных материалов, которые состоят из 5 и более элементов в примерно равных эквимолярных концентрациях [1], которыми и было предложено название высокоэнтропийные сплавы.

Обширное композиционное пространство предстоит исследовать и наметить плодотворные пути выявления областей внутри этого пространства, где могут скрываться сплавы с потенциально интересными свойствами. Число сплавов, которое можно создать с помощью данного подхода, практически безгранично, ввиду большого сочетания нескольких химических элементов и до сих пор исследовались только крошечные области данного пространства.

Высокоэнтропийные сплавы обладают исключительными свойствами, которые превосходят по свойствам традиционные сплавы. В частности, данный вид материалов проявляет сверхсвойства в совокупности, такие как высокая твердость и пластичность [2-6]. Данный вид материалов выделен в отдельную группу, так как процессы образования фаз в них, а также диффузионная подвижность атомов, механизм формирования механических свойств и термическая стабильность существенно отличаются от аналогичных процессов в традиционных сплавах.

К основным механическим свойствам сплавов относят пластичность, прочность, ударную вязкость и другие. Пластичность способствует

изменению геометрических размеров сплава без его разрушения, за счёт чего можно проводить технологические операций над образцами, таких как всесторонняя ковка, равно канальное угловое прессование и другие виды механических обработок, позволяющие принимать детали любую форму заданной конфигурации. Изготовление изделий путём деформирования без разрушения и сохранения прочностных свойств широко используется в машиностроении для создания различных видов машин.

Совсем недавно появился новый подход к предсказанию свойств материалов, который основан на применении одновременно комбинации машинном обучения и генетических алгоритмов [7]. Использование данного класса методик связано с тем, что невозможно теоретически предсказать механические характеристики многокомпонентных сплавов, так как не существует вполне чётко оформленной теории для их прогнозирования, в частности пластичности.

Генетические алгоритмы хорошо зарекомендовали себя при пространстве бесконечных решений. Так что они могут успешно использоваться для решения данной задачи. Методы машинного обучения позволяет по базе данных построить модель, которая может без экспериментальных исследований, вполне удовлетворительно рассчитать значение пластичности материала зная лишь пространство определённых признаков.

1 Генетические алгоритмы

Мы часто оказываемся в условиях выбора между различными решениями. Естественное желание сделать наилучший выбор, будучи формализованным, приводит к постановке задачи оптимизации – поиску решений, превосходящих всех остальных по некому критерию.

Существует огромное множество алгоритмов оптимизации [8]. Одним из наиболее популярных алгоритмов глобальной оптимизации является генетический алгоритм. Задачей как эволюционного алгоритма, так и любого алгоритма оптимизации состоит в нахождении экстремального значения функции в одномерном или многомерном пространстве.

Генетические (эволюционные) алгоритмы – это класс алгоритмов, основной идеей которых является дарвинские принципы эволюции в природе [9]. Имитируя процессы естественного отбора и воспроизводства, эволюционные алгоритмы могут находить высококачественные решения задач, включающих поиск, оптимизацию и обучение. В то же время аналогия с естественным отбором позволяет этим алгоритмам преодолевать некоторые препятствия, встающие на пути традиционных алгоритмов поиска и оптимизации, особенно в задачах с большим числом параметров и сложными математическими пространством поиска.

В последнее время резко возросло число статей, которые посвящены тематике эволюционных алгоритмов и их приложений. Многочисленные результаты исследований показали, что генетические алгоритмы могут получить широкое распространение за счёт объединения с другими методами [10] (так, например, с нейросетевыми технологиями и методами машинного обучения).

Основная блок-схема процесса работы генетического алгоритма приведена на рис. 1.

Рассмотрим основные работы эволюционных алгоритмом.

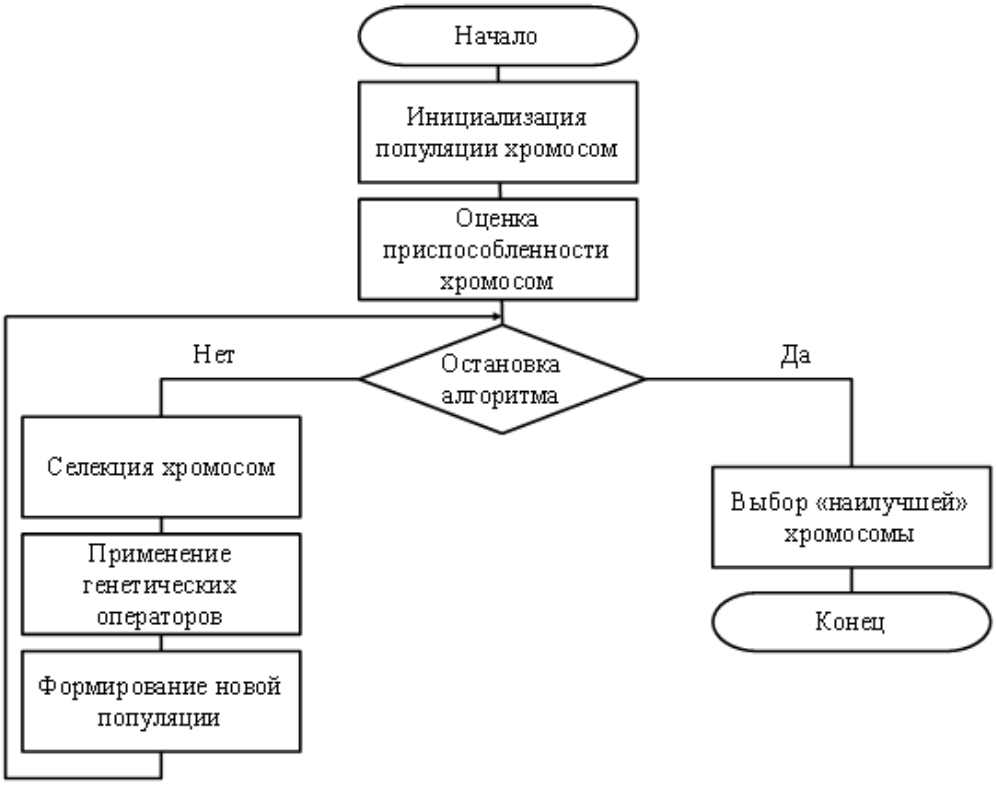


Рис. 1. Блок схема работы генетического алгоритма

Перед непосредственным запуском генетического алгоритма необходимо провести формулировку задачи на машинном языке. С этой целью выполняется переподготовка данных. Задачу формализуют, так что её решение кодируется в виде хромосомы, которая представляет собой совокупность генов. В зависимости от типа решаемой задачи ген может быть битом, числом или неким объектом.

Обычно в эволюционных алгоритмах длина хромосомы имеет постоянную длину, но это требование не является обязательным. Одним из наиболее используемых методов для кодирования хромосом является двоичное кодирование, которое заключается в том, что аллели всех генов равны либо 1 или 0.

Другим вариантом кодирования является вещественное кодирование, где в отличие от битового кодирования в качестве значений аллель гена может выступать любые вещественные числа.

На первой стадии работы эволюционного алгоритма происходит инициализация начальной популяции индивидуумов случайным образом. Начальная популяции представляет собой совокупность индивидуумов. Каждый индивид представляет собой определённую последовательность генов, каждый из которых отвечает за каждый признак особи.

На втором этапе происходит расчёт фитнесс-функции для каждой особи в популяции, таким образом в эволюционных алгоритмах происходит оценивавшие популяции индивидуумов. Оценивавшие необходимо проводить с целью выявления более и менее приспособленных особей. Происходит сравнение функций пригодности для каждого индивидуума из популяции. Особи получают право на размножение и дальнейшее продолжение популяции, если они проходят критерий качества модели.

На третьей стадии эволюционного алгоритма применяют три основные генетические оператора, такие как мутация, селекция и скрещивание.

Оператор скрещивания продуцирует новые особи из двух индивидуумов. Именно этот тип оператора обеспечивает видовое разнообразие особей в популяции. Существуют следующие типы операторов скрещивания: одноточечные, двухточечной, многоточечный, равномерный и неравномерный.

Оператор отбора применяется с целью выбора наилучших особей из популяции и последующим использованием только этих индивидуумов. Отбор носит вероятностный характер и вероятность выбора особи зависит от приспособленности индивидуума. Существуют следующие методы отбора: правило рулетки, стохастическая универсальная выборка, ранжированный отбор.

С целью непопадания популяции или большинство числа индивидуумов в локальный минимум (что приведёт к преждевременной сходимости) сложного неисследованного пространства поиска решений задачи применяют операцию мутации.

Мутация — это генетический оператор, который применяется к каждой особи в популяции, который незначительно модифицирует индивидуум, изменяя лишь либо один ген (мягкая мутация) или большее число генов (жёсткая мутация). Данный генетический оператор обычно используется после применения операторов скрещивания и отбора.

Операция мутация является вероятностной, поэтому она выполняется либо редко, либо с заданной вероятностью, которая определяется на начальной стадии работы эволюционного алгоритма. Данный оператор незначительно применяется к популяции, так как он может ухудшить качество индивидуумов. Существуют несколько видов операций мутаций: обменом, обращением и перетасовкой.

Результаты работы эволюционного алгоритма зависит от его основных параметров. К основным параметрам относят:

1). Размер популяции

2). Количество поколений

3). Разновидность оператора скрещивания

4). Вероятность мутации

5). Разновидность оператора мутации

6). Вероятность скрещивания

К основным преимуществам эволюционных алгоритмов относят:

1). Генетические алгоритмы не требуют хорошо проработанной математической теории.

2). Позволяют решать задачи в пространстве в бесконечномерном пространстве

3). Широкая область реализации

Данный вид алгоритмам применим к задачам, которые трудно описать стандартными методами программирования, в частности задач оптимизации функции одной переменной или более, а также композиций функций в сложном многомерном пространстве.

4). Возможность использования в сочетании с другим алгоритмом.

5). Возможность использования разного рода ограничений на целевую оптимизируемую функцию.

6). Устойчивы при попадании в локальные экстремумы.

7). Могут быть использованы в задачах с динамически внешне меняющейся средой.

К недостаткам эволюционных алгоритмов относят:

1). Медленное время работы при расчёте функции оценки

2). Возможность зацикливания в локальном экстремуме, что не позволяет найти глобальный максимум или минимум в решаемой задаче

3). Необходимость в переподготовке задачи для подготовки данных

Благодаря своим преимуществам генетические алгоритмы широки используются в науке и технике. Эволюционные алгоритмы позволяют осуществлять дизайн новых конструкций антенн, которые являются более эффективными (имеют большую чувствительность и большую дальность передачи сигнала). Литературный обзор показал, что генетические алгоритмы можно использовать совместно с расчётами теории функционала электронной

с целью предсказания кристаллической структуры, зная лишь химический состав материалов. В частности, в 2011 группой учёных был разработан новый метод предсказания кристаллических структур и их стехиометрии по их элементному составу (USPEX), авторами которого являются А.Р. Оганов [11]. Эволюционные алгоритмы находят широкое применение в задачах компоновки, составления расписаний, биоинформатика, широко применяются в машинном обучении, с целью поиска оптимальных моделей машинного обучения или гиперпараметра.

2 Общие сведения о машинном обучение

Машинным обучением называют область исследований, в которой компьютеры могут обучаться непосредственно без программирования [12]. Основным смыслом машинного обучения является использование подходящих признаков и моделей, которые будут оптимально подходить для решения задачи.

Машинное обучение широко используются в алгоритмах веб-поиска, финансах, медицине и маркетинге на протяжении десятилетий. Были разработаны различные модели машинного обучения для решения контролируемых (классификация и регрессия) и неконтролируемых (кластеризация) задач. Использование наиболее подходящего алгоритма для конкретной задачи имеет решающее значение для ускорения открытия новых материалов, поскольку хорошая модель машинного обучения позволяет точно предсказать неизученные материалы.

Лучшей модели машинного обучения для всех проблем с материалами не существует, и для конкретной проблемы требуется своя собственная обученная модель. Например, линейные алгоритмы приводят к простым, легко интерпретируемым моделям, но всегда приводят к плохой производительности. Метод опорных векторов является методом нелинейного анализа, который использует ядра для отображения сложных взаимосвязей и демонстрирует преимущество в возможностях работы с разбросанными выборками, нелинейностью и большим количеством измерений. Случайный лес приводит к высокой точности прогнозов, но не обеспечивает понятную человеку модель и легко приводит к переобучению. Следовательно, набор моделей машинного обучения должен быть построен в соответствии с конкретной проблемой материала и включать больше моделей, которые можно сравнивать и выбирать из них.

Основным объектом при машинном обучении в первую очередь является обучающая выборка данных. Большой объём данных позволяет

использовать сложные модели, оценивать и повышать точность, а также получать надёжные результаты.

На рис.2 представлена обобщённая схема процесса машинного обучения. Рассмотрим основные этапы машинного обучения.

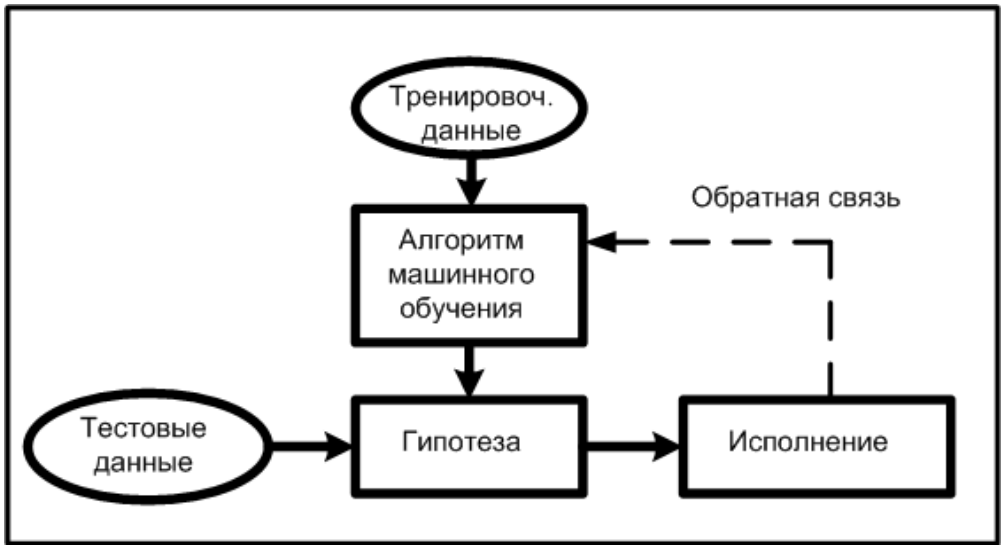


Рис. 2. Основная схема принципа работы машинного обучения

Начальный этап: предобработка данных. Исходные данные редко поступают в том виде и в той форме, которые необходимы для оптимальной работы алгоритма обучения. Поэтому данный этап является важным в любом приложении с использованием искусственного интеллекта.

Например, стандартный набор данных цветка ириса, который часто используется в качестве примера представляет собой совокупность большого числа данных, из которых необходимо выделить специальные признаки и их обработать. С этой целью было придумано множество методик уменьшения размерности пространства признаков.

Уменьшение размерности – большое количество признаков сводится к меньшему, чтобы было удобнее в последующем их визуализировать. Данный вид используется, когда необходимо уменьшить количество признаков до определённого значения, так как часто имеется доступ к большому числу

признаков. Это вызывает большие проблемы, так как необходима большая база данных для охвата небольшой части данных конфигураций признаков.

Выделение признаков для снижения размерности имеет вполне определённую задачу, чтобы в результате получить набор новых признаков, сохранив при этом большинство информации. Другими словами, мы уменьшаем количество признаков с малой потерей способности наших данных генерировать предсказания высокого качества.

Недостатком данного метода является трудно поддающиеся интерпретации получаемые новые признаки. Существует целый класс методов для уменьшения размерности: анализ главных компонент, линейный дискриминационный анализ, уменьшение размерности посредством отбора признаков.

Большинство алгоритмов искусственного интеллекта также требуют, чтобы признаки, которые необходимо отобрать были стандартизированы.

Также признаки в наборе могут быть сильно коррелированными и поэтому могут быть избыточными. Тогда применяют методы понижения размерности с целью уменьшения числа неинформативных признаков.

На втором этапе происходит отбор и тренировка модели машинного обучения. Как мы убедимся, для решения различных практических задач было разработано большое количество разных алгоритмов машинного обучения. Поэтому на практике важно сравнить, по крайней мере, несколько различных алгоритмов, чтобы натренировать и отобрать самую качественную модель. Но прежде чем мы сможем сопоставить различные модели, мы сначала должны принять решение относительно метрики для измерения их качества. Одна из самых распространенных метрик – это доля правильно классифицированных экземпляров от общего числа предсказаний.

Чтобы узнать, какая модель работает хорошо на окончательном тестовом наборе данных и реальных данных, если мы не используем этот тестовый набор для отбора модели можно воспользоваться различными

методами перекрестной проверки, или кросс-валидации, где тренировочный набор данных далее подразделяется на тренировочное и проверочные.

Также известно, что параметры различных алгоритмов обучения, предоставляемых программными библиотеками по умолчанию, оптимальны для конкретной практической задачи. Поэтому в более поздних главах мы часто будем использовать методы гиперпараметрической оптимизации, которые помогут выполнить тонкую настройку качества нашей модели.

На третьей стадии машинного обучения после отбора признаков и подходящих моделей происходит оценка модели на ранее не встречающихся объектах. После отбора модель, подогнанную под тренировочный набор данных, мы можем воспользоваться тестовым набором данных и оценить, насколько хорошо она работает на этих, ранее не встречавшихся ей данных, чтобы вычислить ошибку. Если мы удовлетворены качеством модели, то теперь мы можем использовать ее для прогнозирования новых будущих данных.

При оценивании моделей, кажется, что оценка может выглядеть тривиально: натренировать модель, а потом, вычислить насколько качественно она работает, используя специальные метрики на том же наборе данных. Данная стратегия является ошибочной, так как необходимо оценивать насколько хорошо модель на новых наборах данных. Ошибка на прежних данных может быть минимальна, вследствие эффекта переобучения. С целью преодоления данной проблемы делаются срезы данных для тестирования. В классическом варианте для проверки наблюдения разделяют на два набора, которые называют тренировочным набором и тестовым набором. Стратегия состоит в том, что мы обучаем нашу модель на тренировочном наборе и оцениваем насколько хорошо модель, работает на тестовом наборе.

Однако данный метод имеет 2 недостатка. Во-первых, качество модели зависит от размера тренировочного набора. Во-вторых, модель не тренируется с использованием всех данных и не оценивается по всем имеющимся данным.

Для решения данной проблемы был придуман метод к-блочной перекрёстной проверки. В данном методе данные разделяются на к блокам. Модель обучается на любых к-1 блоков, которые объединены в один тренировочный набор. Данная процедура повторяется к раз. Качество модели определяется путём усреднения всех полученных оценок по к-блокам.

Отметим, что параметры для таких ранее упомянутых процедур, как масштабирование признаков и снижение размерности, получают исключительно из тренировочного набора данных, и те же самые параметры позже применяют повторно для трансформирования тестового набора данных, а также любых новых образцов данных – в противном случае измеренное на тестовых данных качество модели может оказаться сверхоптимистичным.

Если модель не обладает нужным значением критерия качества, то вновь переходим ко второму этапу, в котором мы отбираем признаки и пробуем уже другие алгоритмы. Итеративный процесс заканчивается, когда мы смогли улучшить качество модели для нашей задачи.

Основное направление в области машинного обучения связано с решением задачи следующего вида: имеется множество объектов и множество ответов. Считается, что существует функциональная связь между этими множествами, но её необходимо найти. Известны только пары вида (объект, цель), которые принято называть обучающей выборкой. Задача состоит в приближённом нахождении функциональной зависимости. Такой тип машинного обучения относится к обучению с учителем. В данном случае под учителем подразумевается сама выборка, на которой модель тренируется.

Существует огромное число задач, которые решаются машинным обучением. По способу решения задачи в общем случае делятся на [12]:

1). Задача регрессии – прогноз на основе выборки объектов с различными признаками. Выходным значением в данном типе задач является действительным числом. Задачи данного типа как правило связаны с прогнозированием. Примером может служить предсказание пластичности, прочности или других физических характеристик материалов по химическому

составу и структуре сплава. К основным видам регрессии относят линейную регрессию и т. п.

2). Задача классификации – получение категориального ответа на основе набора признаков. Имеет конечное количество ответов (как правило, в формате «да» или «нет»). Например, присутствует ли данная фаза в материале, кинетически стабилен при определённой температуре или нет.

В машинном обучении задача классификации - это задача разделения множества наблюдений (объектов) на группы, называемые классами, на основе анализа их формального описания. При классификации каждая единица наблюдения относится определенной группе или номинальной категории на основе некоторого качественного свойства.

В контексте машинного обучения классификация относится к обучению с учителем. Такой тип обучения подразумевает, что данные, подаваемые на входы системы, уже помечены, а важная часть признаков уже разделена на отдельные категории или классы. Поэтому сеть уже знает, какая часть входов важна, а какую часть можно самостоятельно проверить. Пример классификации — сортировка различных растений на группы, например, «папоротники» и «покрытосеменные». Подобная задача может быть выполнена с помощью дерева решений — одного из типов классификатора.

3). Задача кластеризации – главным является разбиение данных на группы. К основным целям кластеризации относят понимание структуры данных и разбиение множества объектов на схожие группы с схожими свойствами.

4). Задача выявления аномалий – отделение аномалий от стандартных случаев. На первый взгляд она совпадает с задачей классификации, но есть одно существенное отличие: аномалии – явление редкое, и обучающих примеров, на которых можно натаскать обучающуюся модель на выявление таких объектов исчезающе мало. Поэтому методы классификации здесь не работают.

Рассмотрим четыре типа алгоритма, которые использовались для решения поставленной задачи.

2.1 Задача линейной регрессии

Линейная регрессия является привлекательной моделью, потому что ее представление очень простое.

Представление - это линейное уравнение, объединяющее определенный набор входных значений решений, к которому является прогнозируемый вывод для этого набора входных значений. Таким образом, как значения входных данных, так и выходное значение являются числовыми.

Линейное уравнение присваивает масштабный коэффициент к каждому входному значению. Добавлен также один дополнительный коэффициент, добавляющую дополнительную степень свободы (например, движение вверх и вниз по двумерному участку) и часто называют коэффициентом перехвата или смещения.

Наиболее простая задача регрессии, когда на вход подается одна переменная x и есть одно выходящее значение y. Форма подобной модели будет даваться формулой (2.1)

$$y = \beta_0 + \beta_1 x + \varepsilon$$

(2.1)

В случае многомерных измерений (т.е. когда у нас есть более одной вводной переменной (x)), линия превращается в гиперплоскости. Таким образом, представление представляет собой форму уравнения и конкретные значения, используемые для коэффициентов. Когда говорят о сложности регрессионной модели, такой как линейная регрессия - имеют ввиду количество коэффициентов, используемых в модели.

Когда конкретный элемент коэффициент становится нулевым, он эффективно удаляет влияние входной переменной на модель и, следовательно, влияния на прогноз модели.

Когда у нас есть более одной входной переменной, мы можем использовать метод наименьших квадратов для оценки значений коэффициентов. Процедура наименьших квадратов направлена на минимизацию суммы квадратов остатков. Это означает, что, учитывая регрессионную линию через данные, мы вычисляем расстояние от каждой входной точки x_i до линии регрессии, берем квадрат этого расстояние и суммируем все квадраты подобных расстояний. Данное сумма квадратов для наиболее точной подгонки модели должна быть минимальна.

2.2 Метод опорных векторов

Опорно-векторные машины – один из методов машинного обучения, который решает задачу классификации и регрессии путём проведения в многомерном пространстве размерности n гиперплоскости.

Гиперплоскость – подпространство размерности n-1 в n-мерном пространстве. Если мы хотим разделить трёхмерное пространство, то используем двухмерную гиперплоскость. Опорно-векторные машины классифицируют наборы данных путём нахождения гиперплоскости, которая максимизирует допустимый промежуток между классами в тренировочных данных.

В отношении опорно-векторных регрессоров необходимо отметить, что гиперплоскость по определению линейна. Но часто встречаются классы (данные), которые мы не можем разделить линейно идеально. Поэтому в этих случаях существует баланс между максимизацией промежутка гиперплоскости и минимизации ошибочной классификации. В опорно-векторном классификаторе последние контролируются с помощью гиперпараметра C – штрафа, налагаемого на ошибки.

Опорно-векторный классификатор можно представить, как:

$$f(x)=\beta_0+\sum \alpha_i K(x_i, x_j) \qquad (2.2)$$

, где β_0- смещение, α_i – модельные параметры, которые необходимо заучить; (x_i, x_j) − пары двух опорно-векторных наблюдений. K − ядерная функция, которая сравнивает сходство между двумя наблюдениями. Ядерная функция определяет тип плоскости.

Опорно-векторный классификатор получил своё название благодаря тому, что гиперплоскость определяется относительно небольшим количеством наблюдений, называемыми опорными векторами.

2.3 Метод к-ближайших соседей

Классификатор к ближайшим соседям является одним из простым и наиболее используемым классификатором в контролируемом машинном самообучении. Иногда его называют ленивым учеником, потому что он технически не тренирует модель делать предсказания. Вместо этого предсказывается, что наблюдение классом небольшой доли к ближайшим соседям. Для классификации в методе к-ближайших соседей проводят следующие шаги:

1). Вычисляют расстояние до каждого из объектов обучающей выборки

2). Отбирают k объектов обучающей выборки, расстояние до которых минимально.

3). Классифицируют класс объекта, элементы которого находятся в непосредственной близости.

Преимущество метода к-ближайших соседей является то, что он прост к применяемой задаче, следовательно, может быть применим в качестве затравки для решения исследуемой задачи. Минусам данного метода относят его слабую корректность.

2.4 Постановка задачи исследования

Обзор приведенной литературы показал, что высокоэнтропийные сплавы являются актуальной темой для будущих применений в промышленности в качестве функциональных и конструкционных материалов. Для многих практических приложений необходимы в первую очередь прочностные и пластические свойства.

Данный вид сплавов мало изучен, ввиду большого композиционного пространства поиска. Вследствие чего, сначала легче провести оценки материалов компьютерными методами, а потом выполнить экспериментальные исследования.

Ввиду не разработанности полной теории к определения прочностных и пластических характеристик были предложены методы их определения из недавно быстро развивающейся отрасли, такой как машинное обучение с использованием генетических алгоритмов, которые позволяют исследовать огромные пространства поиска. Поэтому ставлю целью этой работы: разработать модель машинного обучения с использованием генетического алгоритма для предсказания пластических свойств, т.е. пластичности высокоэнтропийных сплавов.

Ставлю перед собой следующие задачи:

1). Создание базы данных сплавов на основание литературных данных

2). Выбор признаков для высокоэнтропийных сплавов с помощью генетического алгоритма, которые могут успешно использоваться для предсказания пластичности данного класса материалов.

3). Подбор подходящих моделей машинного обучения и оценка их качества работы

4). Тестирование модели по предсказанию пластичности на выплавленных сплавах

3. Материалы и методики

3.1 Выбор материала исследования

В качестве материалов исследования были выбраны высокоэнтропийные сплавы, химический состав которых указан в таблице1.

Таблица 1

Химический состав выплавленных сплавов

Alloys	Al	Cr	Nb	Ti	V	Zr
A1	13	0	10	44	28	5
A2	12	2	19	44	16	7
A3	15	11	6	37	20	11
A4	11	7	25	50	14	11
A5	14	13	23	22	26	2
A6	12	1	22	31	20	14
A7	14	3	21	34	4	24
A8	10	0	20	36	13	21

Индексы в сплавах показывают атомную долю компонента. Если цифра не указана – атомная доля компонента равна 1. Сплавы были получены с помощью метода вакуумно-дугового переплава. Перед плавкой навески исходных металлов промывались в ацетоне в ультразвуковой ванне. Для получения химически однородных сплавов образцы переплавляли не менее 5 раз, вес готовых заготовок составил 20 г. После проведения выплавки над образцами производили резку на электроэрозионной установке LaboPol-5 фирмы «Struers» с применением механического шлифования на наждачных кругах P120-P4000. Шлифованные образцы запаивались в кварцевую трубку и

отжигались при 1200^0C в течение 10 часов. После отжига образцы закаливались в воду.

3.2 Испытания на одноосное сжатие

Испытания на осадку образцов высокоэнтропийных сплавов выполняли согласно ГОСТ 8817-82 «Металлы. Метод испытания на осадку». Размер образцов для испытаний имеет следующие геометрические параметры 6×6×9 $мм^3$. Осадка сплавов происходила по методу одноосного сжатия с помощью испытательной машины Instron при комнатной температуре. Была выбрана скорость деформации $10^{-4}с^{-1}$. Поверхность образцов подготавливалась шлифовкой и полировкой. Шлифовка производится на вращающемся круге, на поверхность которого закреплена абразивная бумага.

При шлифовании необходимо переходить от одной наждачной бумаги к другой, уменьшая при этом зернистость. Каждый раз при смене наждачной бумаги образец поворачивают на угол 90°.

Следующим этапом за шлифовкой является полировка, целью которой является устранение царапин, которые возникли от воздействия абразивных частиц, которые образовались во время шлифовки. Для полировки использовались специальные полировочные диски и суспензии с абразивными частицами и коллоидной-кремниевой суспензии OP-S.

Так была подготовлена поверхность образцов для металлографических исследований и в том числе для проведения механических испытаний на осадку.

3.3 Создание обучающего набора, расчет признаков для работы генетического алгоритма

Для того, чтобы решить задачу предсказания пластичности высокоэнтропийных сплавов встала необходимость создания обучающего

набора данных. Для этого были использована база данных [16]. В результате удалось собрать базу данных размером в 100 сплавов, которая представляет собой таблицу химических составов материалов и их пластичности.

Одним из важных факторов, влияющих на точность модели в машинном обучении является набор используемых признаков материалов. С этой целью важно создать такой набор, но поскольку неизвестно какие признаки являются оптимальными необходимо использовать как можно больше «хороших» дескрипторов. Тогда можно будет гарантировать, что выбор признаков объясняет заданное целевое свойство (в нашем случае — пластичность).

Совокупность признаков материалов (сплавов) может быть построено на основе элементарных, термодинамических, кинетических параметрах, а также на основе знания структурной информации. К элементарным параметрам относят такие параметры как атомная масса, период и группу в таблице Менделеева, а также такие свойства элементов как электроотрицательность по Полингу, атомный радиус и другие важные свойства.

Термодинамические и кинетические параметры, такие как свободная энергия, энтропия и энтальпия смешения и коэффициенты диффузии, используются для определения стабильности конкретной фазы или протекания конкретной реакции.

Результаты вычислений теории функционала электронной плотности, включая объем элементарной ячейки, ширину запрещенной зоны, энергию когезии, упругие постоянные, диэлектрические постоянные, электронную структуру и фононные свойства, более многочисленны и обычно внедряются в дескрипторы структуры машинного обучения для повышения производительности моделей машинного обучения. Несмотря на вышеупомянутые общие аспекты построения пространства дескрипторов материалов, конкретные дескрипторы также должны быть добавлены к конкретным задачам.

Мы свели признаки материалов, которые потенциально влияют на пластичность высокоэнтропийных сплавов [13-15]. В литературе предложены

следующие дескрипторы эмпирических материалов: энтальпия смешения, энтропия смешения и дескриптор (Ω), который количественно определяет преобладание энтропии над энтальпией, являются основными термодинамическими дескрипторами. Элементные дескрипторы, включая концентрацию валентных электронов, разницу в электроотрицательности Полинга, электроотрицательности Аллена, разность размеров атомов и другие геометрические дескрипторы, рассчитываются на основе расширенных классических правил Юм-Розери.

Каждый элементарный параметр в нашем пространстве композиций может быть представлен молярным средним значением X элементного параметра X_i как:

$$\overline{X} = \sum_{i=1}^{N} c_i X_i \tag{3.3.1}$$

, где c_i - мольная доля этого элемента. Другой способ описать каждый сплав состоит в вычислении значения рассогласования между элементарными параметрами X_i его компонентов. δX задается формулой:

$$\delta X = \sqrt{\sum_{i=1}^{N} c_i \left(1 - \frac{X_i}{\overline{X}}\right)^2} \tag{3.3.2}$$

, где c_i - мольная доля этого элемента, а $\overline{X}$ - среднее молярное значение, полученное с помощью уравнения (3.3.1) . Таким образом, вводятся 76 дескрипторов материалов, которые могут вносить свой вклад в пластичность высокоэнтропийных сплавов. Всего пространство дескрипторов материалов содержит 76 дескрипторов материалов, которые перечислены в таблице 1. Исходные данные обучения включают информацию о соответствующих химических составах. Все композиции используются для вычисления признаков по формулам, приведенным в таблице, с вводом свойств элементов. Поскольку мы можем легко получить доступ к этим свойствам элементов, все

дескрипторы доступны как для наших обучающих данных, так и для виртуальных данных.

Поскольку дескрипторы материалов различаются по величине, единицам измерения и большинство алгоритмов машинного обучения используют эвклидово расстояние между двумя точками данных, дескрипторы материалов с высокими величинами будут весить больше по сравнению с дескрипторами с низкими величинами. Чтобы привести дескрипторы всех материалов к одному и тому же уровню величин, каждый дескриптор материалов подвергают минимаксному шкалированию. В минимаксном шкалировании для определения новых значений признаков из старых используют наименьшие и наибольшие значения. В частности, минимакс определяют следующим образом:

$$x_i' = \frac{x_i - min\,(x)}{max(x) - min\,(x)}$$

, где $\{x_i\}$ - вектор признаков, x_i' - прошкалированный элемент.

Таблица 2

Набор признаков для обучающего алгоритма

Параметр	Аббревиатура	Формула
Разница атомных размеров	R	$\overline{R} = \sum_{i=1}^{N} c_i R_i$ $\delta R = \sqrt{\sum_{i=1}^{N} c_i \left(1 - \frac{R_i}{\overline{R}}\right)^2}$
	λ	$\lambda = \frac{\Delta S_m}{\delta^2}$
Электроотрицательность	χ_P	

по Полингу		$\overline{X}=\sum_{i=1}^{N} c_i X_i$ $\delta X=\sqrt{\sum_{i=1}^{N} c_i\left(1-\frac{X_i}{\overline{X}}\right)^2}$
Электроотрицательность по Аллену	χ_A	
Электроотрицательность по Оганову	χ_O	
Концентрация валентных электронов	VEC	
Энтальпия смешения	ΔH_{mix}	$\Delta H_{mix}=4\sum c_i c_j H_{ij}$
Энтропия смешения	ΔS_{mix}	$\Delta S_{mix}=-R\sum_{i=1}^{n} c_i \ln c_i$
Комбинационный эффект энтальпии и энтропии смешения	Ω	$\Omega=\frac{T_m \Delta S_m}{\Delta H_m}$
Температура плавления	*MT*	$\overline{X}=\sum_{i=1}^{N} c_i X_i$ $\delta X=\sqrt{\sum_{i=1}^{N} c_i\left(1-\frac{X_i}{\overline{X}}\right)^2}$
Энергия образования твёрдого раствора	*CE*	
Модуль Юнга	*MC*	
Первый ионизационный потенциал	*FIE*	
Второй ионизационный потенциал	*SIE*	
Третий ионизационный потенциал	*TIE*	
Работа выхода	*WF*	
Атомный номер	*AN*	
Квантовое число	*QN*	
Колонка в периодической системе	C	
Релятивистская атомная масса	*RAM*	

Атомный объём	*VA*	
Электроотрицательность (Бацанов)	χ_{MB}	
Электроотрицательность (Альфред Рошоу)	χ_{AR}	
Абсолютная электроотрицательность	χ_{ABS}	
Химический потенциал	*CPM*	
Эффективный заряд ядра	*NCE*	
Температура по Больцману	*TB*	
Энтальпия испарения	*EV*	
Энтальпия плавления	*EM*	
Атомная энтальпия	*EA*	$\overline{X} = \sum_{i=1}^{N} c_i X_i$ $\delta X = \sqrt{\sum_{i=1}^{N} c_i \left(1 - \frac{X_i}{\overline{X}}\right)^2}$
Ионный радиус	*RI*	
Ковалентный радиус	*RC*	
Плотность валентных электронов	*DVE*	
Плотность	*D*	

3.4 Общая схема работы генетического алгоритма, выбор набора признаков и алгоритма машинного обучения, обеспечивающего максимальную точность предсказания

Нетрудно убедиться, что число комбинаций из отобранных признаков неограниченно огромно. С этой целью используем генетический алгоритм, который может достаточно быстро найти оптимальное решение. Каждый признак сплавов шифруется целым числом от 1 до 72, что обозначает номер выбранного признака. Следовательно, набор дескрипторов, подлежащих оптимизации будет описываться строкой, состоящей из вещественных чисел

от 1 до 72. Основная блок-схема эволюционного приведена на рис.3. Определённый набор признаков должен обеспечить наибольшую точность для определения пластичности высокоэнтропийных сплавов.

Таким образом может быть достигнута оптимальная комбинация модели машинного обучения и дескрипторов материалов.

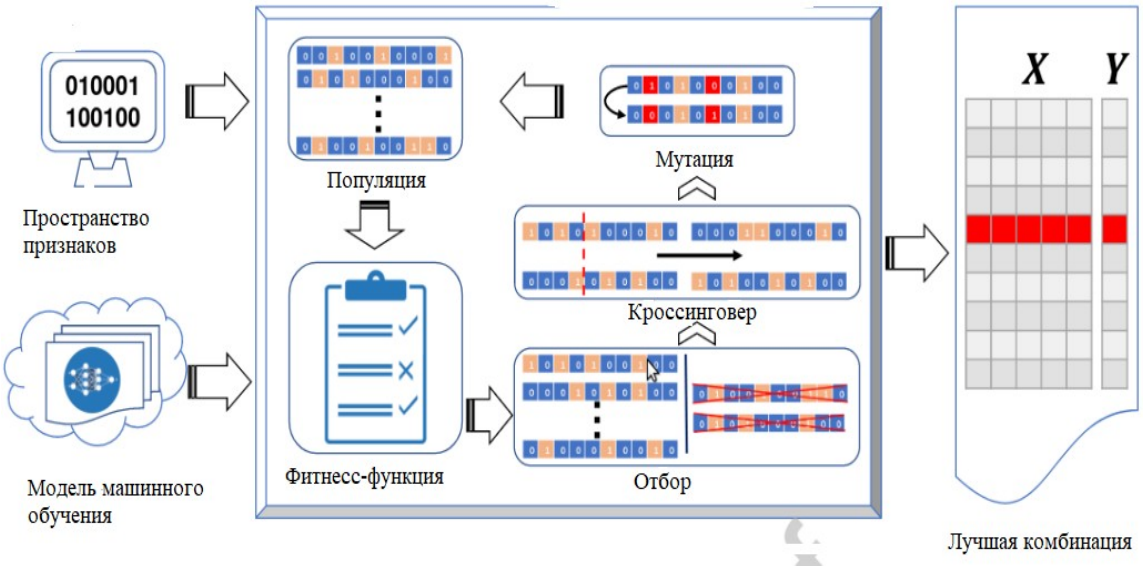


Рис. 3. Схема работы эволюционного алгоритма с машинным обучением

Как видно из блок-схемы работы генетического алгоритма совместно с машинным обучением на начальном этапе случайным образом инициализируется начальная популяция особей, представляющая собой строку генов, которые представляют собой определённый признак, который определён числом от 1 до 72.

Рассчитав значения функции пригодности для начальной популяции, которая представляет собой совокупность строк из признаков применили основные операторы эволюционного алгоритма: отбора, кроссинговера и мутации.

Новая популяция получившихся решений будет являться «начальной» популяцией. Цикл будет повторяться до тех пор, пока не выполнится хотя бы

одно из условий остановки: число поколений достигло максимально заданного количества или была достигнута наилучшая точность.

Для тестирования различных регрессивных моделей машинного обучения данные разбивались случайным образом, так что на 70% происходила тренировка модели, а на 30 % - её тестирование. Данная процедура повторялась 10 раз. Далее полученные результаты тестирования усреднялись, а полученные величины трактовались как результаты работы регрессора.

3.5 Алгоритм машинного обучения

В машинном обучении гиперпараметрами называют такие числовые значения переменных в модели, которые устанавливаются только перед процессом обучения. Гиперпараметры необходимы для контроля процесса обучения. Метод опорных векторов использует несколько гиперпараметров: C и epsilon.

С целью нахождении оптимального гиперпараметра модели признаки необходимо разделить на 3 части:

1). Тренировочный набор данных, который используется для тренировки модели машинного обучения.

2). Валидационный набор данных, который используется для оценки качества обучения модели в процессе перебора обучающихся гиперпараметров.

3). Тестовый набор данных, который используется для оценивания критерия качества работы модели.

Оптимизация гиперпараметра осуществлялась с помощью метода поиска по решётке, который делает перебор значений заданной величины по заданному вручную множеству пространства гиперпараметра метода опорных векторов.

4 Результаты

4.1 Результаты выбора дескриптора материалов для различных моделей машинного обучения

Результатом работы генетического алгоритма является оптимальный набор признаков, который обеспечил минимального среднеквадратичное отклонение на тестовых данных для каждой из 4 моделей машинного обучения.

В данной части работы выбираются определённые дескрипторы материалов для различных моделей обучения. Как было раньше показано одним из основных методов для решения таких задач с огромным числом признаков является генетический алгоритм. Данный метод способен находить наилучшие признаки для различных моделей искусственного интеллекта, применимо к задаче определению пластичности сплавов.

В качестве примера суррогатной модели мы выбрали: метод опорных векторов, алгоритм к-ближайших соседей, байесовский регрессор и линейную регрессию.

На первом этапе эволюционного алгоритма был проведён расчёт точности метода опорных векторов 20 раз с целью получения оптимального числа признаков.

Результатом стало построение графика точности метода опорных векторов для регрессии, который изображён на рис.4. Линия на графике показывает среднеквадратичную ошибку регрессора метода опорных векторов от числа признаков. Заметим, что при числе признаков равным 8 среднеквадратичная ошибка минимальна. Исходя из данного предположения выберем 8 признаков для решения поставленной задачи.

Первоначально набор признаков материалов составляет 300 подмножеств, которые выбираются случайным образом. Эти 300 разных подмножеств являются родительским поколением для следующих. Для каждого подмножества признаков используется метод опорных векторов с rfb ядром.

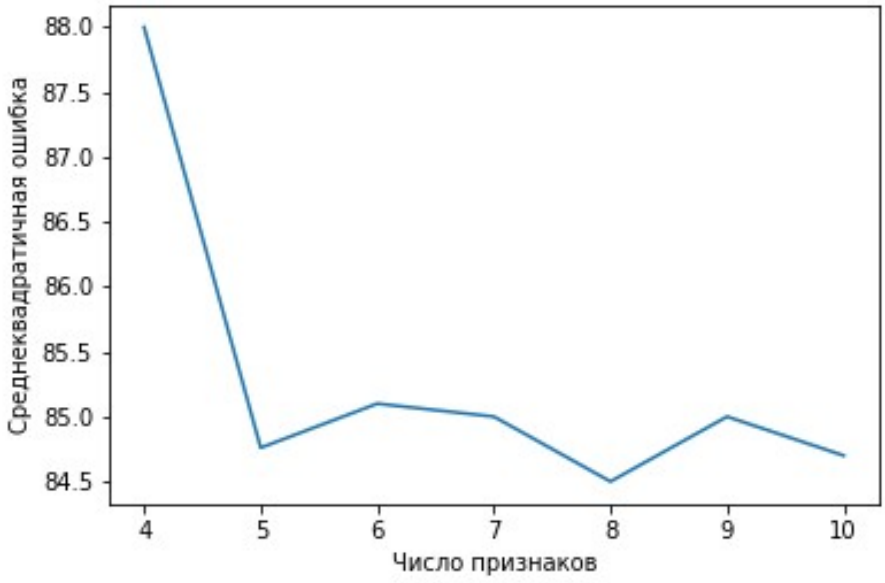


Рис.4. График зависимости среднеквадратичной ошибки от числа признаков для метода опорных векторов

Каждое поколение признаков, которое получается операторами мутации и скрещивания особей (совокупности признаков) должно давать точность не хуже, чем предыдущее поколении. Поэтому с целью проверки правильности проверки параметров эволюционного алгоритма был построен график зависимости среднеквадратичной ошибки от количества итераций в пределах одного прогона эволюционного алгоритма для метода опорных векторов. Синяя сплошная линия, которая показана на рисунке 5, показывает среднеквадратичную оценка для лучших представителей особей (признаков) каждой итерации.

В нашей реализации эволюционного алгоритма, подмножества исходных признаков материалов содержат случайно выбранные дескрипторы. При каждой итерации данные подмножества признаков используются в последующие поколения с целью улучшения точности модели, которые сохраняются там с помощь оператора отбора. Операторы скрещивания (кроссинговера) и мутации быстро точечно исследуют новое поколение, которое предположительно должно иметь большую точность. Из графика

(рис.5) нетрудно заметить, что уже после 5 итераций генетического алгоритма как средняя точность приводит первоначально к росту точности с последующим выходом на насыщение. Следовательно, параметры эволюционного алгоритма можно считать оптимальными для решения поставленной задачи.

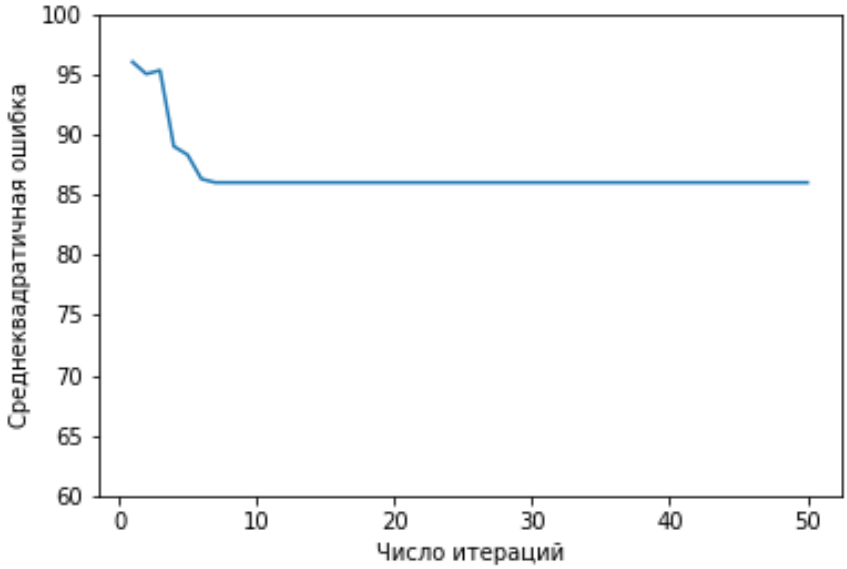


Рис.5. График зависимости среднеквадратичной ошибки в зависимости от числа итераций генетического алгоритма

Средняя точность быстро возрастают за первые несколько итераций и сходятся к оптимальному решению. Повторял вышеописанную процедуру 20 раз с целью получения оптимального решения, так как эволюционные алгоритмы по своей структуре относят к методам стохастического анализа и результат работы зависит как от случайно созданной на начальном этапе популяции, так и от типов операторов мутации и скрещивания. Целью заданной процедуры является нахождение оптимальной комбинации признаков в пределах 20 расчётов.

Результаты вышеописанной процедуры были занесены в таблицу 3. Все подмножества содержат 8 дескрипторов, которые показали наилучшую

точность на каждой из моделей машинного обучения, которые и сведены в таблицу.

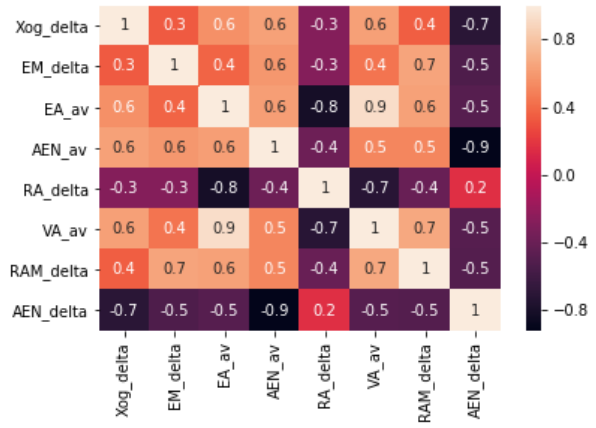
Таблица 3

Точность моделей машинного обучения

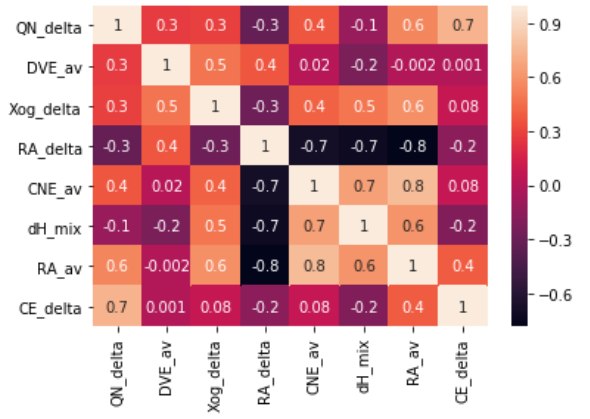
Метод отбора признаков	Модель обучения	Признаки	Ошибка
Генетический алгоритм	Линейная регрессия	$\delta(X_{og})$, $\overline{C}$, $\overline{X_{\text{abs}}}$, dH_{mix} $\overline{X_{MB}}$, $\overline{EV}$, $\delta(CNE)$, δRA	8,2
Генетический алгоритм	Метод опорных векторов	$\overline{C}$, $\overline{WF}$, $\overline{X_{og}}$, $\overline{D}$, $\overline{VA}$, $\overline{SIE}$, $\overline{AN}$, dH_{mix}	8
Генетический алгоритм	Байесовская регрессия	$\overline{X_{\text{abs}}}$, $\delta(EM)$, $\overline{EA}$, $\overline{AEN}$ δRA ,$\overline{VA}$, δ(RAM) , dH_{mix}	8
Генетический алгоритм	К ближайших соседей	δQN, $\overline{DVE}$, δX_{og}, δRA , $\overline{CNE}$, dH_{mix} ,$\overline{RA}$, δCE	7,3

Делая статистический анализ полученных признаков можно сделать вывод, что в 3-ёх моделях – присутствует энтальпия смешения, радиус атомов и электроотрицательность по Оганову. Из таблицы 2 видно, что самым оптимальным алгоритмом для решения поставленной задачи является метод ближайших соседей.

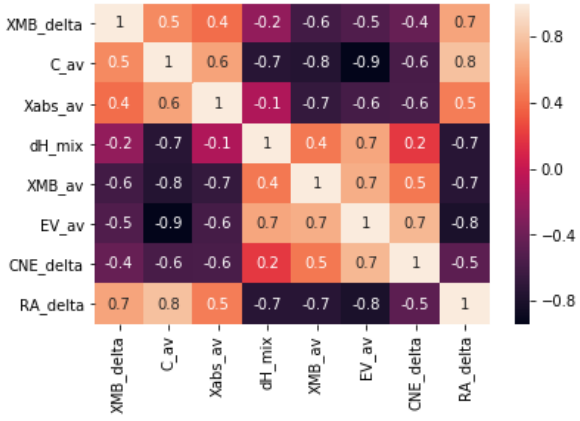
С целью выявления сильно взаимосвязанных признаков были построены матрицы корреляции для 4 алгоритмов машинного обучения, которые показаны на рис.6.



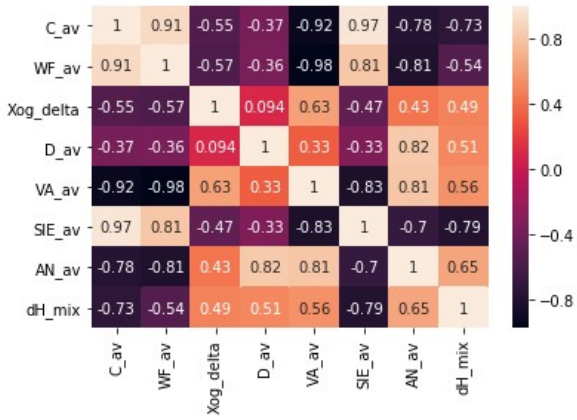
Корреляция по Байесу



Метод ближайших соседей



Линейная регрессия



Метод опорных векторов

Рис.6. Матрицы корреляций признаков

На рис.7. представлена гистограмма, на которой показана точность каждой модели машинного обучения для определённого набора подобранных признаков, которые отобраны с помощью эволюционного алгоритма.

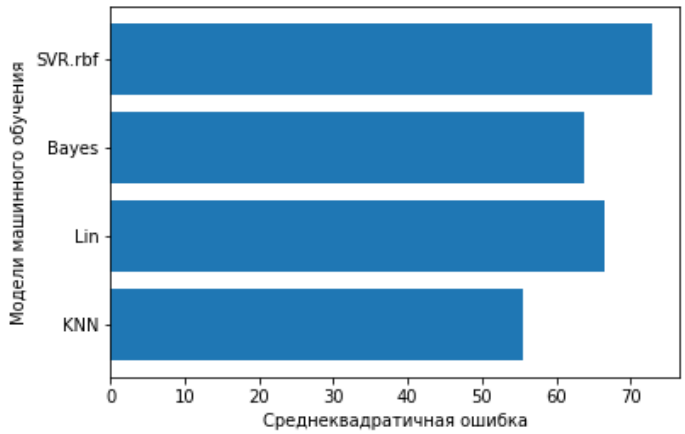


Рис.7. Гистограмма точности моделей машинного обучения

4.2 Результаты работы машинного обучения

В предыдущем подпункте в генетическом алгоритме были использованы модели машинного обучения с найденными оптимальными признаками без настройки гиперпараметров. Так большинство алгоритмов машинного обучения содержат данный тип параметров, которые должны выбираться до начала процесса тренировки модели необходимо провести анализ их влияния на точность моделей.

В нашем решении мы тренируем 2 модели и строим график зависимости точности регрессоров для тренировочного набора и во время перекрёстной проверки по мере увеличения определённого гиперпараметра. На рис.8 показан процесс оценивания модели на тестовых данных до выбора гиперпараметров для метода опорных векторов.

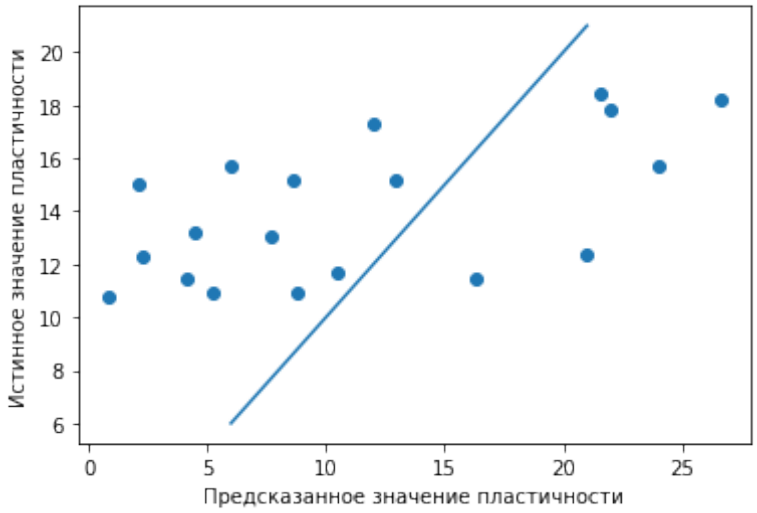


Рис.8. Сравнение предсказанных параметров с тестовыми для метода опорных векторов

К основным гиперпараметрами метода опорных векторов относят параметры C и epsilon. С целью изучения влияния их на точность модели были построены валидационные кривые.

Из валидационных кривых, изображённых на рисунках 9 и 10 видно, что с ростом значения гиперпараметра C как тренировочная оценка, так и

перекрёстно-проверочная оценка уходят на насыщение. Средняя квадратичная ошибка при перекрёстно-проверочной оценке по мере увеличения C сначала падает, достигая своего минимума, а затем увеличивается до насыщения.

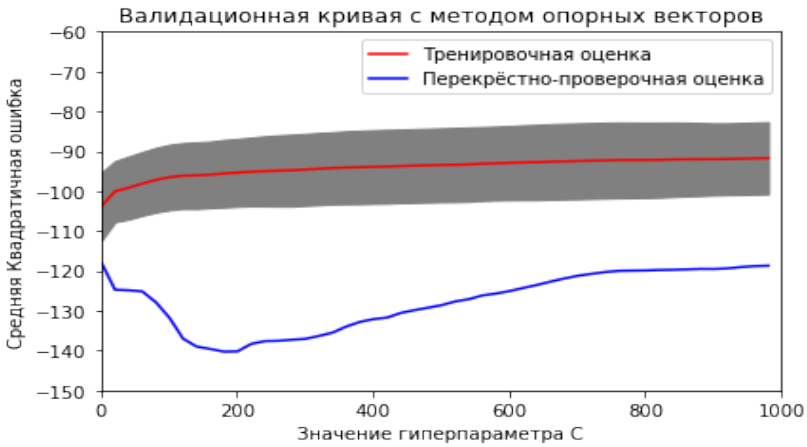


Рис.9 Зависимость точности метода опорных векторов от гиперпараметра C

Рассуждая аналогичным образом можно показать, что значение гиперпараметра epsilon будет оптимально при малых значениях, так средняя квадратичная ошибка падает при увеличении данного гиперпараметра.

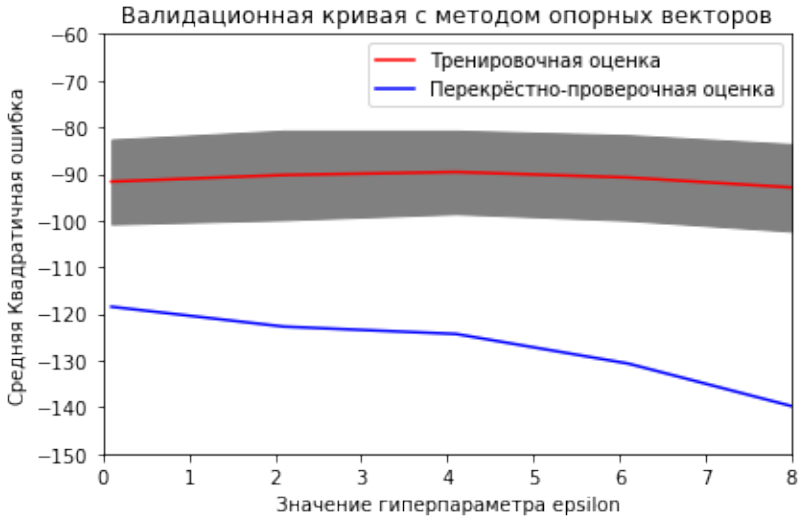


Рис.10 Зависимость точности метода опорных векторов в зависимости от гиперпараметра epsilon.

Оценим точность модели опорных векторов после процедуры назначения гиперпараметров. На рис.11 показан процесс оценивания модели на тестовых данных после выбора гиперпараметров для метода опорных векторов.

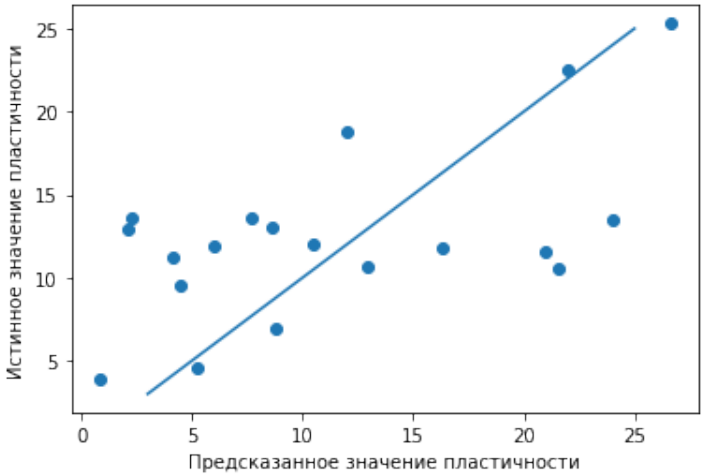


Рис.11 Сравнение предсказанных параметров с тестовыми для метода опорных векторов после отбора лучших гиперпараметров

Нетрудно заметить, что точки пространства истинное значение и предсказанное значение пластичности ближе прилегают к прямой, а значит точность модели больше. Аналогичная процедура выполнена для метода ближайших соседей.

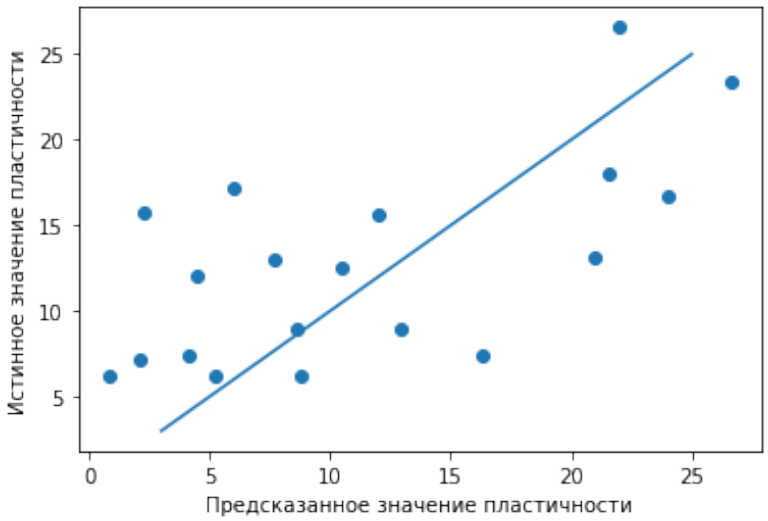


Рис.12 График разброса истинных и тестовых значений для метода ближайших соседей

Существенным гиперпараметрами, влияющим на точность модели KNR является число ближайших соседей. На рис.13 построена валидационные кривая для оценки влияния параметра n_neighbors. Видно, что при увеличении гиперпараметра точность тренировочной оценки уменьшается, а средняя квадратичная ошибка для перекрёстно-проверочной оценки уменьшается. Так как при больших n_neighbors две оценки уходят на насыщение, то за число ближайших соседей составит 8 в оптимальном случае.

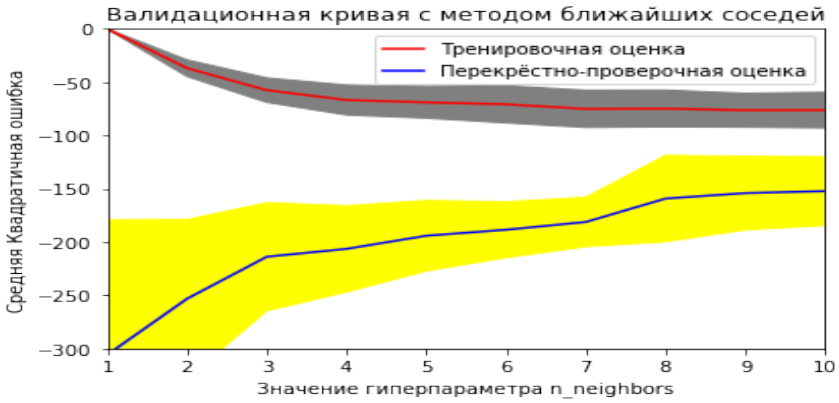


Рис.13 Зависимость точности метода ближайших соседей от числа ближайших соседей

График разброса истинных и предсказанных значений был построен для метода KNR (рис.14) с учётом параметра n_neighbors=8. Видно незначительное улучшение в точности KNR .

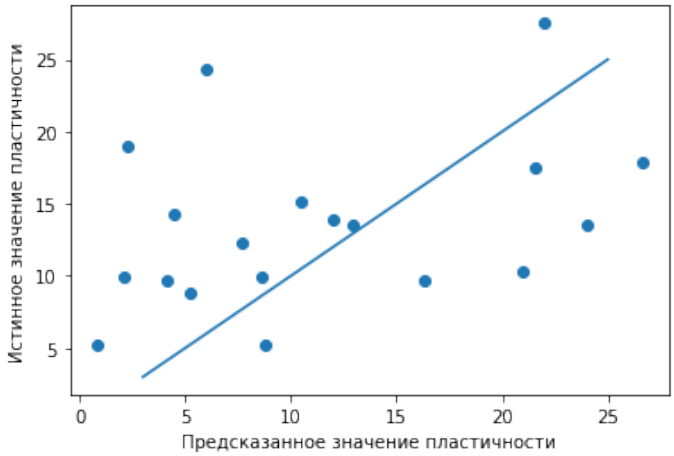


Рис.14 График разброса истинных и тестовых значений для метода ближайших соседей после выбора оптимальных гиперпараметров.

4.3 Получение сплавов и оценка точности модели

Учитывая наши найденные оптимальные признаки и модели машинного обучения был произведён расчёт, с целью тестирования модели для предсказания пластичности новых выплавленных высокоэнтропийных сплавов. Химический состав сплавов, а также результаты расчёта их пластичности приведены в таблице 3.

Таблица 3

Химический состав выплавленных сплавов

Alloys	Al	Cr	Nb	Ti	V	Zr	$\delta_{пр}$,%	δ,%
A1	13	0	10	44	28	5	-	7
A2	12	2	19	44	16	7	-	6
A3	15	11	6	37	20	11	1	8
A4	11	7	25	50	14	11	-	3
A5	14	13	23	22	26	2	12	12
A6	12	1	22	31	20	14	10	7
A7	14	3	21	34	4	24	1	8
A8	10	0	20	36	13	21	1	9

Для проверки соответствующих результатов сплавы выплавили вакуумно-дуговым методом, произвели обработку, шлифовку и полировку. Образцы испытывали на испытательной машине на сжатие. Сплавы A1, A2, A4 разрушились в упругой области. Сплавы A3, A5, A6, A7, A8 имеют пластичность около 1%,12 % ,10% ,1% и 1% соответственно.

Анализ пластичности выплавленных сплавов с полученными теоретическими значениями показал, что может предсказывать пластичности в рамках вышеприведённых стандартных отклонений. Задачу считаю выполненной, но для дальнейшего улучшения точности предсказания моделей необходим более детальный анализ. В частности, необходимо учитывать фазовый состав сплава, его микроструктуру, взаимное расположение частиц, а также плотность дислокаций.

Выводы

В главе 1 и 2 был дан обзор литературы, который посвящён высокоэнтропийными сплавам, были рассмотрены такие методы моделирования как машинное обучение и генетический алгоритм. Рассмотрены основные концепции работы машинного обучения и эволюционного алгоритма.

В главе 3 исходя из выбранной тематики работы и поставки работы были выбраны такие методики, как испытания на осадку, сбор базы данных, конструирование моделей машинного обучения и генетического алгоритма.

В главе 4 даны основные результаты. Было показано, что эволюционный алгоритм при 8 признаках показывает наилучшую точность для решения задачи по предсказанию пластичности высокоэнтропийных сплавов. В качестве моделей машинного обучения использовали метод опорных векторов, метод ближайших соседей, байесовский регрессор и линейная регрессия. Как показано самым оптимальным вариантом выбора модели является метод случайных соседей, который дал минимальную среднеквадратичную ошибку.

Анализ пластичности выплавленных сплавов с полученными теоретическими значениями показа, что может предсказывать пластичности в рамках вышеприведённых стандартных отклонений. Задачу считаю выполненной, но для дальнейшего улучшения точности предсказания моделей необходим более детальный анализ. В частности, необходимо учитывать фазовый состав сплава, его микроструктуру, взаимное расположение частиц, а также плотность дислокаций.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

[1] J.W. Yeh, et al., Formation of simple crystal structures in Cu-Co-Ni-Cr-Al-Fe-Ti-V alloys with multiprincipal metallic elements, Metall. Mater. Trans. A 35 (8) (2004) 2533–2536.

[2] C.-J. Tong, M.-R. Chen, J.-W. Yeh, S.-J. Lin, S.-K. Chen, T.-T. Shun, S.-Y. Chang, Mechanical performance of the AlxCoCrCuFeNi high-entropy alloy system with multiprincipal elements, Metall. Mater. Trans. 36 (2005) 1263e1271.

[3] Y. Zhang, Y.J. Zhou, Solid solution formation criteria for high entropy alloys 561 (2007) 1337e1339.

[4] C. Li, J. Li, M. Zhao, Q. Jiang, Effect of alloying elements on microstructure and properties of multiprincipal elements high-entropy alloys, J. Alloy. Comp. 475 (2009) 752e757.

[5] Y. Zhang, Mechanical Properties and Structures of High Entropy Alloys and Bulk Metallic Glasses Composites, vol. 654, 2010, pp. 1058e1061.

[6] H. Diao, R. Feng, K. Dahmen, P. Liaw, Fundamental deformation behavior in high-entropy alloys: an overview, Curr. Opin. Solid State Mater. Sci. 21 (5) (2017) 252e266.

[7] Yan Zhang , Cheng Wen , Changxin Wang , Stoichko Antonov , Dezhen Xue , Yang Bai , Yanjing Su Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models 185(2019)

[8] Гилл Ф. Практическая оптимизация / Ф. Гилл, У. Мюррей, М. Райт; пер. с англ. – М.: Мир, 1985. – 509 с

[9] Панченко Т.В. Генетические алгоритмы: учеб.-метод. пособие / Т.В. Панченко; под ред. Ю.Ю. Тарасевича. – Астрахань: Астраханский ун-т, 2006. – 89 с.

[10] Вороновский Г. К., Махотило К. В., Петрашев С. Н., Сергеев С. А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Харьков: Основа, 1997.
[11] Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. Journal of Chemical Physics 124, 244704 (2006).
[12] J. Fürnkranz, D. Gamberger, N. Lavrac, Machine Learning and Data Mining, Springer, Berlin Heidelberg, 2012. 1–17
[13] Y. Zhang, Y. Zhou, J. Lin, G. Chen, P. Liaw, Solid-solution phase formation rules for multi-component alloys, Adv. Eng. Mater. 10 (6) (2008) 534–538.
[14] M. Poletti, L. Battezzati, Electronic and thermodynamic criteria for the occurrence of high entropy alloys in metallic systems, Acta Materialia 75 (2014) 297–306.
[15] Z. Wang, Y. Huang, Y. Yang, J. Wang, C. Liu, Atomic-size effect and solid solubility of multicomponent alloys, Scripta Materialia. 94 (2015) 28–31.
[16] S. Gorsse , M.H. Nguyen , O.N. Senkov , D.B. Miracle Corrigendum to database on the mechanical properties of high entropy alloys and complex concentrated alloys, data in brief 21 (2018) 2664-2678.