

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
электроники, фотоники и
молекулярной физики**

В.В. Иванов

	Рабочая программа дисциплины (модуля)
по дисциплине:	Машинное обучение в моделировании материалов
по направлению:	Прикладные математика и физика
профиль подготовки:	Физика перспективных технологий: альтернативная энергетика, научное программирование и функциональные материалы Физтех-школа Электроники, Фотоники и Молекулярной Физики кафедра физики высокотемпературных процессов
курс:	4
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 8 (весенний) - Экзамен

Аудиторных часов: 30 всего, в том числе:

лекции: 15 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Подготовка к экзамену: 30 час.

Всего часов: 90, всего зач. ед.: 2

Количество контрольных работ, заданий: 2

Программу составил: А.В. Шапеев, канд. физ.-мат. наук

Программа обсуждена на заседании кафедры физики высокотемпературных процессов 27.05.2021

Аннотация

Курс «Машинное обучение в моделировании материалов» предназначен для ознакомления обучающихся с основными способами применения алгоритмов машинного обучения в задачах моделирования материалов вообще, и задачах имитационного атомистического моделирования в частности

Задачи курса: формирование понимания о представлении данных о материалах для их обработки алгоритмами машинного обучения и получение практических навыков в применении алгоритмов машинного обучения для предсказания свойств материалов

По результатам освоения курса студент должен:

знать:

- основы построения моделей машинного обучения для предсказания свойств материалов и их отличия от классических моделей;
- задачи, к которым применимы методы машинного обучения;
- принципы применения алгоритмов машинного обучения к таким задачам и оценки их качества.

уметь:

- представлять данные о материале в виде, пригодном для применения алгоритмов машинного обучения;
- применять алгоритмы машинного обучения для предсказания свойств материалов.

владеть:

- практическими навыками использования алгоритмов машинного обучения для предсказания свойств материалов;
- практическими навыками проведения имитационного моделирования с помощью моделей машинного обучения.

Основное содержание курса изложено в следующих разделах:

- 1 Введение
- 2 Модели типа состав-свойство
- 3 Модели типа структура-свойство
- 4 Свойства упругодеформированных материалов
- 5 Машинно-обучаемые межатомные потенциалы
- 6 Активное обучение межатомных потенциалов
- 7 «Решеточные» межатомные потенциалы
- 8 Предсказание погрешности квантово-механических вычислений
- 9 Алгоритмы машинного обучения для построения фазовых диаграмм

1. Цели и задачи

Цель дисциплины

Ознакомление обучающихся с основными способами применения алгоритмов машинного обучения в задачах моделирования материалов вообще, и задачах имитационного атомистического моделирования в частности.

Задачи дисциплины

Формирование понимания о представлении данных о материалах для их обработки алгоритмами машинного обучения и получение практических навыков в применении алгоритмов машинного обучения для предсказания свойств материалов.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
ПК-1 Способен планировать и проводить научные эксперименты (в избранной предметной области) и (или) теоретические (аналитические и имитационные) исследования	ПК-1.1 Владеет фундаментальными понятиями, законами и теориями современной физики

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основы построения моделей машинного обучения для предсказания свойств материалов и их отличия от классических моделей;
- задачи, к которым применимы методы машинного обучения;
- принципы применения алгоритмов машинного обучения к таким задачам и оценки их качества.

уметь:

- представлять данные о материале в виде, пригодном для применения алгоритмов машинного обучения;
- применять алгоритмы машинного обучения для предсказания свойств материалов.

владеть:

- практическими навыками использования алгоритмов машинного обучения для предсказания свойств материалов;
- практическими навыками проведения имитационного моделирования с помощью моделей машинного обучения.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение	1	1		
2	Модели типа "состав-свойство"	2	2		2
3	Модели типа "структура-свойство"	2	2		6
4	Свойства упругодеформированных материалов	2	2		6
5	Машинно-обучаемые межатомные потенциалы	2	2		6
6	Активное обучение межатомных потенциалов	2	2		6

7	«Решеточные» межатомные потенциалы	2	2		4
8	Предсказание погрешности квантово-механических вычислений	1	1		
9	Алгоритмы машинного обучения для построения фазовых диаграмм	1	1		
Итого часов		15	15		30
Подготовка к экзамену		30 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 8 (Весенний)

1. Введение

Введение в машинное обучение. Признаки и метки (features and labels). Обучение с учителем, задача регрессии. Линейная регрессия, ядерная регрессия, нейросети. Обучающая выборка. Функция потерь. Валидационная выборка. Погрешности обучения и валидации. Оценка неопределенности предсказаний величин. Байесовский подход и гауссовские процессы.

Основы физики твердого тела. Преобразования Фурье и Блоха. Решетка Браве.

Основы атомистического моделирования. Атомистические системы. Модели межатомного взаимодействия. Статистический ансамбль. Термостат. Молекулярная динамика. Эквilibрирование структуры. Метод Монте-Карло. Расчет термодинамических свойств.

2. Модели типа "состав-свойство"

Примеры моделей типа состав-свойство. Представление данных. Решение типичной задачи.

3. Модели типа "структура-свойство"

Примеры моделей типа состав-свойство. Представление данных. Решение типичной задачи.

4. Свойства упругодеформированных материалов

Тензор деформации, деформация кристаллической решетки, упругая и неупругая деформация. Полносвязные и сверточные нейронные сети. Задача классификации. Предсказание электронных свойств деформированного материала.

5. Машинно-обучаемые межатомные потенциалы

(Полу-)эмпирические потенциалы межатомного взаимодействия. Локальные и нелокальные межатомные потенциалы. Атомное окружение. Реактивные и нереактивные межатомные потенциалы. Машинно-обучаемые межатомные потенциалы, их основные виды. Moment Tensor Potentials (МТП-потенциалы). Пассивное обучение. Расчет констант упругости с помощью МТП-потенциалов и оценка неопределенности их предсказания.

6. Активное обучение межатомных потенциалов

Отличие пассивного и активного обучения. Стратегия отбора конфигураций по ансамблю потенциалов. Критерий D-оптимальности. Алгоритм MaxVol (нахождение подматрицы максимального объема). Активное множество. Интерполяция и экстраполяция. Оценка степени экстраполяции конфигурации. Пример активного обучения МТП-потенциала в ходе молекулярно-динамического моделирования.

7. «Решеточные» межатомные потенциалы

Модели межатомного взаимодействия «на решётке». Метод кластерного разложения. Малоранговые потенциалы “на решётке”, Low-rank potentials (LRP-потенциалы). Разложение в «тензорной поезд». Кэширование данных потенциала. Исследование стабильности многокомпонентных сплавов методом канонического Монте-Карло на основе LRP.

8. Предсказание погрешности квантово-механических вычислений

Данные квантово-механических расчетов как Гауссовские процессы. Значение и неопределенность предсказания энергии как функции параметра решетки. Дифференцирование гауссовского процесса и производные свойства в равновесном состоянии.

9. Алгоритмы машинного обучения для построения фазовых диаграмм

Построение фазовых диаграмм по функциям свободной энергии фаз. Представление свободной энергии как Гауссовского процесса. Значение и неопределенность предсказания свободной энергии как функции термодинамических переменных. Значение и неопределенность элементов фазовой диаграммы таких как тройная точка, критическая точка, кривые и области сосуществования.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

6. Перечень рекомендуемой литературы

Основная литература

1. Метод Монте-Карло в статистической термодинамике / В.М. Замалин, Г.Э. Норман, В.С. Филинов, Москва, Наука, 1977. 228 с.
2. Введение в физику твердого тела / Ч. Киттель, Пер. с англ. Физматгиз, 1963.
3. Физико-химические основы материаловедения / Г. Готтштайн, под ред. В.П. Зломанова, 2009.

Дополнительная литература

1. Machine Learning Meets Quantum Physics / K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda, K.-R. Müller, Springer, 2020.
2. Pattern recognition and machine learning / C.M. Bishop. Springer, 2006.
3. Gaussian Processes for Machine Learning / C. E. Rasmussen and C. K. I. Williams. MIT Press, 2006.
4. Understanding Molecular Simulation From Algorithms to Applications / D. Frenkel, B. Smit, Computational Science Series 1, 2001.
5. Introduction to Statistical Learning / G. James, D. Witten, T. Hastie, R. Tibshirani. Springer, 2013.
6. Deep elastic strain engineering of bandgap through machine learning / Z. Shi, E. Tsybalov, M. Dao, S. Suresh, A.V. Shapeev, J. Li, Proceedings of the National Academy of Sciences 116 (10), 4117-4122, 2019.
7. Accelerating high-throughput searches for new alloys with active learning of interatomic potentials / K. Gubaev, E.V. Podryabinkin, G.L.W. Hart, A.V. Shapeev, Computational Materials Science 156, 148-156, 2019.
8. Active learning of linearly parametrized interatomic potentials / E.V. Podryabinkin, A.V. Shapeev, Computational Materials Science 140, 171-180, 2017.
9. Accurate representation of formation energies of crystalline alloys with many components / A.V. Shapeev, Computational Materials Science 139, 26-30, 2017.
10. Short-range order in face-centered cubic VCoNi alloys / T. Kostiuchenko, A.V. Ruban, J. Neugebauer, A.V. Shapeev, F. Körmann, Physical Review Materials 4 (11), 113802, 2020.

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. Новиков И. С., Шапеев А. В. Практическое руководство по использованию программного пакета для работы с машинно-обучаемыми потенциалами МТР.
<https://gitlab.com/ashapeev/mlip-2-tutorials/-/wikis/>.
2. The MLIP package: Moment tensor potentials with MPI and active learning / I.S. Novikov, K. Gubaev, E.V. Podryabinkin, A.V. Shapeev, Machine Learning: Science and Technology 2 (2), 025002, 2020.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Для занятий может потребоваться следующее программное обеспечение:

Интернет-браузер, MS Word, MS Power Point, LaTeX, Adobe Reader, Anaconda Python distribution, Компилятор или среда разработки языка C++.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения и понятия, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Физика перспективных технологий: альтернативная энергетика, научное программирование и функциональные материалы Физтех-школа Электроники, Фотоники и Молекулярной Физики кафедра физики высокотемпературных процессов
курс:	<u>4</u>
квалификация:	бакалавр
Семестр, формы промежуточной аттестации: 8 (весенний) - Экзамен	
Разработчик:	А.В. Шапеев, канд. физ.-мат. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
ПК-1 Способен планировать и проводить научные эксперименты (в избранной предметной области) и (или) теоретические (аналитические и имитационные) исследования	ПК-1.1 Владеет фундаментальными понятиями, законами и теориями современной физики

2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение в моделировании материалов» обучающийся должен:

знать:

- основы построения моделей машинного обучения для предсказания свойств материалов и их отличия от классических моделей;
- задачи, к которым применимы методы машинного обучения;
- принципы применения алгоритмов машинного обучения к таким задачам и оценки их качества.

уметь:

- представлять данные о материале в виде, пригодном для применения алгоритмов машинного обучения;
- применять алгоритмы машинного обучения для предсказания свойств материалов.

владеть:

- практическими навыками использования алгоритмов машинного обучения для предсказания свойств материалов;
- практическими навыками проведения имитационного моделирования с помощью моделей машинного обучения.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

С целью контроля освоения обучающимися учебного материала проводится устный опрос в начале занятия по теме прошлой лекции или в конце занятия по пройденной теме.

В течение семестра студентами выполняются и сдаются преподавателю два практических задания.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Вопросы к экзамену:

1. Построить модель машинного обучения по обучающей и валидационной выборке для предсказания соответствующего свойства материалов.
2. Для задачи регрессии построить зависимость ошибки обучения полносвязной нейронной сети от глубины и ширины сети, найти оптимальные значения ее гиперпараметров. Провести валидацию полученной нейронной сети на отложенной выборке.

3. Провести ресемплирование обучающей выборки для задачи классификации с несбалансированными классами. Обучить полносвязную нейронную сеть для задачи многоклассовой классификации.
4. Исследовать зависимость ошибки обучения МТР-потенциала от его уровня, а также от числа его параметров, найти оптимальный уровень и число параметров. Провести кросс-валидацию МТР-потенциала с оптимальным числом параметров.
5. Провести оценку неопределенности предсказания энергий и сил с помощью МТР-потенциалов.
6. Провести активное обучение МТР-потенциалов с различным числом параметров в ходе молекулярной динамики. Исследовать зависимость величины базы данных от числа параметров.
7. Провести молекулярно-динамическое моделирование при разных температурах (от комнатной до температуры плавления материала), в ходе которого активно обучить МТР-потенциал. Сравнить количество конфигураций в итоговых базах данных при разных температурах.
8. Найти оптимальный ранг LRP-потенциала по обучающей и валидационной выборкам.
9. Обучить ансамбль LRP потенциалов на данных обучающей и валидационной выборках с заданным рангом. По результатам моделирования каноническим методом Монте-Карло определить температуры соответствующие максимальной неопределенности (погрешности) результатов.
10. Восстановить тестовую гладкую функцию и ее производную с помощью гауссовского процесса по “зашумленным” значениям ее производной в некоторых точках.
11. Рассчитать критическую точку модельной системы с помощью гауссовского процесса, оценить неопределенность ее предсказания.

Примеры экзаменационных билетов:

Пример 1:

1. Провести активное обучение МТР-потенциалов с различным числом параметров в ходе молекулярной динамики. Исследовать зависимость величины базы данных от числа параметров.
2. Провести молекулярно-динамическое моделирование при разных температурах (от комнатной до температуры плавления материала), в ходе которого активно обучить МТР-потенциал. Сравнить количество конфигураций в итоговых базах данных при разных температурах.

Пример 2:

1. Для задачи регрессии построить зависимость ошибки обучения полносвязной нейронной сети от глубины и ширины сети, найти оптимальные значения ее гиперпараметров. Провести валидацию полученной нейронной сети на отложенной выборке.
2. Провести ресемплирование обучающей выборки для задачи классификации с несбалансированными классами. Обучить полносвязную нейронную сеть для задачи многоклассовой классификации.

Критерии оценивания

Оценка отлично 10 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 9 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 8 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо 7 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо 6 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо 5 баллов - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно 4 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно 3 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно 2 балла - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно 1 балл - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Прием экзамена проводится по билетам. В каждом билете представлено два теоретических вопроса и одна задача. При проведении устного экзамена обучающемуся предоставляется 40 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.