МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ В.Ф. УТКИНА»

Кафедра «Вычислительная и прикладная математика»

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

«Проектирование рекомендательных систем»

Направление подготовки – 09.04.04 Программная инженерия

ОПОП академической магистратуры

«Программно-алгоритмическое обеспечение систем искусственного интеллекта»

Квалификация (степень) выпускника – магистр

Форма обучения – очная (2 года)

Рязань .

1. СПИСОК ТЕОРЕТИЧЕСКИХ ВОПРОСОВ К ЗАЧЕТУ/ЭКЗАМЕНУ

1. Рекомендательная система: определение, типы, задачи.
2. Неперсонализированные рекомендации. Ассоциативные правила. Алгоритм Apriori.
3. Матрица предпочтений (оценок), явные и неявные методы получения оценок.
4. Меры сходства, применение мер сходства в методе коллаборативной фильтрации и в методе фильтрации на основе содержания.
5. Коллаборативная фильтрация на основе сходства пользователей и на основе сходства объектов. Проблема холодного старта.
6. Фильтрация на основе содержания, модель TF-IDF.
7. Матричная факторизация. Алгоритм SVD.
8. Гибридные рекомендательные системы.
9. Методы оценки качества рекомендательных систем.
10. Ранжирование. Классификация алгоритмов обучения ранжированию.
11. Социальные сети: определение, типы, задачи. Использование данных, полученных из социальной сети, при формировании рекомендаций.
12. Графовые базы данных.
13. Клика, сообщество. Алгоритм распространения меток.
14. Задача определения влиятельности узлов: постановка, связь с рекомендательными системами. Применение алгоритма PageRank для определения влиятельности узлов.
15. Паттерны поведения пользователей в социальных сетях.
16. Графовые рекомендательные системы. Примеры использования. Представление объектов в виде графа, представление пользователей и объектов в виде двудольного графа.
17. Прогнозирование связей в графе как задача построения рекомендаций. Классификация методов прогнозирования связей в однодольном графе.
18. Методы прогнозирования связей в однодольном графе на основе топологии сети.
19. Методы прогнозирования связей в однодольном графе на основе случайного блуждания.
20. Групповые рекомендательные системы
21. Использование нечеткой логики в рекомендательных системах.

2. ПЕРЕЧЕНЬ ЛАБОРАТОРНЫХ/ПРАКТИЧЕСКИХ РАБОТ ПО ДИСЦИПЛИНЕ

Лабораторная работа 1.1. Сборка и запуск учебного веб-сайта.
Цель работы.
Подготовка рабочего места для ЛР 1.2-1.14.
Задание: скачать исходные коды учебного веб-сайта из репозитория курса, изучить их, выполнить настройку окружения в соответствии с инструкцией, запустить учебный веб-сайт и проверить работоспособность.

Лабораторная работа 1.2. Неперсонализированные рекомендации
Цель работы.
Изучение методов формирования неперсонализированных рекомендаций.
Задание: доработать серверную и клиентскую часть учебного веб-сайта для отображения неперсонализированных рекомендаций: новых и самых популярных объектов.

Лабораторная работа 1.3. Алгоритм Apriori для поиска ассоциативных правил.

Цель работы.
Изучение алгоритмов поиска ассоциативных правил.
Задание: реализовать алгоритм Apriori, применить его к данным из базы данных учебного веб-сайта, сохранить полученные ассоциативные правила в файл.

Лабораторная работа 1.4. Формирование рекомендаций на основе ассоциативных правил.

Цель работы.
1. Получение навыков формирования рекомендаций на основе ассоциативных правил.
2. Получение навыков интеграции методов формирования рекомендаций в рамках создания веб-сайтов.

Задание: доработать серверную и клиентскую часть учебного веб-сайта для отображения рекомендаций, сформированных на основе ассоциативных правил.

Лабораторная работа 1.5. Сбор логов о действиях пользователя на страницах веб-сайта.

Цель работы.
Получение навыков программирования сборщиков данных на стороне клиента.

Задание: спроектировать таблицу базы данных, доработать серверную и клиентскую часть учебного веб-сайта для сбора логов о действиях пользователя на веб-страницах: кликов, наведений курсора, скроллинга.

Лабораторная работа 1.6. Расчет неявных оценок на основе логов о действиях пользователя на страницах веб-сайта.

Цель работы.
Изучение формул расчета неявных оценок на основе логов о действиях пользователя на веб-страницах.

Задание: для каждого пользователя сгенерировать логи, рассчитать неявные оценки, сохранить полученную матрицу неявных оценок в файл.

Лабораторная работа 1.7. Алгоритм коллаборативной фильтрации на основе сходства пользователей.

Цель работы.
1. Изучение мер близости.
2. Изучение алгоритма коллаборативной фильтрации на основе сходства пользователей.

Задание: с использованием готовых библиотек (numpy, sklearn) реализовать алгоритм коллаборативной фильтрации на основе сходства пользователей, определяемого с помощью косинусной меры близости, применить его к данным из базы данных учебного веб-сайта, сохранить полученную матрицу предпочтений в файл.

Лабораторная работа 1.8. Алгоритм коллаборативной фильтрации на основе сходства объектов.

Цель работы.

1. Изучение мер близости.
2. Изучение алгоритма коллаборативной фильтрации на основе сходства объектов.

Задание: с использованием готовых библиотек (numpy, sklearn) реализовать алгоритм коллаборативной фильтрации на основе сходства объектов, определяемого с помощью косинусной меры близости, применить его к данным из базы данных учебного веб-сайта, сохранить полученную матрицу предпочтений в файл.

Лабораторная работа 1.9. Составление профиля объекта.
Цель работы.

1. Получение навыков применения модели TF-IDF в рекомендательных системах.
2. Подготовка к выполнению ЛР 1.10 и ЛР 1.11.

Задание: для каждого объекта в учебной базе данных составить вектор с использованием модели TF-IDF, сохранить полученные результаты в базу данных.

Лабораторная работа 1.10. Составление профиля пользователя.
Цель работы.

1. Получение навыков применения модели TF-IDF в рекомендательных системах.
2. Подготовка к выполнению ЛР 1.11.

Задание: для каждого пользователя в учебной базе данных составить вектор следующим образом: найти объекты, которые пользователь оценил выше, чем на заданную оценку, умножить вектор каждого объекта на соответствующую оценку пользователя, сложить векторы поэлементно.

Лабораторная работа 1.11. Алгоритм фильтрации на основе содержания

Цель работы:
Изучение алгоритма фильтрации на основе содержания.

Задание: используя косинусную меру близости и результаты ЛР 1.9-1.10, заполнить матрицу предпочтений значениями сходства профиля пользователя и профиля объекта, сохранить полученную матрицу предпочтений в файл.

Лабораторная работа 1.12. Алгоритмы матричной факторизации.

Цель работы:

1. Изучение алгоритмов матричной факторизации.
2. Изучение готовых библиотек, реализующих алгоритмы матричной факторизации.

Задание: выбрать алгоритм матричной факторизации (SVD, SVD++, NMF), применить функцию из библиотеки scikit-surpize, реализующую выбранный алгоритм, к данным из базы данных учебного веб-сайта, сохранить полученную матрицу предпочтений в файл.

Лабораторная работа 1.13. Гибридный рекомендательный алгоритм.

Цель работы.

Изучение гибридных рекомендательных алгоритмов.

Задание: реализовать простейший гибридный алгоритм: если количество оценок, поставленных пользователем, меньше заданного, использовать алгоритм фильтрации на основе содержания, иначе – алгоритм коллаборативной фильтрации на основе сходства пользователей, применить реализованный гибридный алгоритм к данным из базы данных учебного веб-сайта, сохранить полученную матрицу предпочтений в файл.

Лабораторная работа 1.14. Формирование рекомендаций на основе матрицы предпочтений.

Цель работы.

1. Получение навыков формирования рекомендаций на основе матрицы предпочтений.
2. Получение навыков интеграции методов формирования рекомендаций в рамках создания веб-сайтов.

Задание: для каждого пользователя сформировать рекомендации на основе матрицы предпочтений, полученной в любой из предыдущих лабораторных работ, доработать серверную и клиентскую часть учебного веб-сайта для отображения рекомендаций.

Лабораторная работа 2.1. Графовая СУБД.

Цель работы.
Формирование навыков работы с графовой базой данных и СУБД.
Задание: установить графовую СУБД, выполнить подключение к базе данных учебной социальной сети, изучить структуру базы данных.

Лабораторная работа 2.2. Выполнение запросов к графовой базе данных.

Цель работы.
Формирование навыков работы с графовой базой данных и СУБД.
Задание: выполнить запросы по варианту к базе данных учебной социальной сети.

Лабораторная работа 2.3. Подготовка демографических данных пользователей для кластеризации.

Цель работы.
Получение навыков обработки демографических данных.
Задание: выбрать алгоритм кластеризации, выполнить все необходимые преобразования демографических данных пользователей для последующей кластеризации.

Лабораторная работа 2.4. Кластеризация пользователей по демографическим данным.

Цель работы.
Получение навыков сегментации пользователей.
Задание: выполнить кластеризацию пользователей по демографическим данным выбранным в ЛР 2.3 алгоритмом.

Лабораторная работа 2.5. Визуализация графа учебной социальной сети.

Цель работы.
Получение навыков работы с инструментами визуализации больших графов

Задание: выбрать инструмент визуализации (networkX, igraph, gephi, graph-tool или аналогичный), визуализировать граф учебной социальной сети.

Лабораторная работа 2.6. Алгоритмы выделения сообществ.
Цель работы.

1. Изучение алгоритмов выделения сообществ в графах.
2. Получение навыков сегментации пользователей.

Задание: выбрать алгоритм выделения сообществ, выбрать библиотеку анализа графов, реализующую выбранный алгоритм (networkX, igraph, gephi, graph-tool или аналогичную), применить выбранный алгоритм к графу учебной социальной сети, визуализировать полученный результат.

Лабораторная работа 2.7. Коллаборативная фильтрация с учетом сегмента пользователя.
Цель работы.
Получение навыков использования данных, извлекаемых из социальной сети, в задаче формирования рекомендаций.

Задание: заполнить матрицу оценок для холодных пользователей (холодными считать пользователей, поставивших менее заданного количества оценок). Для этого определить сегмент холодного пользователя (дескрипторы сегментов получить по результатам ЛР 2.4 или 2.6 на выбор) и применить алгоритм коллаборативной фильтрации на основе сходства пользователей ко всем пользователям из того же сегмента, что и рассматриваемый холодный пользователь.

Лабораторная работа 2.8. Коллаборативная фильтрация с учетом сегмента пользователя.
Цель работы.
Получение навыков использования данных, извлекаемых из социальной сети, в задаче формирования рекомендаций.

Задание: заполнить матрицу оценок для нехолодных пользователей (нехолодными считать пользователей, поставивших более заданного количества оценок). Для этого определить сегмент нехолодного пользователя (дескрипторы сегментов получить по результатам ЛР 2.4 или 2.6 на выбор) и применить алгоритм коллаборативной фильтрации на основе сходства пользователей к k ближайшим пользователям из того же сегмента, что и рассматриваемый нехолодный пользователь.

Лабораторная работа 2.9. Алгоритм node2vec.
Цель работы.

1. Изучение алгоритма node2vec.
2. Получение навыков формирования рекомендаций в графовых рекомендательных системах.

Задание: реализовать первый этап задачи рекомендации друзей в социальной сети: выбрать библиотеку, реализующую алгоритм node2vec, и применить этот алгоритм к узлам в графе учебной социальной сети.

Лабораторная работа 2.10. Прогнозирование связей в графе социальной сети.

Цель работы.

1. Получение навыков использования алгоритмов классификации в задачах прогнозирования.

2. Получение навыков формирования рекомендаций в графовых рекомендательных системах.

Задание: реализовать второй этап задачи рекомендации друзей в социальной сети: выбрать алгоритм классификации, выбрать библиотеку, реализующую выбранный алгоритм, и применить этот алгоритм к узлам в графе учебной социальной сети (вектор признаков узлов взять из ЛР 2.9).

Лабораторная работа 2.11 Формирование рекомендаций по результатам прогнозирования.

Цель работы.

1. Получение навыков интерпретации результатов прогнозирования связей в графе для формирования рекомендаций.
2. Получение навыков формирования рекомендаций в графовых рекомендательных системах.

Задание: на основе результатов ЛР 2.10 сформировать ТОП-N наиболее вероятных друзей заданного пользователя.

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Изучение дисциплины проходит в течение одного семестра. Основные темы дисциплины осваиваются в ходе аудиторных занятий, однако важная роль отводится и самостоятельной работе студентов.

Самостоятельная работа как вид учебной работы может использоваться на лабораторных работах, а также иметь самостоятельное значение – внеаудиторная самостоятельная работа обучающихся – при подготовке к лабораторным работам, при подготовке к дифференцированному зачету.

Самостоятельная работа включает в себя следующие этапы:

- изучение теоретического материала (работа над конспектом лекции);
- самостоятельное изучение дополнительных информационных ресурсов (доработка конспекта лекции);
- выполнение заданий текущего контроля успеваемости (подготовка к лабораторным работам);
- итоговая аттестация по дисциплине (подготовка к дифференцированному зачету/ экзамену).

4. РЕКОМЕНДАЦИИ ПО ПЛАНИРОВАНИЮ И ОРГАНИЗАЦИИ ВРЕМЕНИ, НЕОБХОДИМОГО ДЛЯ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ. ОПИСАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ ДЕЙСТВИЙ СТУДЕНТА («СЦЕНАРИЙ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ»)

Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины.

Для освоения лекционного материала следует: изучить конспект лекции в тот же день, после лекции: 10 – 15 минут, повторно прочитать конспект лекции за день перед следующей лекцией: 10 – 15 минут. Также следует изучить теоретический лекционный материал по рекомендуемому учебнику/учебному пособию: 1 час в неделю.

Следует максимально использовать лекционное время для изучения дисциплины, понимания лекционного материала и написания конспекта лекций. В процессе лекционного занятия студент должен уметь выделять важные моменты и основные положения. При написании *конспекта лекций* следует придерживаться следующих правил и рекомендаций.

1. При ведении конспекта рекомендуется структурировать материал по разделам, главам, темам. Вести нумерацию формул. Выделять по каждой теме постановку задачи, основные положения, выводы. Кратко записывать те пояснения лектора, которые показались особенно важными. Это позволит при подготовке к сдаче зачёта не запутаться в структуре лекционного материала.

2. Лекционный материал следует записывать в конспект лишь после того, как излагаемый лектором тезис будет вами дослушан до конца и понят.

3. При конспектировании следует отмечать непонятные, на данном этапе, положения, доказательства и пр.

4. Рекомендуется по каждой теме выразить свое мнение, комментарий, вывод.

Доработка конспекта лекции с применением учебника, методической литературы, дополнительной литературы, интернет-ресурсов: этот вид самостоятельной работы студентов особенно важен в том случае, когда одну и ту же задачу можно решать различными способами, а на лекции изложен только один из них. Кроме того, рабочая программа предполагает рассмотрение некоторых относительно несложных тем только во время самостоятельных занятий, без чтения лектором.

Подготовка к лабораторным работам состоит в теоретической подготовке (изучение конспекта лекций, методических указаний к данной лабораторной работе и дополнительной литературы) и выполнении индивидуального задания. Выполнение каждой из запланированных работ заканчивается предоставлением отчета. Требования к форме и содержанию отчета приведены в методических указаниях к лабораторным работам или определяются преподавателем на первом занятии. Допускаясь к лабораторной работе, каждый студент должен представить преподавателю «заготовку» отчета, содержащую: оформленный титульный лист, цель работы, задание, проект решения, полученные результаты, выводы.

Важным этапом является защита лабораторной работы. В процессе защиты студент отвечает на вопросы преподавателя, касающиеся теоретического материала, относящегося к данной работе, и проекта, реализующего его задание, комментирует полученные в ходе работы результаты. При подготовке к защите лабораторной работы рекомендуется ознакомиться со списком вопросов по изучаемой теме и попытаться самостоятельно на них ответить, используя конспект лекций и рекомендуемую литературу.

Подготовка к сдаче дифференцированного зачета.

Зачет – форма промежуточной проверки знаний, умений, навыков, степени освоения дисциплины. Главная задача зачета состоит в том, чтобы у студента по окончанию изучения данной дисциплины сформировались определенное представление об общем содержании дисциплины, определенные теоретические знания и практические навыки, определенный кругозор. Готовясь к зачету, студент приводит в систему знания, полученные на лекциях, на практических и лабораторных занятиях, разбирается в том, что осталось непонятным, и тогда изучаемая им дисциплина может быть воспринята в полном объеме с присущей ей строгостью и логичностью, ее практической направленностью.

Зачеты дают возможность преподавателю определить теоретические знания студента и его практические навыки при решении определенных прикладных задач. Оцениваются: понимание и степень усвоения теоретического материала; степень знакомства с основной и дополнительно литературой, а также с современными публикациями; умение применить теорию к практике, решать определенные практические задачи данной предметной области, правильно проводить расчеты и т. д.; знакомство с историей данной науки; логика, структура и стиль ответа, умение защищать выдвигаемые положения.

Значение зачета не ограничивается проверкой знаний, являясь естественным завершением обучения студента по данной дисциплине, они способствуют обобщению и закреплению знаний и умений, приведению их в стройную систему, а также устранению возникших в процессе обучения пробелов.

Подготовка к зачету – это тщательное изучение и систематизация учебного мате-

риала, осмысление и запоминание теоретических положений, формулировок, формул, установление и осмысление внутрипредметных связей между различными темами дисциплины, закрепление теоретических знаний путем решения определенных задач.

Планируйте подготовку к зачету, учитывая сразу несколько факторов: неоднородность в сложности учебного материала и степени его проработки в ходе обучения, свои индивидуальные способности. Рекомендуется делать перерывы в занятиях через каждые 50-60 минут на 10 минут. После 3-4 часов занятий следует сделать часовой перерыв. Чрезмерное утомление приведет к снижению тонуса интеллектуальной деятельности. Целесообразно разделять весь рабочий день на три рабочих периода – с утра до обеда, с обеда до ужина и с ужина до сна. Каждый рабочий период дня должен заканчиваться отдыхом не менее 1 часа. Работая в сессионном режиме, студент имеет возможность увеличить время занятий с 10 (как требовалось в семестре) до 12 часов в сутки.

Подготовку к зачету следует начинать с общего планирования своей деятельности. С определения объема материала, подлежащего проработке, необходимо внимательно сверить конспекты с программой дисциплины, чтобы убедиться, все ли разделы отражены в лекциях, отсутствующие темы изучить по учебнику. Второй этап предусматривает системное изучение материала по данному предмету с обязательной записью всех выкладок, выводов, формул. На третьем этапе – этапе закрепления – полезно чередовать углубленное повторение особенно сложных вопросов с беглым повторением всего материала.

5. РЕКОМЕНДАЦИИ ПО РАБОТЕ С ЛИТЕРАТУРОЙ

Теоретический материал курса становится более понятным, когда дополнительно к прослушиванию лекции и изучению конспекта изучаются и книги по данному предмету. Литературу по дисциплине рекомендуется читать как в бумажном, так и в электронном виде (если отсутствует бумажный аналог). Полезно использовать несколько учебников и пособий по дисциплине. Рекомендуется после изучения очередного параграфа ответить на несколько вопросов по данной теме. Кроме того, полезно мысленно задать себе следующие вопросы (и попробовать ответить на них): «о чем этот параграф?», «какие новые понятия введены, каков их смысл?», «зачем мне это нужно по специальности?».

Рекомендуется самостоятельно изучать материал, который еще не прочитан на лекции и не применялся на лабораторном или практическом занятии, тогда занятия будут гораздо понятнее. В течение недели рекомендуется выбрать время (1 час) для работы с литературой.

Оператор ЭДО ООО "Компания "Тензор"

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

ПОДПИСАНО ЗАВЕДУЮЩИМ КАФЕДРЫ	**ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ	**12.10.26** 09:03 (MSK)	Простая подпись
ПОДПИСАНО ЗАВЕДУЮЩИМ ВЫПУСКАЮЩЕЙ КАФЕДРЫ	**ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ	**12.10.26** 09:03 (MSK)	Простая подпись