**Бородкин Л.И.**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Межфакультетский курс по выбору**

**Наука о данных и методы искусственного интеллекта в исторических исследованиях**

Аннотация курса

Методы и технологии искусственного интеллекта уже более полувека привлекает внимание представителей различных областей гуманитарного знания, включая историков. Этот термин получил широкое распространение с начала 2000-х гг., когда привычные методы обработки и анализа данных, основанные на методах математической статистики, стали активно дополняться методами и технологиями искусственного интеллекта (ИИ), а затем и больших данных. Статистические методы и сегодня часто рассматриваются как важная компонента науки о данных, но неотъемлемой ее частью стали методы обучения машин, искусственные нейросети и другие технологии искусственного интеллекта. И хотя большинство специалистов в области ИИ работают в сфере бизнес-приложений, спрос на них растет и в социально-гуманитарных исследованиях.

Сегодня искусственный интеллект – широкая область теоретических исследований и прикладных разработок, она включает целый ряд направлений. Для гуманитарных наук актуальными являются машинное обучение (наиболее популярный его вид связан с использованием искусственных нейронных сетей); распознавание образов; экспертные системы (основанные на базах знаний); репрезентация знаний; обработка естественного языка (NLP); нечеткие модели и др. Имеют ли историки опыт применения методов искусственного интеллекта? Да, такой опыт имеется, и он берет начало еще в прошлом веке. Актуальной задачей по-прежнему является распознавание отсканированных рукописных и старопечатных текстов – ведь ежедневно делаются доступными пользователю тысячи страниц оцифрованных, но не являющихся машиночитаемыми источников. В археологии методы машинного обучения чаще всего применяются в задачах классификации и типологии. В области сохранение культурного наследия проходят апробацию различные подходы к использованию машинного обучения и искусственных нейросетей в задачах атрибуции объектов культурного наследия, их типологии, ландшафтных виртуальных реконструкций и т.д. Отдельный интерес представляет вопрос о Big Data, возможностях применения технологий больших данных в гуманитарных исследованиях.

В рамках предлагаемого курса будут затронуты вопросы об эволюции структуры науки о данных, задачах машинного обучения, об искусственном интеллекте и нейронных сетях. Отдельное внимание уделяется «Большим данным», технологиям хранения и обработки больших данных. Рассматриваются дискуссионные вопросы об использовании концепций Big Data в исторических исследованиях. Приводятся примеры применения методов и технологий ИИ в исторических и – шире – гуманитарных приоложениях.

Место дисциплины (модуля) в структуре ОПОП ВО

Межфакультетский курс по выбору студента «Наука о данных и методы искусственного интеллекта в исторических исследованиях

относится к вариативной части ОПОП ВО.

Входные требования для освоения дисциплины, предварительные условия:

при освоении дисциплины обучающийся владеет элементарными математическими знаниями; владеет первичными навыками обобщения и систематизации научной информации.

Объем дисциплины

Объем дисциплины составляет 1 з.е., в том числе 20 академических часа, отведенных на контактную (онлайн) работу обучающихся с преподавателем, 8 академических часа на самостоятельную работу обучающихся, 4 часа зачет.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Наименование и краткое содержание разделов и тем дисциплины (модуля),**  **Форма промежуточной аттестации по дисциплине (модулю)** | **Всего**  **(часы**) | В том числе | | | |
| **Контактная работа  (работа во взаимодействии с преподавателем)**  **Виды контактной работы, часы** | | | **Самостоятельная работа обучающегося,**  **часы**  *(виды самостоятельной работы – эссе, реферат, контрольная работа и пр. – указываются при необходимости)* |
| Занятия лекционного типа | Занятия семинарского типа | **Всего** |  |
| Тема 1. Наука о данных. Эволюция структуры Data Science. Математическая статистика, методы многомерного анализа данных, data mining, искусственный интеллект, Большие данные.  Примеры применения методов искусственного интеллекта в исторических исследованиях 1980-х - 1990-х гг. Экспертные системы в исторических и археологических исследованиях, когнитивные методы анализа историко-политических текстов. | 2 | 2 |  | 2 |  |
| Тема 2. Задачи машинного обучения. Искусственный интеллект и нейронные сети. Терминология машинного обучения. Примеры задач машинного обучения.  Примеры применения методов искусственного интеллекта в исторических исследованиях последнего десятилетия. | 4 | 4 |  | 4 |  |
| Тема 3. Методология машинного обучения. Нейронные сети и глубокое обучение. Типология машинного обучения | 2 | 2 |  |  |  |
| Тема 4.  Необходимые условия применения методов искусственного интеллекта. Полнота, чистота, достоверность данных. Контроль качества данных. Культура постановки задач. Предметная экспертиза. Понятие о «data-driven» проектах. Культура анализа данных. Актуальность средств визуализации данных. | 2 | 2 |  | 2 | 2  реферат |
| Тема 5. О мифах искусственного интеллекта.  «Сейчас наблюдается прорыв в области Искусственного Интеллекта».  «Скоро будет создан Общий Искусственный Интеллект (AGI)».  «Тот, кто станет лидером в сфере ИИ, будет властелином мира».  «ИИ является альтернативой классическому моделированию». | 2 | 2 |  |  |  |
| Тема 6.  «Большие данные». Определение, технологии хранения и обработки больших данных. Дискуссионные вопросы об использовании концепций Big Data в исторических исследованиях. Примеры использования больших данных в исторических исследованиях. | 2 | 2 |  | 2 |  |
| Тема 7. Использование методов и технологий искусственного интеллекта в работах по 3D-моделированию и виртуальной реконструкции объектов культурного наследия | 2 | 2 |  | 2 | 2  эссе |
| Промежуточная аттестация: зачет | 4 |  | | | 4 |
| **Итого** | *20* | *24* | | | 4 |

Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине (модулю)

Типовые контрольные задания для проведения текущего контроля (реферата):

1. Data Science: эволюция структуры науки о данных. Чем характеризуется современный этап этой эволюции?
2. В чем специфика применения методов науки о данных в исторических исследованиях?
3. Три этапа процесса математизации научного знания. Общее и особенное в применении математических методов в исторических исследованиях и в гуманитарных науках в целом.
4. Статистические методы и модели в исторических исследованиях, примеры использования.
5. Регрессионный анализ, примеры использования в исторических исследованиях.

Кластерный анализ, примеры использования в исторических исследованиях.

Типовые контрольные задания или иные материалы для проведения промежуточной аттестации (зачета):

1. Data mining: особенности поиска релевантных данных в цифровой историко-ориентированной среде.
2. Основные этапы формирования методов искусственного интеллекта.
3. Экспертные системы в исторических и археологических исследованиях, когнитивные методы анализа историко-политических текстов.
4. Машинное обучение: как обеспечить эффективность их применения?
5. Методы глубокого обучения: в чем особенность их применения?
6. Дайте характеристику искусственных нейросетей.
7. Приведите примеры применения методов искусственного интеллекта в исторических исследованиях 1980-х - 1990-х гг.
8. Приведите примеры применения методов искусственного интеллекта в исторических исследованиях последнего десятилетия.
9. Определение «больших данных», характеристика методов и технологий их хранения и обработки.
10. Каковы основные дискуссионные вопросы об использовании концепций Big Data в исторических исследованиях?

Ресурсное обеспечение:

Перечень основной и дополнительной литературы

Основная литература:

Воронцов К. В. Лекции по машинному обучению. www.MachineLearning.ru, 2004-2020.

|  |
| --- |
| Бородкин Л.И. Методы искусственного интеллекта: новые горизонты исторического познания // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории РАН. 1992. № 5. С. 4-16. |
|  | Бородкин Л.И. Становление исторической информатики в России: первые шаги историков на пути «микрокомпьютерной революции» // Историческая информатика. – 2017. – №3. – С. 155-172. DOI: 10.7256/2585-7797.2017.3.24709 URL: https://nbpublish.com/library\_read\_article.php?id=24709 |
|  |  |

Дополнительная литература:

Бенджио И., Гудфеллоу Я., Курвилль А. Глубокое обучение. ДМК-Пресс, 2018.

Николенко С. И., Кадурин А. А., Архангельская Е. О. Глубокое обучение. Питер, 2018.

Коэльо Л. П., Ричарт В. Построение систем машинного обучения на языке Python. 2016.

Мерков А. Б. Распознавание образов. Введение в методы статистического обучения. 2011.

Мерков А. Б. Распознавание образов. Построение и обучение вероятностных моделей. 2014.

Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014.

Bishop C. M. Pattern Recognition and Machine Learning. - Springer, 2006.

Домингос П. Верховный алгоритм. 2016.

Перечень профессиональных баз данных и информационных справочных систем

1) Библиотека электронных ресурсов Исторического факультета МГУ: <http://www.hist.msu.ru/ER/index.html>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет» (при необходимости)

1) Цифровой ресурс «Динамика экономического и социального развития России в XIX – начале ХХ вв.»

<http://hist.msu.ru/Dynamics/>

Журнал «Историческая информатика» (2017 - 2020 гг.) в открытом доступе:

<https://www.nbpublish.com/e_istinf/contents_2020.html>

Язык преподавания: русский

Автор программы

Зав. кафедрой исторической информатики исторического факультета МГУ, д.и.н., проф., член-корр. РАН Л.И. Бородкин