



Магистерская программа «Машинное обучение и анализ больших данных»

**Кафедра системного
программирования ЮУрГУ**

<https://sp.susu.ru/>

Карта профильных курсов

Семестр	Машинное обучение	Обработка данных	Интеллектуальный анализ данных	Высокопроизводительные вычисления
I	<ul style="list-style-type: none">• Основы машинного обучения	<ul style="list-style-type: none">• Теоретические основы разработки систем управления большими данными	<ul style="list-style-type: none">• Язык Python для анализа данных	
II	<ul style="list-style-type: none">• Нейронные сети	<ul style="list-style-type: none">• Хранение и обработка больших данных• Администрирование и оптимизация корпоративных СУБД		<ul style="list-style-type: none">• Архитектура распределенных программных систем• Технологии параллельного программирования
III	<ul style="list-style-type: none">• Поиск, обработка и распознавание аудио, видео и графической информации	<ul style="list-style-type: none">• NoSQL системы	<ul style="list-style-type: none">• Интеллектуальный анализ больших данных	<ul style="list-style-type: none">• Технологии параллельного программирования• Квантовые вычисления• Облачные вычисления
IV	<ul style="list-style-type: none">• Обработка естественного языка			

Магистерская программа «Машинное обучение и анализ больших данных»

Направление подготовки

- ▶ 02.04.02 «Фундаментальная информатика и информационные технологии»

Аннотация

- ▶ Машинное обучение и анализ больших данных является одним из главных трендов в IT-индустрии в настоящее время, именно поэтому профессионалы в этой области очень востребованы на современном рынке труда. Магистерская программа «Машинное обучение и анализ больших данных» готовит специалистов высшей квалификации в области искусственного интеллекта и анализа больших данных. В процессе обучения студенты осваивают современные методы машинного обучения, разработки нейронных сетей, интеллектуального анализа данных. Теоретическое обучение сочетается с большим количеством практических занятий. Выпускник магистратуры подготовлен к деятельности, требующей использования самых современных методов машинного обучения и анализа данных.

Ключевые учебные курсы

- ▶ Математические основы машинного обучения
- ▶ Машинное обучение
- ▶ Применение искусственных нейронных сетей
- ▶ Хранение и обработка больших данных
- ▶ Современные методы анализа больших данных
- ▶ Теоретические основы разработки систем управления большими данными
- ▶ Python для решения задач интеллектуального анализа данных
- ▶ Современные NoSQL-системы
- ▶ Высокопроизводительные вычисления



Руководитель программы:
проф., д.ф.-м.н.
Л.Б. Соколинский
leonid.sokolinsky@susu.ru

ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ



Кандидат физ.-мат. наук
О.Л. Ибряева
ibriaevaol@susu.ru

Формат

- Место в учебном плане: 1 семестр
- Лекции: 32 часов
- Практика: 16 часов

Краткая аннотация

- Понимание многих алгоритмов машинного обучения немыслимо без уверенного владения базовым математическим аппаратом. Цель данной дисциплины – изложить математические и концептуальные основы линейной алгебры, теории вероятностей и теории информации, численных расчетов, необходимые для успешного выполнения исследовательских проектов в области глубокого обучения. В ходе курса рассматриваются случайные величины, их распределения, характеристики и структурные вероятностные модели. Из основ линейной алгебры выводится один из алгоритмов машинного обучения – метод главных компонент (principal components analysis). Обсуждается оптимизация градиентным методом и такие функции, как логистическая сигмоида, softplus, softmax в контексте их будущего применения в проектировании нейронных сетей.

Нейронные сети



Формат

- Место в учебном плане: 2 семестр
- Лекции: 32 часов
- Практика: 16 часов

Краткая аннотация

- Машинное обучение является магистральным направлением в области современных информационных технологий. Цель настоящего курса – дать студентам теоретические знания и практические навыки по использованию методов машинного обучения для решения практических задач, связанных с использованием методов искусственного интеллекта. В ходе обучения студенты знакомятся с основными классами и методами машинного обучения. Особое внимание уделяется строению биологического нейрона человеческого мозга. Рассматривается математическая модель нейрона и глубокие нейронные сети. Даются основные методы обучения нейронных сетей: градиентный спуск и метод обратного распространения ошибки. Излагаются принципы организации сверточных нейронных сетей
- На практических занятиях студенты применяют полученные знания к созданию на языке Python искусственной нейронной сети, распознающей рукописные цифры

Доктор физ.-мат. наук,
профессор
Л.Б. Соколинский
leonid.sokolinsky@susu.ru



Ст. преподаватель
Р.С. Федянина
raisa.fedyanina@susu.ru

Поиск, обработка и распознавание аудио, видео и графической информации



Ст. преподаватель
К.Ю. Никольская
nikolskaiaki@susu.ru

Формат

- Место в учебном плане: 3 семестр
- Лекции: 32 часов
- Практика: 32 часов

Краткая аннотация

- Цель настоящего курса – дать студентам теоретические знания и практические навыки по практическому применению глубоких нейронных сетей. В ходе обучения студенты знакомятся существующими библиотеками для обучения глубоких нейронных сетей: TensorFlow, Keras, CNTK, PyTorch, Caffe, а также учатся применять данные библиотеки для решения практических задач обработки и распознавания аудио, видео и графической информации. В рамках курса студенты используют бесплатную облачную платформу Google Colaboratory с мощными GPU для обучения глубоких нейронных сетей
- На практических занятиях студенты осваивают практическое применение полученных знаний для создания проектов на языке Python: распознавание человека по походке, определение психотипа по Юнгу (анализ мимики, жестов), задачи распознавания и классификации автомобильного трафика и пешеходов по видеопотоку, анализ медицинских изображений и др.

Обработка естественного языка



Кандидат физ.-мат. наук
С.А. Иванов
saivanov@susu.ru

Формат

- Место в учебном плане: 4 семестр
- Лекции: 16 часов
- Практика: 32 часов

Краткая аннотация

- Современное состояние развития области машинного обучения позволяет находить применение в различных областях человеческой деятельности. Цель данного курса – дать студентам теоретические и практические навыки по обработке естественного языка методами машинного обучения, включая нейронные сети. В ходе обучения студенты познакомятся с основными задачами обработки естественного языка: морфологический анализ, задачи информационного поиска, автоматическое реферирование, синтаксический анализ, анализ тональности, извлечение информации, вопросно-ответный поиск, лексическая семантика, машинный перевод.
- На практических занятиях студенты осваивают практическое применение полученных знаний для решения описанных выше задач на языке Python.

Язык Python для анализа данных



Ст. преподаватель
Р.С. Федянина
raisa.fedianina@susu.ru

Формат

- Место в учебном плане: 1 семестр
- Лекции: 16 часов
- Практика: 32 часов

Краткая аннотация

- Язык программирования Python в последнее время все чаще используется для анализа данных, как в науке, так и коммерческой сфере. Этому способствует простота языка, а также большое разнообразие открытых библиотек. В рамках курса будут рассмотрены основы Python для работы с данными. Основные библиотеки для анализа данных: numpy, scipy, pandas. Статистика в Python. Предобработка данных и feature selection. Визуализация данных в Python. Обзор библиотек: matplotlib, pandas, seaborn, plotly, bokeh. Базовые типы визуализаций: графики, столбчатые диаграммы, гистограммы, точечные диаграммы (scatter plots). Сбор и подготовка данных. Работа с динамическими сайтами, со структурированными данными, с API различных сервисов.

Хранение и обработка больших данных



Ст. преподаватель
Е.Р. Вальковская
(планируется)

Формат

- Место в учебном плане: 2 семестр
- Лекции: 32 часов
- Практика: 32 часов

Краткая аннотация

- Теория и практика работы с большими объемами данных на учебном Hadoop-кластере с использованием Hadoop MapReduce (Java), Apache Spark (Scala) и Apache Hive (SQL). Анализ данных с помощью средств Apache Hadoop и Apache Spark.

Интеллектуальный анализ больших данных



Кандидат физ.-мат. наук,
доцент
М.Л. Цымблер
mzym@susu.ru

Формат

- Место в учебном плане: 3 семестр
- Лекции: 32 часов
- Практика: 32 часов

Краткая аннотация

- Курс включает в себя следующие разделы. Понятие интеллектуального анализа данных. Технологический цикл аналитической обработки данных. Основные задачи анализа данных: поиск шаблонов (ассоциативных правил), классификация, кластеризация. Основные алгоритмы поиска частых наборов: Apriori, Eclat, FP-Growth. Меры полезности шаблонов: support, confidence, lift и др. Поиск шаблонов в Больших данных. Основные методы классификации: деревья решений, Байесовская классификация, правила «если-то». Меры оценки качества классификации. Методы повышения точности классификации: bagging, boosting, random forests. Основные методы кластеризации: разделительная кластеризация (алгоритмы k-Means, k-Medoids), иерархическая кластеризация (алгоритмы AGNES, DIANA), плотностная кластеризация (алгоритм DBSCAN). Меры оценки качества кластеризации. Кластеризация данных большой размерности. Кластеризация графов и социальных сетей.



Преподаватель
Я.А. Краева
kraevaya@susu.ru

Архитектура распределенных программных систем



Кандидат физ.-мат. наук,
доцент
Г.И. Радченко
gleb.radchenko@susu.ru

Формат

- Место в учебном плане: 2 семестр
- Лекции: 32 часов
- Практика: 16 часов

Краткая аннотация

•Целью курса является изучение основных концепций разработки распределенных программных систем и практическое освоение методов разработки приложений для таких систем. В ходе изучения данной дисциплины студенты знакомятся с основными подходами к классификации систем с общей и распределенной памятью, изучат историю становления вычислительных систем, ключевые методы классификации таких систем. Будут разобраны ключевые аспекты протоколов организации связи с таких системах, включая как протоколы сетевого и транспортного уровня (TCP и UDP), так и прикладные протоколы, включая протоколы сериализации и представления данных (XML, JSON, Google Protocol Buffers и др.). Основная часть курса посвящена рассмотрению и анализу наиболее распространенных концепций организации архитектур распределенных приложений. Рассматриваются такие подходы как использование методов удаленного вызова процедур (RPC) и удаленного вызова методов (RMI), многослойные клиент-серверные технологии, технологии и алгоритмы организации одноранговых (Peer-to-peer) систем, технологии организации связи с использованием очередей сообщений, облачные вычисления. Подробно рассматривается парадигма сервис-ориентированной архитектуры (SOA) приложений, включая RPC API, API сообщений, API ресурсов (REST) и графовый API. Обсуждаются проблемы проектирования распределенных вычислительных систем, включая проблемы достижения консенсуса, согласованности и доступности данных (CAP-теорема).

Технологии параллельного программирования



Кандидат физ.-мат. наук
Т.Ю. Маковецкая
tatyana.makovetskaya@susu.ru

Формат

- Место в учебном плане: 2,3 семестр
- Лекции: 32 часов
- Практика: 32 часов

Краткая аннотация

- Большинство актуальных вычислительных задачи науки и промышленности имеют большой объем и требуют длительного времени выполнения. На текущий момент основным способом сокращения времени вычислений является распараллеливание программ и выполнение их на многопроцессорной технике. Цель настоящего курса – дать студентам теоретические знания в области высокопроизводительных вычислений и практические навыки создания параллельных программ. В ходе обучения студенты знакомятся с существующими архитектурами параллельных вычислительных систем и способами оценки их производительности. Особое внимание уделяется влиянию используемой архитектуры на структуру программ и алгоритмов решения задач. Излагаются принципы разработки параллельных алгоритмов и методы анализа их трудоемкости, рассматриваются примеры параллельных алгоритмов решения стандартных математических задач. Дается классификация существующих инструментов для создания параллельных программ, подробно рассматриваются широко распространенные технологии разработки параллельных программ MPI и OpenMP.
- На практических занятиях студенты применяют полученные знания к созданию параллельных программ для различных аппаратных архитектур.

Квантовые вычисления



Кандидат физ.-мат. наук,
доцент
И.И. Клебанов
klebanovii@susu.ru

Формат

- Место в учебном плане: 3 семестр
- Лекции: 32 часов
- Практика: 32 часов

Краткая аннотация

- Квантовые вычисления – новое направление междисциплинарных исследований на стыке теоретической информатики и квантовой механики. Цель курса – познакомить студентов с основами теории квантовых вычислений и простейшими практическими приемами квантового программирования. Рассматриваются история идей и достижений в области квантовой информатики, основы квантовой механики для инженеров-программистов, теория квантовых алгоритмов, квантовой информации, концептуальные основы построения квантового компьютера, подходы к физической реализации квантового компьютера и перспективы построения полноценного квантового компьютера в обозримом будущем. Рассматриваются основы теории квантовых нейронных сетей.

Облачные вычисления



Кандидат физ.-мат. наук,
доцент
Г.И. Радченко
gleb.radchenko@susu.ru

Формат

- Место в учебном плане: 3 семестр
- Лекции: 32 часов
- Практика: 16 часов

Краткая аннотация

- Целью дисциплины является изучение основных концепций организации облачных вычислений, а также методов разработки приложений для облачных систем. В ходе изучения дисциплины студенты знакомятся с понятием облачных вычислений, историей их появления и развития (начиная с концепции коммунальных вычислений, заканчивая «Облаками 2.0»). Рассматривается архитектура облачных вычислительных систем, уровни предоставления облачных услуг (Infrastructure-as-a-Service, Platform-as-a-Service, Software-as-a-Service). Изучается архитектура как частных (VMware vCloud, Apache CloudStack, OpenStack и др.) так и публичных (IBM Bluemix, Amazon Web Services, Microsoft Azure и др) облачных систем. Подробно рассматриваются основные технологии виртуализации и контейнеризации, лежащие в основе облачных систем. Изучается платформа Docker для управления доставкой и развертывания облачных приложений. Изучаются механизмы оркестрации облачных сервисов, включая такие системы как Docker Swarm и Kubernetes. Изучается концепция микросервисной архитектуры, методы и подходы к проектированию облачных приложений на основе микросервисов, основные паттерны микросервисных приложений.

Теоретические основы разработки систем управления большими данными



Доктор физ.-мат. наук,
профессор
Л.Б. Соколинский
leonid.sokolinsky@susu.ru

Формат

- Место в учебном плане: 1 семестр
- Лекции: 32 часов

Краткая аннотация

- Цифровизация всех сфер человеческой жизнедеятельности привела к появлению проблемы больших данных, заключающейся в том, что многообразие, объем и скорость прироста цифровых данных многократно превосходят возможности классических систем управления базами данных (СУБД). Цель настоящего курса – познакомить студентов с основными принципами и методами обработки больших данных. Рассматриваются теоретические основы создания реляционных СУБД: логическая оптимизация, В-деревья, оценка стоимости реляционных операций, статистические характеристики данных, реализация исполнителя запросов, алгоритмы соединения. Дается модель распределенных вычислений MapReduce. Обсуждаются современные методы обработки и анализа больших данных

Администрирование и оптимизация корпоративных СУБД



Кандидат физ.-мат. наук,
доцент
А.Т. Латипова
latipovaat@susu.ru

Формат

- Место в учебном плане: 2 семестр
- Лекции: 32 часов
- Практика: 32 часов

Краткая аннотация

- Основной целью курса является изучение реляционных баз данных и хранилищ данных. В рамках курса рассматриваются основы реляционных баз данных, дается обширный практикум по SQL и хранилищам данных (OLAP, ROLAP), оптимизация SQL запросов. Студенты получают навыки создания, настройки и подключения к корпоративным СУБД. Также изучаются такие вопросы, как стратегии страхового копирования и восстановления БД, системные БД, разработка сервисных триггеров, оптимизация работы сервера БД
- В рамках курса предусмотрены практические задания по настройке учетных записей к серверу, пользователей БД.

NoSQL-системы



Кандидат физ.-мат. наук
Е.В. Иванова
elena.ivanova@susu.ru



Ст. преподаватель
Н.С. Силкина
silkinans@susu.ru

Формат

- Место в учебном плане: 3 семестр
- Лекции: 16 часов
- Практика: 32 часов

Краткая аннотация

- Цель настоящего курса – дать студентам теоретические знания и практические навыки по использованию нереляционных систем баз данных. В ходе обучения студенты знакомятся с особенностями появления и развития NoSQL-систем. Особое внимание уделяется классам NoSQL хранилищ и области их применения.
- На практических занятиях студенты применяют полученные знания к созданию модельного клиентского приложения для NoSQL СУБД.

Спасибо за внимание!