

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

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

	Рабочая программа дисциплины (модуля)
по дисциплине:	Глубокое обучение в обработке естественного языка
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Физтех-школа Прикладной Математики и Информатики кафедра дискретной математики
курс:	3
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 75 час.

Всего часов: 135, всего зач. ед.: 3

Количество контрольных работ, заданий: 1

Программу составил: А.А. Сорокин, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры дискретной математики 04.03.2020

Аннотация

В последние годы нейронные сети под новым именем глубокого обучения (deep learning) показали свою эффективность в разных задачах, известность методы глубокого обучения получили после победы нейронных сетей в соревновании ImageNet. Но менее известен тот факт, что методы deep learning побили state-of-the-art результаты на тот момент в другой области раньше - в области распознавания речи еще в 2009 году. В настоящий момент deep learning проникает во все области, связанные с машинным обучением. Где-то его победа уже неоспорима (например, в области компьютерного зрения), а где-то, как в области обработки естественного языка, наступление пока только в разгаре.

Курс познакомит слушателя с современными подходами в области обработки естественного языка, а также с методами deep learning, которые более универсальны и пригодны для применения в различных областях. На курсе предполагается просмотр лекций, выполнение квизов и домашних заданий, а также выполнение курсового проекта

1. Цели и задачи

Цель дисциплины

Освоение глубокого обучения в обработке естественного языка.

Задачи дисциплины

- освоение студентами базовых знаний (понятий, концепций, методов и моделей) в области глубокого обучения в обработке естественного языка;
- приобретение теоретических знаний и практических умений и навыков в области глубокого обучения в обработке естественного языка;
- оказание консультаций и помощи студентам в проведении собственных теоретических исследований в области глубокого обучения в обработке естественного языка.

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-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.3 Знает основные требования информационной безопасности

ОПК-3 Способен составлять и оформлять научные и (или) технические (технологические, инновационные) отчеты (публикации, проекты)	ОПК-3.1 Знает основные правила оформления научных публикаций и научно-технической документации, в том числе с использованием прикладного программного обеспечения
	ОПК-3.2 Владеет на практике методологией составления научно-технических отчетов (проектов)
	ОПК-3.3 Владеет методами визуального и графического представления результатов научной (научно-технической, инновационной технологической) деятельности в виде отчетов, научных публикаций
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценить качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива
	ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- ☐ фундаментальные понятия, законы, теории глубокого обучения в обработке естественного языка;
- ☐ современные проблемы соответствующих разделов глубокого обучения в обработке естественного языка;
- ☐ понятия, аксиомы, методы доказательств и доказательства основных теорем в разделах, входящих в базовую часть цикла;
- ☐ основные свойства соответствующих математических объектов;
- ☐ аналитические и численные подходы и методы для решения типовых прикладных задач глубокого обучения в обработке естественного языка.

уметь:

- ☐ понять поставленную задачу;
- ☐ использовать свои знания для решения фундаментальных и прикладных задач ЭК;
- ☐ оценивать корректность постановок задач;
- ☐ строго доказывать или опровергать утверждение;
- ☐ самостоятельно находить алгоритмы решения задач, в том числе и нестандартных, и проводить их анализ;
- ☐ самостоятельно видеть следствия полученных результатов;
- ☐ точно представить математические знания в области сложных вычислений в устной и письменной форме.

владеть:

- ☐ навыками освоения большого объема информации и решения задач (в том числе, сложных);
- ☐ навыками самостоятельной работы и освоения новых дисциплин;
- ☐ культурой постановки, анализа и решения математических и прикладных задач, требующих для своего решения использования математических подходов и методов ЭК;
- ☐ предметным языком глубокого обучения в обработке естественного языка и навыками грамотного описания решения задач и представления полученных результатов.

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 в обработку естественного языка и глубокое обучение.	6	6		12
2	Более сложные векторные представления слов: модели языка, softmax, одноуровневые нейронные сети.	6	6		12
3	Нейронные сети и обратное распространение ошибки в деталях.	6	6		15
4	Рекуррентные нейронные сети в применении к моделированию языка и другим задачам.	6	6		18
5	Сверточные нейронные сети в применении к классификации предложений.	6	6		18
Итого часов		30	30		75
Подготовка к экзамену		0 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 6 (Весенний)

1. Введение в обработку естественного языка и глубокое обучение.

Простые векторные представления слов: word2vec, GloVe.

2. Более сложные векторные представления слов: модели языка, softmax, одноуровневые нейронные сети.

Нейронные сети и обратное распространение ошибки в приложении к распознаванию именованных сущностей.

3. Нейронные сети и обратное распространение ошибки в деталях.

Практические советы: проверки на градиент, переобучение, регуляризация, функции активации.

4. Рекуррентные нейронные сети в применении к моделированию языка и другим задачам.

GRU и LSTM в применении к машинному переводу.

5. Сверточные нейронные сети в применении к классификации предложений.

Будущее глубокого обучения для обработки естественного языка: сети с динамической памятью.

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

Стандартная учебная аудитория с проектором.

6. Перечень рекомендуемой литературы

Основная литература

1. Нейронные сети : полный курс = Neural Networks. A Comprehensive Foundation, [учебное пособие] / Саймон Хайкин ; [перевод с английского]. Санкт-Петербург, Диалектика, 2019

Дополнительная литература

1. Нейронные сети: история развития теории [Текст]. Кн. 5 : учеб. пособие для вузов / под общ. ред. А. И. Галушкина, Я. З. Цыпкина. — М. : ИПРЖР, 2001. — 840 с.

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

1. I. Goodfellow et al. Deep Learning. <http://deeplearningbook.org>
2. С. Николенко и др. Введение в глубокое обучение. - СПб. - Питер, 2017.
3. D. Jurafsky and J. H. Martin. Speech and Language Processing (3rd ed. draft). <https://web.stanford.edu/~jurafsky/slp3/>
4. Y. Goldberg. A Primer on Neural Network Models for Natural Language Processing. <http://u.cs.biu.ac.il/~yogo/nnlp.pdf>

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

На лекционных занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

В процессе самостоятельной работы обучающихся возможно использование таких программных средств, как Mathcad, MATLAB, Maple и др.

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

1. Рекомендуется успешно сдавать контрольные работы, так как это упрощает итоговую аттестацию по предмету.
2. Для подготовки к итоговой аттестации по предмету лучше всего пользоваться материалами лекций.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Информатика и вычислительная техника

профиль подготовки: Физтех-школа Прикладной Математики и Информатики
кафедра дискретной математики

курс: 3

квалификация: бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Разработчик: А.А. Сорокин, канд. физ.-мат. наук, доцент

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-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
	ОПК-1.3 Способен определять границы применимости полученных результатов
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.3 Знает основные требования информационной безопасности
ОПК-3 Способен составлять и оформлять научные и (или) технические (технологические, инновационные) отчеты (публикации, проекты)	ОПК-3.1 Знает основные правила оформления научных публикаций и научно-технической документации, в том числе с использованием прикладного программного обеспечения
	ОПК-3.2 Владеет на практике методологией составления научно-технических отчетов (проектов)
	ОПК-3.3 Владеет методами визуального и графического представления результатов научной (научно-технической, инновационной технологической) деятельности в виде отчетов, научных публикаций
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценить качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива

2. Показатели оценивания компетенций

В результате изучения дисциплины «Глубокое обучение в обработке естественного языка» обучающийся должен:

знать:

- ☐ фундаментальные понятия, законы, теории глубокого обучения в обработке естественного языка;
- ☐ современные проблемы соответствующих разделов глубокого обучения в обработке естественного языка;
- ☐ понятия, аксиомы, методы доказательств и доказательства основных теорем в разделах, входящих в базовую часть цикла;
- ☐ основные свойства соответствующих математических объектов;
- ☐ аналитические и численные подходы и методы для решения типовых прикладных задач глубокого обучения в обработке естественного языка.

уметь:

- ☐ понять поставленную задачу;
- ☐ использовать свои знания для решения фундаментальных и прикладных задач ЭК;
- ☐ оценивать корректность постановок задач;
- ☐ строго доказывать или опровергать утверждение;
- ☐ самостоятельно находить алгоритмы решения задач, в том числе и нестандартных, и проводить их анализ;
- ☐ самостоятельно видеть следствия полученных результатов;
- ☐ точно представить математические знания в области сложных вычислений в устной и письменной форме.

владеть:

- ☐ навыками освоения большого объема информации и решения задач (в том числе, сложных);
- ☐ навыками самостоятельной работы и освоения новых дисциплин;
- ☐ культурой постановки, анализа и решения математических и прикладных задач, требующих для своего решения использования математических подходов и методов ЭК;
- ☐ предметным языком глубокого обучения в обработке естественного языка и навыками грамотного описания решения задач и представления полученных результатов.

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 прошлого занятия.

Темы курсовых работ:

Рекуррентные нейронные сети и классификация текстов

Языковые модели

Задача Seq2Seq и машинный перевод

Современные архитектуры NLP: BERT, Transformer, ELMO

Диалоговые системы

Суммаризация текста

Введение в обработку аудиосигналов

Распознавание речи

Генерация речи

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

1. Введение в обработку естественного языка и глубокое обучение.
2. Простые векторные представления слов: word2vec, GloVe.

3. Более сложные векторные представления слов: модели языка, softmax, одноуровневые нейронные сети.
4. Нейронные сети и обратное распространение ошибки в приложении к распознаванию именованных сущностей.
5. Нейронные сети и обратное распространение ошибки в деталях.
6. Практические советы: проверки на градиент, переобучение, регуляризация, функции активации.
7. Рекуррентные нейронные сети в применении к моделированию языка и другим задачам.
8. GRU и LSTM в применении к машинному переводу.
9. Сверточные нейронные сети в применении к классификации предложений.
10. Будущее глубокого обучения для обработки естественного языка: сети с динамической памятью.

Критерии оценивания

- оценка «отлично (10)» 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 (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений
- оценка «хорошо (7)» 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 (6)» выставляется студенту, если он 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)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины.

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

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины.