

Министерство образования Республики Беларусь
Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК _____

Рудь Максим
Витальевич

Программный модуль синтеза сюжетов сказок

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1-40 80 02 «Системный анализ, управление и обработка
информации»

Научный руководитель

Герман Олег Витольдович

кандидат техн. наук, доцент

Минск 2019

ВВЕДЕНИЕ

Автоматическая генерация историй имеет давние традиции в области искусственного интеллекта. Возможность создавать истории по запросу имеет огромный потенциал для сфер развлечения и образования. Одним из примеров могут быть современные компьютерные игры, которые становятся все более захватывающими, содержащими множество сюжетных линий и сотни персонажей. Это существенно увеличивает объем работы, необходимой для производства каждой игры. Позволив игре написать свою собственную сюжетную линию, разработчик может снять с себя эту необходимость, а игра останется привлекательной для игрока.

Способность генерировать истории автоматически также актуальна в образовании, где истории используются для обучения детей. В настоящее время создание историй опирается на авторов-людей, а процесс написания требует значительных усилий и времени, чтобы истории были информативными и интересными. Написание компьютерных историй может значительно снизить зависимость от авторов-людей.

Искусство рассказывания историй – это то, что люди часто принимают как должное, но оно формирует неотъемлемую часть нашего культурного наследия. Мы сталкиваемся с историями в большом количестве повседневных ситуаций. Мы используем их, чтобы общаться друг с другом, чтобы создавать примеры для обучения, однако, достаточно сложно определить навыки, необходимые для создания и понимания простой истории. Взять, к примеру, известную басню «Заяц и черепаха», приписываемую Эзопу:

“Черепаха и заяц спорили, кто из них быстрее. Назначили они для состязания время и место и разошлись. Но заяц, полагаясь на свою природную резвость, не старался бежать, а улегся возле дороги и заснул. А черепаха понимала, что двигается она медленно, и потому бежала без передышки. Так обогнала она спящего зайца и получила победную награду.”

Упрощенная природа этого текста классифицирует его как детскую историю, но необходимо учитывать, что чтение этой истории включает сложные мыслительные процессы. Помимо языковой обработки, эта история требует, чтобы мы делали выводы, основываясь на нашем знании, что зайцы быстрые, а черепахи медленные. Мы должны правильно интерпретировать фабулу басни из сюжета. Это часто включает в себя способность «смотреть между строк», используя умозаключения, мировые знания и личный опыт, чтобы заполнить пробелы в повествовании. Мы также должны учитывать, что басня содержит моральные сообщения и то, как ситуация, которую

разыгрывают персонажи истории, применима к нашей собственной жизни, учитывая наш собственный личный опыт. В результате это делает задачу понимания и создания истории очень сложной для компьютеров.

Есть несколько общих проблем, с которыми столкнется любой гипотетический генератор историй. Возможно, наибольшее беспокойство вызывает количество мировых знаний, необходимых для создания хороших историй. По этой причине предыдущие системы были в основном изготовлены вручную. Количество требуемых знаний варьируется, но большинство подходов, как правило, требуют информации о сущностях, появляющихся в истории, о том, как они взаимодействуют и как история может быть структурирована, чтобы включить их. Вся эта информация должна быть полной, чтобы алгоритмы создавали успешные планы. Использование данных, созданных вручную, устанавливает предел для доменов, которые может охватывать генератор историй. Процесс создания необходимой базы знаний для этих систем будет трудоемким, что устраняет любые стимулы для их расширения для охвата новых жанров.

Многие компоненты истории, такие как персонажи и темы, повторяются в аналогичных формах в других историях. Например, практически в каждой истории присутствуют Герой или Злодей, и, хотя они, возможно, не будут одинаковыми в каждой истории, у них можно выделить много схожих свойств. Именно эту избыточность мы хотим использовать для создания новых историй, обнаруживая закономерности в большой выборке созданных вручную. Вместо того, чтобы вручную кодировать всю информацию в системе для каждого жанра, мы хотим, чтобы система получала эту информацию из историй, которые она уже видела. Это позволит портировать генератор историй в разные жанры без необходимости ручного кодирования.

Есть много различных жанров историй, которые можно было бы генерировать (романы, сценарии или даже детективы). Мы хотим начать с простого, чтобы увидеть, является ли наша гипотеза осуществимой. По этой причине мы решили сосредоточиться на сказках. Сказки предлагают несколько преимуществ в качестве основы для технико-экономического обоснования. Во-первых, словарный запас и содержание каждой истории будут ограничены, поскольку они должны соответствовать знаниям, которыми владеют дети. Во-вторых, они имеют большой потенциал для образовательных интерактивных систем. У этих систем обычно есть вводная история, которую дети стремятся развивать. Мы считаем, что, если бы такая система могла генерировать истории сама, она могла бы дать детям гораздо более богатые примеры, когда они пытаются написать свои собственные.

ОБЩАЯ ХАРАКТЕРИСТИКА

В работе рассматривается проблема автоматической генерации историй.

Целью настоящей работы является создание программного модуля синтеза сюжетов сказок с использованием обучаемых компонентов вероятностной генерации естественного языка (NLG) и оценочных функций ранжирования на основе корпуса сказок.

Задачами настоящей работы являются:

- выполнение анализа существующих подходов, подходящих для реализации целей и задач настоящей работы;

- изучение системы вероятностной генерации естественного языка (NLG);

- извлечение сущностей, синтаксических конструкций и грамматических отношений из корпуса сказок;

- проектирование обучаемых компоненты NLG стохастически, без обширного ручного кодирования;

- разработка двух оценочных функций для оценки согласованности и интереса сгенерированных историй.

Общими требованиями к программному модулю являются:

- использование свободно распространяемого программного обеспечения с открытым исходным кодом – система должна иметь возможность свободного использования и доработки без лицензионных ограничений для учебных, образовательных и научных целей;

- масштабируемость системы – возможность переобучения программного модуля для различных корпусов, отличающихся по лексике, стилю, жанру и области;

- высокая производительность и отказоустойчивость системы, достигнутая с помощью отсеивания заведомо неудачных историй и ограничения пространства возможных историй.

Заключение

В результате работы были выполнены следующие задачи:

- на начальном этапе был выполнен анализ существующих подходов к проблеме генерации историй, были сформулированы их основные достоинства и недостатки;

- изучена система вероятностной генерации естественного языка (NLG), её структура и модули;

- на основании полученной информации были спроектированы обучаемые компоненты NLG;

- был реализован программный модуль синтеза сюжетов сказок на основе спроектированных компонентов NLG и корпуса сказок с динамическим созданием сюжета;

- были разработаны две оценочные функции, которые оценивают сгенерированные истории с точки зрения интереса и согласованности.

Результаты показывают, что подход, основанный на генерации и ранжировании, является жизнеспособным в создании рассказов, которые демонстрируют повествовательную структуру. Поскольку программный модуль может быть легко переобучен для разных корпусов, он может генерировать истории, которые различаются по лексике, стилю, жанру и области.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Рудь М.В. Вероятностная генерация естественного языка // Студенческий вестник: научный журнал. № 21(71). – М., Изд. «Интернаука», 2019.