

ОТЗЫВ

на выпускную квалификационную работу магистра Доржиева Зоригто Жаргаловича
«Анализ аудиоданных и распознавание событий для систем безопасности»

Работа Доржиева З.Ж. посвящена актуальной проблеме распознавания звуковой информации. В настоящее время большие успехи в этой области достигнуты в распознавании речи, однако распознавание звуковых событий различной природы все еще далеко от совершенства.

Цель работы – разработка и экспериментальное исследование методов классификации аудиоинформации, способных распознавать в реальном времени в аудиоданных один или несколько заранее заданных классов звуковых событий.

Поскольку любые классификаторы требуют для своего обучения большие наборы размеченных данных, то в работе подробно описывается наиболее представительный набор аудиоданных AudioSet, содержащий широкий спектр звуков человека и животных, музыкальных инструментов и жанров, а также повседневных звуков окружающей среды. Автор описывает разработанную им процедуру получения из этого датасета данных, необходимых для своего исследования.

В работе достаточно подробно описываются различные характеристики аудиосигнала, которые могут быть использованы в качестве признаков для классификации. Затем предлагается алгоритм предварительной обработки аудиосигнала, с помощью которого формируется размеченный массив извлеченных признаков для всех полученных ранее данных. Дополнительно проводится разметка каждого класса, который рекурсивно проверяется на наличие подклассов, после чего создаётся соответствующая иерархия. Этот структурированный массив признаков используется затем для обучения классификаторов.

Автор рассматривает результаты применения классических классификаторов (KNN, SVM) и предлагает исследовать классификаторы, построенные на основе сверточных нейронных сетей. В работе предлагается две архитектуры неронной сети. Первая — достаточно простая сверточная сеть на основе архитектуры Ле-Куна, состоящая из двух сверточных слоев с max-pooling и полносвязного слоя. При этом на основе анализа особенностей звукового сигнала автор предложил использовать окна сверток большего размера по сравнению с исходной архитектурой. В работе также рассматриваются методы, использующиеся при обучении сетей, позволяющие повысить их эффективность и качество распознавания. Вторая модель имеет VGG-подобную архитектуру. При этом в дальнейшем исследовались две модификации этой архитектуры.

Доржиев З.Ж. реализовал предложенные модели на языке Python с помощью фреймворка PyTorch, используя свободные реализации соответствующих архитектур нейронных сетей. Обучение моделей и эксперименты с ними проводились в среде Google Colab, которая в том числе предоставляет удаленный доступ к GPU ресурсам. После обучения моделей на наборе данных AudioSet их веса сохранялись и были использованы для работы с другим, менее универсальным датасетом Urban Sound Classification, содержащим звуки городской среды.

В работе приводятся экспериментальные данные по оценке качества для всех исследуемых моделей, как для первого, так и второго набора данных. Их анализ показал, что модели построенные по архитектуре VGG имеют более высокие показатели. Однако для всех рассмотренных моделей они снижаются для второго небольшого и менее универсального набора данных, но более подходящего для распознавания звуковых событий в системах безопасности. При этом полученные показатели качества для первой архитектуры не уступают показателям качества традиционных методов. Модели на базе VGG архитектуры позволили достичь существенно лучших значений показателей качества распознавания. При этом классификация

событий на аудиозаписях происходит со скоростью, достаточной для работы в реальном времени.

В ходе работы Доржиев З.Ж. показал хорошую теоретическую подготовку, квалификацию в области глубоких нейронных сетей и способность применять полученные знания на практике.,

Считаю, что работа Доржиева З.Ж. заслуживает оценки **“Отлично”**.

Научный руководитель

доцент кафедры КММС, к.т.н.



В.М. Гришкин