**P. A. Nosulenko\***

Student of the Department of Applied Informatics

**D. V. Butenina\***

Кандидат физико-математических наук, доцент

\* St. Petersburg state University of aerospace instrumentation

**DEVELOPMENT OF A SYSTEM FOR VIDEO SUBTITLING**

**Abstract**

The article is devoted to the development of an automated system for creating subtitles for video content based on neural networks. For successful operation of the system it is planned to develop a complex of several neural networks.

**Keywords:** quality, metrology, artificial intelligence, recurrent neural networks, subtitles.

УДК 008.4

**П. А. Носуленко\***

Студент кафедры прикладной информатики

**Д. В. Бутенина\***

Кандидат физико-математических наук, доцент

\*Санкт-Петербургский государственный университет аэрокосмического приборостроения

**РАЗРАБОТКА СИСТЕМЫ ДЛЯ СОЗДАНИЯ СУБТИТРОВ К ВИДЕО**

**Аннотация**

Статья посвящена разработке автоматизированной системы создания субтитров для видеоконтента на основе нейронных сетей. Для успешной работы системы планируется разработать комплекс из нескольких нейронных сетей.

**Ключевые слова:** качество, метрология, искусственный интеллект, рекуррентные нейронные сети, субтитры.

Использование видеоконтента становится все более популярным, что создает потребность в субтитрах. Ручное создание субтитров трудоемко и времязатратно, поэтому возрастает необходимость в разработке автоматизированной системы создания субтитров. Система сможет облегчить процесс, увеличить его скорость и обеспечить высокое качество субтитров, делая видео контент более доступным для широкой аудитории, включая людей с нарушениями слуха или изучающих иностранные языки.

Для нашей задачи нужно разработать сложную систему из нескольких нейронных сетей, каждая из которых отвечает за отдельную задачу. На сегодняшний день существует большое количество архитектур нейронных сетей под самые разные задачи [1].

Одной из таких архитектур является архитектура рекуррентной нейросети. Рекуррентные нейронные сети — популярные модели, используемые в обработке естественного языка [2]. Во-первых, они оценивают произвольные предложения на основе того, насколько часто они встречались в текстах. Это дает меру грамматической и семантической корректности. Такие модели используются в машинном переводе [3]. Во-вторых, языковые модели генерируют новый текст.

Для работы нейронной сети, ее нужно сперва обучить на собранном до этого наборе данных или датасете. Целью сбора данных является создание разнообразного набора записей, включающего различные голоса, акценты, темп речи и фоновые шумы [4]. Дополнительно к аудиофайлам, нам также понадобятся текстовые файлы, где будет указано, что говорится в каждой записи, кто это говорит и тайминги предложений. Эти текстовые файлы будут использоваться для связи аудиофайлов с точными транскрипциями.

Задача первой нейронной сети – распознавание и предобработка речи. Содержание видеофайла не важно, это может быть фильм или научная лекция. На видео могут присутствовать посторонние звуки, например, шумы и музыка, люди могут говорить с различным акцентом, скоростью и дикцией, спикеров может быть много. Первая часть системы обрабатывает полученный аудиофайл и возвращает текст, с таймингом, когда этот текст произносился. Однако если просто получить из потока речи текст, то получится просто набор слов, без запятых и точек, без логической группировки слов в предложения.

Поэтому вторая нейросеть должна будет обработать полученный набор слов, удалить мусор, расставить знаки препинания, составить предложения. Для такой задачи восстановления пунктуации достаточно взять текстовый корпус, исходными данными будет идти текст без пунктуации и на нем нейросеть обучится её восстанавливать. На вход нейросети текст поступает в токенизированном виде. Такое разбиение не слишком мелкое, чтобы длина последовательности не сильно увеличилась, но и не слишком крупное, чтобы избежать проблемы, когда токена нет в словаре. На выходе модели после каждого слова метка: ставить ли тот или иной символ пунктуации.

Третья нейросеть – заключительная, ее задача просто перевести полученный текст на нужный язык. На этом этапе все стандартно, при переводе предложение разбивается на словарные сегменты. Затем с помощью специальных декодеров система определяет «вес» каждого сегмента в тексте. Далее вычисляется максимально вероятные значения и перевод сегментов. Последний этап — соединить переведённые сегменты с учётом грамматики.

Таким образом мы имеем систему способную создать субтитры к видео для комфортного просмотра. Система возвращает файл с текстом и таймингами, а также сохраняет смысл и грамотно расставляет пунктуацию. Однако даже такая нейросеть еще далека до совершенства и существует множество путей ее развития.

**Библиографический список:**

1) Разнообразие нейронных сетей. // [Электронный ресурс] URL: https://tproger.ru/translations/neural-network-zoo-2 (дата обращения: 01.11.2023)

2) Рекуррентные нейронные сети: типы, обучение, примеры и применение // [Электронный ресурс] URL: https://neurohive.io/ru/osnovy-data-science/rekurrentnye-nejronnye-seti/ (дата обращения: 21.11.2023)

3) Н. С. Москалев. Виды архитектур нейронных сетей // Молодой учёный №29, 2016 г.

4) Как Яндекс помогает преодолеть языковой барьер: нейросетевой перевод видео, картинок и текста // [Электронный ресурс] URL: https://habr.com/ru/companies/yandex/articles/576438/ (дата обращения: 22.11.2023)