



го анализа казахского языка / Вестник Новосибирского гос. университета. Серия: Информационные технологии, 2014. – Т.12.– № 2. – С. 25-32.

3. Тукеев У.А., Тургынова А. Морфологический анализ казахского языка на основе полной системы окончаний / Труды международной конференции по компьютерной и когнитивной лингвистике. Сер. "Интеллект. Язык. Компьютер" 2016. – С. 225- 231.

4. Мамырбаев О. Ж., Хайрова Н. Ф., Мухсина К. Ж. Қазақ тіліндегі мәтіндердегі қылмыстық мәнді коллакцияларды анықтау / Вестник КазАТК им. М. Тынышпаева, рекомендуемый ККСОН МОН РК. – №3(110). – 2019. – 170 -175 с.

5. Тузов В.А. Компьютерная семантика русского языка. — СПб.: Изд-во СПбГУ, 2004. — 400 с.

6. Қазақ әдеби тілінің сөздігі. Он бес томдық. / Құраст. Т.Жанұзақ, С.Омарбеков, Ә.Жүнісбек және т.б. – Алматы, 2011.

7. «Қазақ тілінің омонимдер сөздігі», М.Белбаева, 1988 – 193 с.

Д.С. Лебедева, А.И. Максимов

ДИАГНОСТИКА ФАЛЬСИФИКАЦИИ ФОНОГРАММ НА ОСНОВЕ СИАМСКИХ СЕТЕЙ

(Самарский университет)

Введение

Поскольку в настоящее время фонограммы активно используются как доказательство в судебных процессах, задача проверки аутентичности аудиозаписей в рамках криминалистической экспертизы является злободневной и актуальной. Под криминалистической экспертизой звукозаписи [1] понимается исследование магнитных или каких-либо других записей звука с целью установления фактов, имеющих доказательственное значение.

В данной работе исследуются возможности применения методов нейронных сетей и машинного обучения [2] для решения поставленной задачи при помощи сравнения фрагментов фоновых шумов аудиозаписей.

Фоновый шум [1] является частью общего шума, поступающего от подвижных или стационарно расположенных источников, при отключении известных источников.

В случае реальной диагностики цифровой фонограммы с использованием сравнения фрагментов фоновых шумов, эксперту необходимо выделить в анализируемой аудиозаписи фрагменты шума. В случае существенного различия метрик, характеристик, параметров взятых фрагментов шума, можно заключить, что в аудиозаписи присутствует вставка – ее части записаны в различных условиях. Для исследовательских целей упростим постановку задачи – пусть имеется пара фрагментов фонограмм фонового шума, необходимо установить, записаны ли эти фрагменты в одинаковых условиях.



Формирование датасета для исследования

Для данного исследования был самостоятельно создан тестовый датасет, представляющий из себя набор фрагментов фонограмм, полностью состоящих из фоновых шумов. Фоновые шумы были записаны в различных условиях – с использованием различных записывающих устройств, в различных помещениях и в различное время суток. Датасет состоит из 137 фонограмм, из которых образовано 9316 пар аудиозаписей. 1544 пары содержат записи, полученные в идентичных условиях, 7772 пары, соответственно, содержат записи, полученные в различных условиях. При сборе данного датасета было использовано 3 записывающих устройства, запись производилась в 2 различных помещениях в различное время суток.

Для выделения характеристик фрагментов фонограмм была использована мел-спектрограмма. Из аудиозаписей формировались мел-спектрограммы размером 128x128 для подачи на вход нейронной сети.

Поскольку мел-спектрограмма шумового сигнала является двумерным сигналом (изображением), было решено исследовать возможности синергии предназначенных для обработки изображений нейросетевых моделей и обработки звуковых сигналов.

Выбор архитектуры нейронной сети

Для решения выбранной задачи подходит архитектура сиамских нейронных сетей [3]. Сиамская нейронная сеть состоит из двух идентичных нейронных подсетей, которые принимают на вход различные данные. Сиамская сеть высчитывает отображение входных данных в вектора, считает расстояние между ними и функцию потерь, после чего оценивает различие между входными данными. Сиамские сети используются для так называемого one-shot распознавания, когда делается предсказание на основе одного экземпляра для каждого класса.

В данной работе нейросеть строилась на основе модели VGG16 [4]. Ее архитектура представлена на рисунке 1.

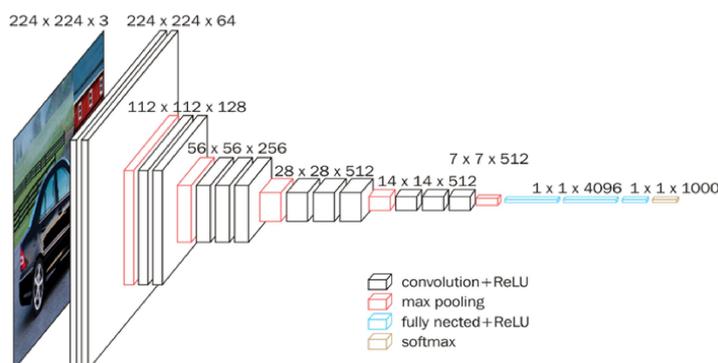


Рис.1. Архитектура сети VGG16

Данная модель содержит в себе 16 слоев, два последних слоя которой отвечают непосредственно за классификацию входных изображений.

Для подсетей сиамской сети были взяты две идентичные предобученные сети VGG16 без последних двух слоев. На выходе из подсети получается вектор-дескриптор исходных данных размерность 4096. Дополнительно была про-



изведена перенастройка подсетей с целью корректно подать на вход данные: был изменен входной тензор, чтобы нейросеть принимала на вход матрицу 128×128 .

Архитектура полученной сиамской нейронной сети представлена на рис. 2:

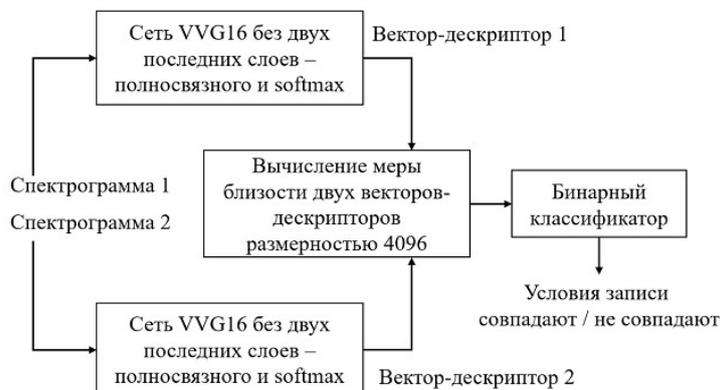


Рис. 2. Архитектура разработанной для исследования сиамской нейросети
Экспериментальное исследование методов классификации пар векторов-дескрипторов

Как видно из рисунка 1, после получения двух векторов-дескрипторов, следуют этапы вычисления меры близости векторов и классификация пары на два класса – «условия записи совпадают» либо «условия записи не совпадают». Данный этап работы модели может быть реализован при помощи разных методов. В ходе данной работы было проведено исследование таких методов.

Одним из наиболее простых вариантов сравнения дескрипторов является вычисление евклидова расстояния между ними. На основе евклидова расстояния был обучен бинарный классификатор [2]. В результате валидации модели были получены следующие значения:

1. Значение accuracy = 0,83,
2. Значение f1-score = 0,46.

Также сравнение пары векторов-дескрипторов с последующей ее классификацией на два класса можно реализовать при помощи многослойного перцептрона (Multilayer perceptron, MLP) [5]. В данном исследовании на вход перцептрона подавалась конкатенация двух векторов-дескрипторов. В ходе валидации полученной модели были получены следующие результаты:

1. Значение accuracy = 0,88,
2. Значение f1-score = 0,60.

Аналогичным образом было реализовано применение градиентного бустинга (Gradient boosting) [6]. В результате были получены следующие значения:

1. Значение accuracy = 0,86,
2. Значение f1-score = 0,40.



Также была исследована возможность применения случайного леса (Random forest) [7] к решению данной задачи. В результате тестирования такой модели были получены следующие результаты:

1. Значение accuracy = 0,87,
2. Значение f1-score = 0,47.

Заключение

В данной работе рассмотрена задача диагностики фальсификации фонограмм путем сравнения фрагментов их фоновых шумов. Для ее решения была предложена новая архитектура сиамской сети на основе модели VGG16.

В ходе работы проведено сравнительное исследование методов классификации пар векторов-дескрипторов, полученных из подсетей сиамской сети. Исследование проведено на собственном датасете фонограмм фоновых шумов.

Как видно из приведенных данных, лучшие результаты продемонстрировало использование MLP для сравнения векторов-дескрипторов.

Литература

1. Каганов А.Ш. Криминалистическая экспертиза звукозаписей / А.Ш. Каганов. – М.: Юрлитинформ, 2005. – 272 с.
2. Рашка С. Python и машинное обучение / Рашка С.; пер. Логунов А.В.; ред. Мовчан Д.А. – ДМК-Пресс, 2017. – 418с.
3. I. Melekhov. Siamese network features for image matching/ I. Melekhov, J. Kannala, E. Rahtu// XXIII International Conference on Pattern Recognition. -2016. - p. 378-383.
4. T. Srikanth. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images// International Journal of Scientific and Research Publications. – 2019. –Vol 9. –p. 143-150.
5. D.E. Rumelhart. Learning representations by back-propagating errors / D.E. Rumelhart, G.E. Hinton, R.J. Williams // Nature. –1986. – Vol.323. – p.533-536.
6. J.H. Friedman. Stochastic gradient boosting // Computational Statistics & Data Analysis. –2002. –Vol.38(4). – 367-378.
7. Т.К. Хо. Random Decision Forests // Proceedings of the 3rd International Conference on Document Analysis and Recognition. –1995. – p. 278–282.