

Фамилия, имя, номер группы:

.....

Мы выживем назло Вселенной. Нас не сломить.

(Элли из Last of us перед тем как решить эту контрольную)

Работа состоит из открытых вопросов на разные темы. На каждый из них вам необходимо дать краткие, но ёмкие ответы. Около каждого вопроса указано количество баллов, которое можно за него получить. Если у вопроса несколько подпунктов, баллы разделяются между ними равномерно.

Вопрос 1. (2.5 балла)

Начался зомби-апокалипсис. Гриб кордицепс мутировал из-за глобального потепления и полностью подчиняет себе нервную систему заражённых людей.

Билл урылся от зомби у себя дома. Он превратил его в крепость и **живёт один**. Ему скучно. Мимо проходил выпускник ИАДа, то есть вы. Билл достал свою двустолку и вежливо попросил вас обучить ему рекомендательную систему для фильмов, чтобы можно было коротать время.

1. Получится для одного Билла обучить user-based колаборативную фильтрацию? А content-based? А матричную факторизацию?
2. Что такое контентные рекомендации? Как можно обучить такую рекомендательную систему для Билла?
3. Что такое проблема холодного старта? Будет ли она актуальна для рекомендательной системы Билла? Как её решить?
4. Что такое feedback loop? Будет ли эта проблема актуальна для рекомендательной системы Билла? Как её решить?

Вопрос 2. (2.5 балла)

Начался зомби-апокалипсис. Гриб кордицепс мутировал из-за глобального потепления и полностью подчиняет себе нервную систему заражённых людей.

Люди концентрируются в городах и огораживают их огромными стенами. В городах процветает чёрный рынок. В том числе чёрный рынок нейросетей. Вас, выпускника ИАДа, наняли как эксперта, чтобы разобраться с проблемами:

1. У торговцев есть GAN, обученный на датасете со многими видами животных, но он генерирует только картинки котиков. Какая проблема произошла при обучении? Проиллюстрируйте ее для двумерного случая.
2. Один из торговцев нашел на флэшке дискриминатор из другого GAN. Он предлагает после обычного обучения «дообучить» ваш GAN с его помощью. Как это может помочь и в чём опасность такого подхода?
3. Чёрные дельцы предлагают для нормализационного потока использовать нейронную сеть, которая принимает на вход картинки 256×256 и возвращает вектор размерности 64. Почему такую модель не получится обучить?

Вопрос 3. (2.5 балла)

Начался зомби-апокалипсис. Гриб кордицепс мутировал из-за глобального потепления и полностью подчиняет себе нервную систему заражённых людей.

Чтобы лучше взаимодействовать между собой, зомби создали социальную сеть. Они поймали выпускника ИАДа, то есть вас и просят помочь сделать её лучше. Пока вы им помогаете, вас обещали не есть.

Зомби хотят реализовать внутри социальной сети механизм рекомендации нового друга. В этом им могут помочь графы. Каждая вершина – это пользователь социальной сети. Ребро между ними означает, что пользователи добавили друг-друга в друзья.

1. Опишите алгоритм Node2Vec для обучения представлений вершин графа.
2. Предположим, что мы используем в качестве функции схожести вероятность перехода из одной вершины в другую при случайном блуждании по графу. Опишите как вы будете готовить выборку для обучения Node2Vec, используя случайное блуждание.
3. Как при обучении представлений вершин графа можно учесть дополнительные факторы? Например, информацию со странички пользователя, его пол, возраст и тп.
4. Вы обучили для каждой вершины графа эмбединги. Теперь вам нужно обучить модель, которая по этим эмбедингам будет предсказывать, что два конкретных человека добавят друг-друга в друзья. Как собрать данные для обучения такой модели? Что будет таргетом? Что будет факторами? Какую модель вы обучите и как будете строить предсказание?

Вопрос 4. (2.5 балла)

Начался зомби-апокалипсис. Гриб кордицепс мутировал из-за глобального потепления и полностью подчиняет себе нервную систему заражённых людей.

Лучшие умы планеты (выпускники ИАДа) расшифровали язык зомби и пытаются обучить speech2text для него. Вы руководитель исследовательской группы.

1. Опишите пайплайн, для решения такой задачи. Как вы предобработаете аудиозаписи перед обучением модели? Какую архитектуру будете использовать? Как будете разбивать данные на обучающие и тестовые?
2. Какие метрики качества для задачи распознавания текста в аудио вы знаете? Как их можно вычислить на тестовой выборке?
3. Вы обучили модель и после этого выяснили, что все данные для обучения собирались в идеальных лабораторных условиях. В полевых условиях микрофон улавливает кучу шума и распознавание речи работает на порядок хуже. Бюджета на сбор новых данных нет. Как бы вы решили эту проблему, используя старые данные?