

Что с нами будет? Письма ученых о самом важном

Это письмо о том, как работают рекомендательные системы

На этой неделе вышел новый выпуск нашего шоу «[Заходит ученый в бар](#)». В нем географ-урбанист Руслан Дохов, разработчица Ксения Шашкина и комик Илья Куруч обсуждают, как современные города адаптируется к технологиям, почему все едут в Краснодар и для чего нужны беспилотные лодки. Смотрите новый эпизод и подписывайтесь на наш [ютюб-канал](#)!



Всем привет!

Меня зовут Иван Ямщиков. Я исследователь искусственного интеллекта, доцент ВШЭ в Петербурге и руководитель [Лаборатории естественного языка](#), где мы анализируем данные в компьютерной лингвистике и разрабатываем методы deep learning для генерации текстов. Еще я веду подкаст «[Прогрессимся](#)». Один из его [выпусков](#), кстати, посвящен работе рекомендательных систем. В этом письме речь тоже пойдет о них. А именно:

- От чего зависит эффективность рекомендательных алгоритмов?
- Как системы учатся на своих ошибках и обратной связи?
- С какими целями сервисы максимизируют время, которое мы в них проводим?
- И что будет с рекомендательными системами в будущем?

Приступим.

Как известно, чтобы сделать хороший продукт, нужно быть в нужное время в нужном месте. Сейчас об этом мало кто помнит, но в начале 2010-х создатели Twitter запустил сервис Vine для публикации зацикленных видеороликов длительностью до шести секунд. Идея была следующая: пользователи будут снимать видео и шерить их между собой — получится такой Twitter для видео. Ничего не напоминает?

Vine [закрылся](#) в 2016 году. Оказалось, что тогда к такому приложению массовая аудитория не была готова — проникновение мобильного интернета было достаточно низким, как и качество мобильного видео. Vine опередил свое время.

Ему на смену пришел TikTok, которым сейчас пользуются миллиарды людей по всему миру. Компания ByteDance, по сути, взяла ту же идею, добавив в свой сервис качественный рекомендательный алгоритм и выбрав подходящий момент для запуска.

Как работают рекомендательные системы

Разные системы решают разные задачи, но базовый принцип работы рекомендательных алгоритмов основан на трех простых и понятных идеях:

1. Есть вещи, которые созданы друг для друга.
2. Похожим людям нравятся похожие вещи.
3. Хорошего много не бывает.

Базовый рекомендательный сервис — это надпись: «Часто с этим продуктом покупают вот это». Если вы купили зубную щетку, может быть, вам нужна и паста, а при покупке машины может потребоваться зимний комплект резины. Неважно, кто вы, где живете и кем работаете. Это очень понятный механизм.

Другая простая схема сводится к мысли: «Вашим друзьям понравилось вот это». Если вы дружите, значит, предполагается, что вы принадлежите к одной социальной

группе, внутри которой у вас могут быть похожие вкусы и интересы.

И третья идея: «Вам понравилось видео с котиками, вот вам еще несколько таких». В этом случае система анализирует ваши предпочтения. Если вам что-то уже понравилось — значит, надо предоставить еще больше подходящего контента.

Важно отметить, что во всех этих случаях существует возможность обратной связи. То есть когда алгоритм что-то посоветовал, а вам не понравилось или наоборот. За счет обратной связи система еще точнее настраивается под ваши требования.

В этом процессе есть много тонкостей, связанных с тем, что отрицательный сигнал иногда может быть важнее, чем положительный, или наоборот. Например, в Tinder есть пользователи, которые в основном свайпают влево, и те, кто свайпает вправо. Соответственно, в зависимости от того, привередлив пользователь или неприхотлив, в одном случае важен положительный сигнал, когда что-то понравилось привереде, а в другом — отрицательный, потому что, даже если неприхотливому человеку что-то не подошло, значит, алгоритм сильно ошибся и надо это учесть.

Как собирают данные для запуска рекомендательного алгоритма

Когда мы говорим о рекомендательных продуктах, самый тонкий и сложный момент — собрать первичные данные, чтобы система начала работать. Это так называемая проблема холодного старта.

Есть разные способы ее решить, но один из самых простых подходов такой: показывать пользователям случайные рекомендации и смотреть, на что они кликают, а на что нет. Подобные рекомендательные алгоритмы запускаются после того, как сервис начал работать, и уже в процессе собирают данные для обучения рекомендательных систем.

Чем больше данных, тем лучше работает алгоритм. При этом можно с высокой долей уверенности сказать, что какая-нибудь рекомендательная система лучше, чем ничего. Однако от системы, которая работает как-то, до алгоритма, от которого пользователи в восторге, — огромный путь.

В своей книге *Machine Learning Yearning* ученый-информатик Эндрю Ын пишет среди прочего, что любой алгоритм машинного обучения бывает двух типов: или это абсолютно бесполезная штука, которую вообще непонятно для чего создали, или это суперудобная вещь, после появления которой люди не понимают, как раньше без нее жили. Книгу Ына, кстати, перевели на русский и ее можно почитать на [«Хабре»](#).

Приведу пример не из области рекомендательных систем. Когда «Яндекс» семь лет назад впервые [запустил](#) «Яндекс.Диктовку» — распознавание голоса и написание текста на русском языке, — особых восторгов по этому поводу не было. А сейчас люди используют голосовые интерфейсы постоянно: просят «Алису» построить маршрут, подсказать, сколько мяса положить в пельмени, включить музыку.

Это происходит автоматически, и мы не задумываемся о том, что в этот момент происходит магия: микрофон нашего телефона улавливает нашу речь, специальный алгоритм преобразует эти слова в текст, а потом запускает процесс поиска. И эта система работает даже для маленького ребенка — алгоритм понимает его лепет, потому что обучился на огромном количестве примеров речи.

То же самое происходит и с рекомендательными системами. Алгоритмы обучаются и становятся эффективнее и полезнее.

К слову, о том, как некоторые технологии незаметно проникают в жизнь современных городов, говорят участники нового выпуска нашего шоу [«Заходит ученый в бар»](#). Как телеграм-боты могут помочь с уборкой мусора? Где в России testируют роботов-курьеров, беспилотные такси и лодки? И какую социальную функцию выполняют ларьки у вашего дома? Смотрите в новом эпизоде.

Насколько сильно алгоритмы влияют на нас

Часто говорят, что при помощи рекомендательных алгоритмов соцсети и онлайн-кинотеатры пытаются удержать наше внимание. На мой взгляд, формулировка о том, что нас где-тодерживают, — неверная. Людейдерживают в тюрьме или в автозаке, а не в TikTok. Мне кажется, некорректно так говорить.

Да, существует такое понятие, как экономика внимания. И действительно, сервисы пытаются максимизировать время, которое мы на них проводим. Будь то соцсеть, YouTube или Netflix.

YouTube, например, максимизирует время для того, чтобы вы посмотрели больше рекламы. Поскольку вы не готовы платить ему деньги, сервис пытается увеличить время вашего присутствия на сайте, показать вам больше рекламы и окупить стоимость своих серверов. Как говорится, если вы не платите за товар, значит, товар — это вы.

На мой взгляд, это более чем справедливая сделка. Только представьте: у вас есть безлимитный доступ к контенту со всего мира, любой человек может снять видео, сделать собственное шоу и охватить любую аудиторию. Это магия.

Netflix — платный сервис. Вы платите за подписку, поэтому рекомендации Netflix имеют другую цель. Этот сервис удерживает ваше внимание, чтобы доказать вам свою ценность. Если бы у него не было хорошей рекомендательной системы, в какой-то момент вы бы перестали за него платить. То есть задача Netflix — показать, что он полезен и регулярно поставляет вам контент, который вам нравится.

Поэтому, когда говорят, что сервисы нас удерживают, складывается впечатление, что речь идет о неразумных детях, а не о взрослых людях. На мой взгляд, нормально, что есть платформы, которые зарабатывают на времени пользователя. Но есть и те, кто этого не делает. Альтернативные модели тоже существуют, и их надо развивать. У нас в подкасте, например, нет рекламных интеграций, мы работаем с помощью Patreon.

Что будет дальше

Объем контента растет, и рекомендательные системы совершенно точно нужны. Будучи технооптимистом, я надеюсь, что люди будут глубже понимать те технологии, которые существуют, и использовать их грамотнее.

Грубо говоря, зайдите на YouTube, посмотрите, что он вам рекомендует и решите, нравится вам это или нет. Например, у вас там «[Заходит ученый в бар](#)», [лекции](#) Роберта Сапольски и [видео](#) про котиков. Если вы всё это смотрите и вам

нравится — отлично. А если вам кажется, что видео про котиков много, а содержательных вещей мало, покликайте на лекции, посмотрите их, алгоритм это учит. То есть можно просто занять проактивную позицию и скорректировать свою ленту.

По мере роста компьютерной грамотности пользователей алгоритмы будут к ней адаптироваться. И мы сможем еще больше контролировать рекомендательный контент. Уже сейчас мы это видим. Есть рекомендательные системы с режимом *discovery*, его можно применять, если хочешь посмотреть что-то новенькое. Я думаю, что подобных механизмов появится больше. Настройки алгоритмов будут расширяться, и ваша лента станет формироваться в зависимости от того, что нужно именно вам.

Что еще почитать, посмотреть и послушать о работе рекомендательных систем

- Выпуск подкаста «[Проветримся](#)», в котором мы обсуждаем рекомендательные системы с Евгением Росинским из онлайн-кинотеатра *Ivi* и Анастасией Августовской, которая руководит работой с клиентами и агентствами в «Яндекс.Дзене».
- [Лекция](#) о том, как работают рекомендательные системы «Яндекс.Музыки» и «Яндекс.Дзена».
- [Пост](#) *vas3ik* с рассказом о принципах работы рекомендательных систем.

На этом всё,
Иван

Science Bar Hopping — это научный фестиваль, который организуют [Фонд инфраструктурных и образовательных программ \(Группа РОСНАНО\)](#) и [«Бумага»](#). Также мы делаем [научную рассылку](#) и YouTube-шоу [«Заходит ученый в бар»](#).

[Отписаться](#)