

СФЕРЫ ИСПОЛЬЗОВАНИЯ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ

Лепешко Р.О. ¹, студент гр.053504

*Белорусский государственный университет информатики и радиоэлектроники¹
г. Минск, Республика Беларусь*

Марков А.Н. – старший преподаватель кафедры информатики

Аннотация. В данной статье исследуются возможности генеративных моделей. Генеративные модели способны имитировать человеческий интеллект, что позволяет им успешно анализировать данные и создавать их. Они применяются в различных сферах от сферы услуг до проектирования деталей космического корабля.

Ключевые слова. Генеративная модель, информация, информационные технологии, сферы, генерация, анализ.

Генеративная модель – это новый вид искусственного интеллекта, который может как анализировать данные, так и создавать новые на основе ранее изученных. Данная технология пытается имитировать человеческий интеллект во многих задачах: перевод текста, написание программ, распознавание изображений, создание изображений. Его можно обучить любой технологии и языку. Например, модель GPT-3 была обучена с использованием более 175 миллиардов параметров. Инженеры OpenAI обучили ее на более чем 45 терабайтах данных из различных источников таких, как статьи в интернете, книги, материалы из Википедии и других.

Генеративные модели при генерации данных пытаются предсказывать признаки по определенной метке. При обучении модели главное дать ей понимание того как конструируются данные, чтобы в будущем она могла создавать уникальные данные.

Генеративные модели также можно использовать для анализа и исследования различных данных. Они способны обобщать информацию, обнаруживать тенденции и закономерности, наблюдаемые в наборах данных. Например, при анализе изображений животных и модели регистрируют такие переменные, как различные формы ушей, глаз, хвостов и кожи. Они изучают признаки и их взаимосвязи, чтобы понять, как выглядят различные животные в целом. Затем они могут создавать новые изображения животных, которых не было в наборе обучающих данных.

Базовые модели – это модели машинного обучения, обученные на широком спектре обобщенных и немаркированных данных, они используют изученные закономерности и взаимосвязи для прогнозирования каждого элемента последовательности [1]. Данные модели являются результатом последних достижений в области машинного обучения.

Большие языковые модели (LLM) – это один из классов базовых моделей. LLM специально ориентированы на языковые задачи, такие как обобщение, генерация текста, классификация, открытая беседа и извлечение информации.

В отличие от предыдущих рекуррентных нейронных сетей, которые последовательно обрабатывают входные данные, трансформеры обрабатывают целые последовательности параллельно. Это позволяет специалистам по обработке данных использовать графические процессоры для обучения LLM на основе трансформеров, что значительно сокращает время обучения.

Архитектура нейронной сети трансформера позволяет использовать очень большие модели, часто с сотнями миллиардов параметров. Такие сверхбольшие модели могут получать огромные объемы данных.

LLM, такой как GPT-3, может учитывать миллиарды параметров и генерировать контент, используя небольшое количество входных данных. Кроме того, изучая данные в масштабе Интернета во всех их различных формах и множестве закономерностей, большие языковые модели учатся применять свои знания в самых разных контекстах. Ниже представлена схема работы генеративной модели на рисунке 1.

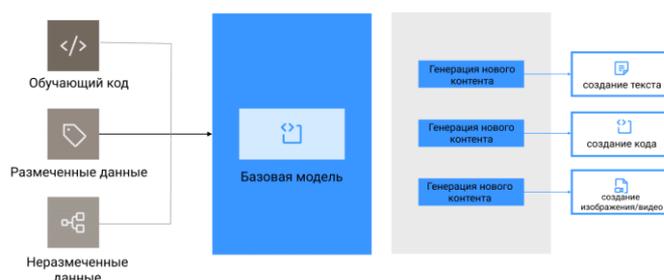


Рисунок 1 – Схема работы генеративной модели

Большие языковые модели невероятно гибкие. Одна модель может выполнять совершенно разные задачи, такие как ответы на вопросы, обобщение документов, языковые переводы и составление предложений. LLM могут кардинально повлиять на создание контента и использованию людьми поисковых систем и виртуальных помощников.

Хотя LLM и не идеальны, они демонстрируют способность делать точные прогнозы на основе относительно небольшого количества подсказок или входных данных. LLM можно использовать для генеративного искусственного интеллекта для создания контента на основе подсказок к вводу на естественном языке.

Приложения на основе генеративных моделей привлекают всеобщее внимание и поражают воображение. С их помощью можно изменить формы взаимодействия с клиентами и разработать новые, невиданные ранее приложения, а также помочь клиентам достичь невероятных показателей производительности.

Генеративные модели способны генерировать различный контент, от изображений до фильмов и сериалов. Что в теории позволяет запустить производство сериала, который будет создаваться в режиме реального времени без необходимости в персонале, кроме технического специалиста. Они могут длиться годами и десятилетиями без остановки. Также можно создать модель, которая может встраиваться в видеостриминговые сервисы и может создавать сериалы, адаптированные под вкус конкретного зрителя. Что может сделать опыт каждого пользователя уникальным. Например, в данный момент транслируется сериал *Nothing, Forever*, запущенный в тестовом режиме и автоматически генерируемый генеративной моделью, где участие человека минимально. В данный момент генеративные модели уже встроены в популярный видеосервис такой, как Youtube, что позволяет ему подбирать видео по интересам пользователя и предоставлять каждому пользователю уникальный пользовательский опыт.

Генеративные модели обычно обучены на огромном количестве статей и книг, что позволяет им понимать контекст сообщения, что позволяет им отлично переводить тексты на различных языках. Модели находят применения во многих переводчиках таких, как Google Translate, Яндекс Переводчик, Bing Microsoft Translator. Однако профессиональные переводчики необходимы для перевода текста из специализированных отраслей, имеющих сложный для генеративной модели контекст. Если привести в пример Google Translate, то внедрение генеративной модели-трансформера позволило достичь среднего прироста BLEU (метрика качества перевода) на 5 пунктов по сравнению с предыдущими моделями, данный прирост последний раз был достигнут 4 года назад при переходе от фразового перевода к нейросетевому [2].

Модели умеют хорошо анализировать данные, что может позволить находить последовательности в исследовательских данных. Их может помочь при разработке новых ферментов, антител и генной терапии. Их можно использовать для моделирования синтетических данных о пациентах, что необходимо для моделирования различных клинических испытаний или изучения редких заболеваний, что может помочь для создания лекарств и вакцин.

Также их можно применять для анализа рентгеновских снимков, МРТ, КТ, что поможет выявить признаки заболевания. Модели можно обучить на множестве данных с изображениями органов больных и здоровых людей, что позволит выявлять патологические ткани, определить стадию заболевания, спрогнозировать риск развития различных осложнений. Также в данный момент созданы чат-боты, которые могут помочь пациентам самостоятельно ставить диагноз, предоставляя информацию о симптомах. Например, Woebot Health умеет производить приведенную операцию. Однако компания-разработчик заявляет, что ее компьютерный ассистент не ставит точный диагноз, потому что порой люди могут ввести чат-бот в заблуждение, путем предоставления неверных показаний, а также следует учитывать некоторую несовершенство генеративных моделей [3].

В промышленности генеративные модели можно использовать для различных целей. Например, они могут использоваться при проектировании различных деталей, продукции для разных целей: уменьшения веса конструкции, уменьшения аэродинамического сопротивления, увеличения прочности и другого. В данный момент NASA использует генеративные модели при проектировании деталей космических кораблей, что позволяет снизить коэффициенты напряжения почти в 10 раз ниже, чем у деталей, созданных профессионалом. Данный вид проектирования называется генеративным проектированием. При использовании генеративного проектирования инженеру необходимо задать размеры детали, места креплений, требуемые нагрузки и другие параметры, после чего модель рассчитает оптимальную форму детали и произведет отбор модели детали с соблюдением необходимых параметров [4].

Генеративные модели могут быть использованы для анализа изображений готовой продукции для выявления дефектов и брака, что позволяет улучшить контроль качества и снизить количество брака, повысить доходность предприятия. Для улучшения качества продукции также можно моделировать ситуации для тестирования продукции в различных ситуациях.

В сфере обслуживания генеративные модели могут быть использованы для разработки чат-ботов, голосовых и виртуальных помощников для посетителей сайта, что позволит повысить качество обслуживания клиентов и увеличить их вовлеченность. В данный момент существует сервис HelpDeskEddy, который помогает настроить собственного чат-бота на основе GPT, нужного для

конкретной организации. Что позволяет создать чат-бот, который часто может решать различные вопросы, возникающие у клиентов, без участия оператора.

С помощью моделей можно создать рекомендательный сервис, который будет анализировать покупки потребителя, отзывы, понравившиеся позиции, просматриваемые товары, содержимое корзины, категории просматриваемых товаров, ценовой диапазон просматриваемых товаров и другое, чтобы рекомендовать товары и услуги, которые им нужны. На рисунке 2 представлена схема работы рекомендательного сервиса. Такие рекомендательные сервисы используют такие приложения как Wildberries, Ozon, Amazon и другие.

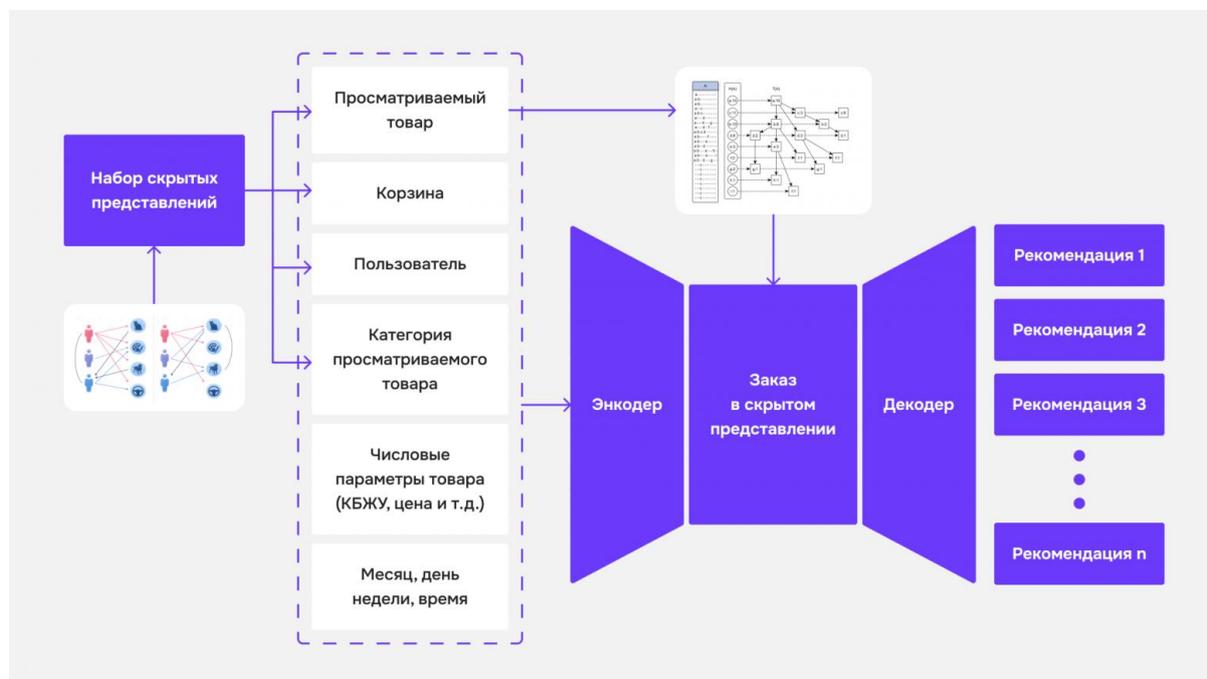


Рисунок 2 – Схема работы рекомендательного сервиса

В сфере информационных технологий генеративные модели совершили революцию путем автоматической генерации кода. Модели могут генерировать код из простых описаний требуемого функционала, что увеличивает скорость разработки и посмотреть на несколько решений одной проблемы, что позволяет посмотреть на задачу под разными углами. Генерация кода активно внедряется в инструменты разработки, тем самым позволяет использовать генеративные модели при написании кода в реальном времени. Например, Copilot, используемый в Visual Studio помогает программистам писать код на основе огромной выборки. Он способен рекомендовать код на основе написанной программы, завершать строки кода, которые программист уже начал писать, генерировать целые блоки, которые разработчик может использовать. Также он всегда проверяет код, который генерирует, что позволяет свести ошибки генерации к минимуму. Copilot – это мощный инструмент, которые позволяет повысить продуктивность и эффективность разработчика, однако он не может быть заменой программиста, так как не может заменить знания и опыт человека.

Модели способны искать ошибки в коде, такие как опечатки, синтаксические и логические ошибки. Однако на данный момент генеративные модели не могут гарантировать, то что они найдут все ошибки в коде, поэтому все еще необходимо вмешательство человека для поиска ошибок. Но можно с уверенностью утверждать, что это крайне полезные инструмент для первичного поиска ошибок в коде. Что может позволить ускорить поиск и устранение ошибок, но требует контроля со стороны специалиста.

При составлении документации приложения модели крайне полезны, они позволяют составлять документацию из исходного кода, после чего специалист проверяет правильность написания документации и вносит некоторые коррективы. Что позволяет уменьшить затраченное время на ее написании и позволяет программисту сфокусироваться на более важных и сложных задачах.

Для тестировщиков генеративные модели представляют собой инструмент для генерации разнообразных тестовых данных, которые помогают комплексно проверить приложение на наличие ошибок, что позволяет повысить надежность и отказоустойчивость приложения. Также они упрощают написание тестов путем их генерации. Кроме того, в данный момент имеется возможность генерировать тесты, используя Copilot, который встроен в Visual Studio, что помогает быстрее писать типовые тесты.

Генеративные модели – это быстроразвивающаяся область исследований. Совершенствуя модели можно будет расширить области обучения, увеличить наборы входных данных, что дает большое количество возможностей для множества сфер [5]. Это мощный инструмент, который помогает генерировать данные, теоретически моделировать различные данные, производить их анализ и многое другое. Однако следует учитывать потенциальные риски при работе с ними и проверять данные, которые они производят. Однако наиболее важный потенциал генеративных моделей в том, что в процессе обучения мы наделяем компьютер пониманием нашего мира, что до этого было просто невозможно.

Список использованных источников:

1. Amazon AWS / <https://aws.amazon.com/ru/>
2. 4. Google Research / <https://blog.research.google/2020/06/recent-advances-in-google-translate.html>
3. Woebot Health / <https://woebothealth.com/?refer=https%3A%2F%2Faicenter.ai>
4. 3D news / <https://3dnews.ru/1082005/v-nasa-nachali-ispolzovat-generativnoe-proektirovanie-detaley-kosmicheskikh-korabley-no-ne-speshat-delat-eto-na-postoyannoy-osnove>
5. Habr / <https://habr.com/ru/companies/wunderfund/articles/334568/>

UDC 004.8

AREAS OF USE OF GENERATIVE MODELS

*Lepeshko R.O.*¹

Belarusian State University of Informatics and Radioelectronics¹, Minsk, Republic of Belarus

Markov A.N. - Senior Lecturer

Annotation. This article explores the capabilities of generative models. Generative models are able to mimic human intelligence, allowing them to successfully analyze and create data. They are used in various fields from the service sector to the design of spacecraft parts.

Keywords. Generative model, information, information technology, spheres, generation, analysis.