**Исследование методов диагностики фальсификации фонограмм путем сравнения фрагментов фоновых шумов**

***Д.С. Лебедева, А.И. Максимов***

***Самарский Национальный Исследовательский университет имени академика С.П. Королёва,***

***кафедра геоинформатики и информационной безопасности,***

***факультет информатики***

**Аннотация**

В работе рассматривается задача проверки наличия в фонограмме вставок при помощи анализа фрагментов ее фоновых шумов. Исследуются методы проведения такого анализа на основе статистических критериев, методов машинного обучения, а также при помощи нейронных сетей.

**Введение**

Поскольку в настоящее время фонограммы активно используются как доказательство в судебных процессах, задача проверки аутентичности аудиозаписей в рамках криминалистической экспертизы [1] является злободневной и актуальной. Под криминалистической экспертизой звукозаписи понимается исследование магнитных или каких-либо других записей изображения или звука с целью установления фактов, имеющих доказательственное значение, и составление по данным указанного исследования экспертного заключения для использования последнего в судопроизводстве. [1]

Под аудиозаписью или фонограммой понимается звуков исполнения или других звуков либо отображения звуков, кроме звуков в форме записи, включенной в кинематографическое или иное аудиовизуальное произведение [2].

В данной работе исследуются возможности применения методов обработки сигналов и математической статистики к решению задачи диагностики фальсификации цифровой фонограммы, основанных на сравнении фрагментов фоновых шумов. Были исследованы как отдельные статистические критерии согласия, так и их совокупность. Помимо статистических методов исследованы методы «классического» машинного обучения, а также искусственные нейронные сети.

**Сравнение фрагментов фоновых шумов**

Фоновым шумом является частью общего шума, поступающего от подвижных или стационарно расположенных источников, при отключении известных источников. [3]

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.

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 58 фонограмм, из которых образовано 1653 пар аудиозаписей. 253 пары содержат записи, полученные в идентичных условиях, 1400 пар, 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.

**Исследование использования статистических критериев**

Статистические критерии – это некое правило, используемое для проверки статистической гипотезы о виде распределения величины. [4]

В данной работе исследовалось 3 критерия согласия:

1. критерий согласия хи-квадрат Пирсона [5],
2. критерий Стьюдента [6],
3. критерий Манна-Уитни-Вилкоксона[7].

Статистические критерии применялись для решения задачи диагностики фонограмм следующим образом: для диагностируемой пары аудиозаписей формировалась спектрограмма [8][9], в рамках одного частотного среза фонограмм вычислялся один из трех выбранных статистических критериев, полученные значения усреднялись по всем частотным срезам.

Обобщенное математическое выражение для данной процедуры можно записать в следующем виде:

 (1)

где  - усредненный по всем частотным срезам критерий,

 - вычисляемый по частотному срезу статистический критерий,

 - обозначение, отвечающее за тип используемого статистического критерия,

- номер частотного среза,

 - число частотных срезов,

Также исследовалось усредненное значение полученных критериев:

 (2)

где  - усредненное значение статистических критериев,

 - усредненное по числу частотных срезов значение критерия Стьюдента,

- усредненное по числу частотных срезов значение критерия Манна-Уитни-Вилкоксона,

- усредненное по числу частотных срезов значение критерия хи-квадрат Пирсона.

В данном исследовании полагалось, что две фонограммы идентичны, если значение усредненного критерия превысило порог 0,78, полученный экспериментальным образом.

С учетом установленного порога, были получены следующие результаты точности работы статистических критериев:

1. критерий Стьюдента - 0,30,
2. критерия Манна-Уитни-Вилкоксона – 0,13,
3. критерий хи-квадрат Пирсона – 0,15,
4. среднее по трем критериям – 0,14.

**Исследование использования методов машинного обучения**

В данной работе также исследовалась возможность использование методов машинного обучения для решения задачи. Был построен бинарный классификатор, в который в качестве вектора признаков подавался вектор из значений статистических критериев - . Для обучения собранный датасет был разбит на обучающую и тестовую выборки в пропорции 7:3.

Для определения качества классификации, помимо точности, было использовано значение f-score [7].

 (3)

где *precision* - точность классификатора [10], показывающая, сколько из объектов, которые он отнес к этому классу, действительно относятся к этому классу.

*recall* – полнота [10], показывает, сколько объектов конкретного класса классификатор смог определить правильно.

В результате обучения классификатора были получены следующие значения:

1. значение точности = 0.30,

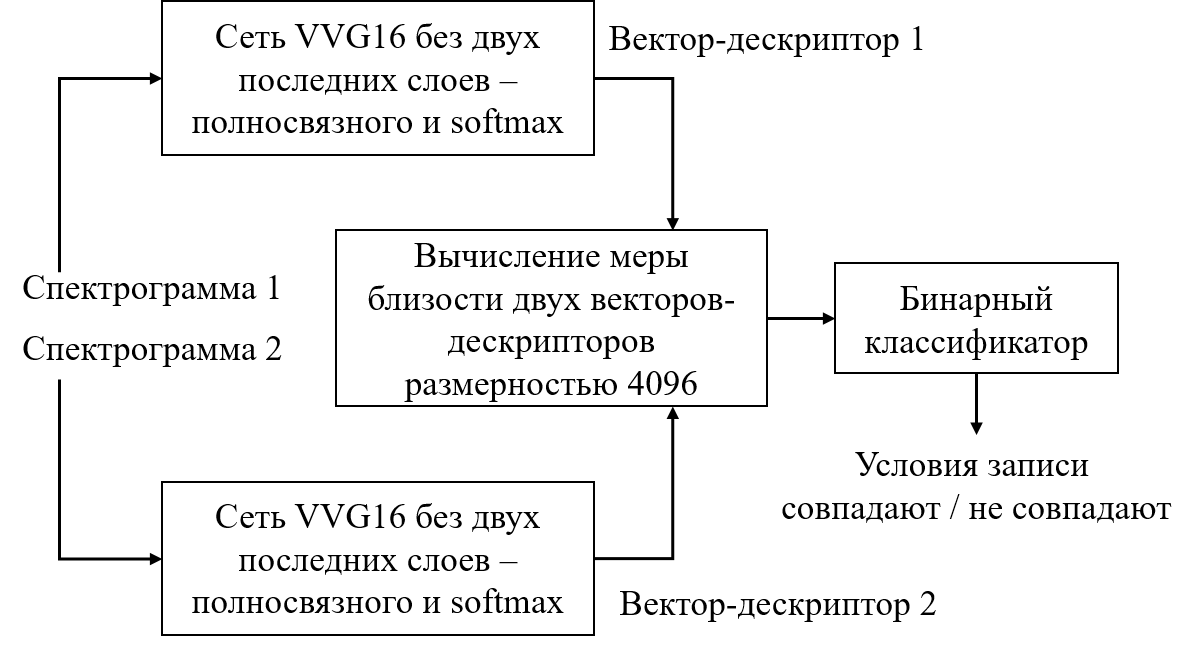
2. значение f-score = 0.08.

**Исследование использования нейронных сетей**

Поскольку применение классического машинного обучения не вызвало существенного повышения точности решения поставленной задачи, были исследованы возможности нейросетевых моделей.

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

Для решения выбранной задачи подходит архитектура сиамских нейронных сетей [11]. Сиамская нейронная сеть состоит из двух одинаковых нейронных подсетей, которые принимают на вход различные данные. Сиамская сеть высчитывают отображение входных данных в вектора, считает расстояние между ними и функцию потерь, после чего оценивает различие между входными данными [11]. Сиамские сети используются для так называемого one-shot распознавания [12], когда делается предсказание на основе одного экземпляра для каждого класса. Классическая архитектура сиамской нейронной сети представлена на рис. 1:



*Рис. 1. Архитектура разработанной для исследования сиамской нейросети*

В данной работе нейросеть строилась на основе модели VGG16 [13]. Данная модель содержит в себе 16 слоев, последние два слоя. На выходе каждой подсети получался вектор-дескриптор входной спектрограммы аудио сигнала размерностью 4096. Следующая задача – вычислить расстояние между полученными векторами, для исследования было выбрано евклидово расстояние. Полученные расстояния подавались на вход бинарному классификатору [14]. Для обучения собранный датасет был разбит на обучающую, тестовую и валидационную выборки в пропорции 6:3:1.

В результате валидации модели были получены следующие значения:

1. Значение точности = 0,77,

2. Значение f1-score = 0,31.

**Анализ результатов и направление дальнейших исследований**

Классификация на основании вектора признаков, полученного по результатам применения статистических критериев, показала значение точности 0,3 и меру f-score 0,08. Неудачу применения такого метода можно связать как с перекошенностью датасета (пар различных шумов больше, чем пар идентичных шумов), так и с тем, что выбранные признаковое пространство было достаточно примитивным.

Результат классификации на основе сиамской нейронной сети показал точность 0,77 – лучший результат среди исследованных методов. Данные результат планируется улучшить в будущих исследований. Для этого предполагается провести исследование нейросетевых моделей и метрик близости векторов-дескрипторов. Также планируется продолжать пополнять датасет фрагментов фоновых шумов.

**Заключение**

В данной работе рассмотрена задача диагностики фальсификации фонограмм путем сравнения фрагментов фоновых шумов. Были исследованы методы основанные на мат.статистике, машинном обучении и нейронных сетях. Исследования были проведены на самостоятельно собранном датасете.

Наилучшие результаты среди статистических критериев демонстрирует критерий Стьюдента с значением точности 0,3. Использование усреднения по трем критериям не оказывает существенного положительного эффекта.

Обученный на векторах из статистических критериев Байесовский классификатор продемонстрировал точность 0,3. Однако оценка качества его работы с помощью метрики f-score, которая является более подходящей для случая обучения на перекошенном датасете, показала низкое значение – 0,08.

Использование сиамской нейросети на основе модели VGG16 с использованием евклидового расстояния в качестве меры близости между векторами-дескрипторами спектрограмм продемонстрировала точность 0,77, что показывает перспективность дальнейших исследований таких ахитектур.

**Список литературы:**

1 Ищенко Е.П. Криминалистика. Учебник/ Ищенко Е.П., Топорков А.А. – М.: Инфра-М Контракт, 2010. – 781 с.

2 Каганов А.Ш. Криминалистическая экспертиза звукозаписей / А.Ш. Каганов. - М.: Юрлитинформ, 2005. - 272 c.

3 ГОСТ 23337-2014 Шум. Методы измерения шума на селитебной территории и в помещениях жилых и общественных зданий (с Поправкой).

4 Семенов В. А. С30 Теория вероятностей и математическая статистика: Учебное пособие. Стандарт третьего поколения. — СПб.: Питер, 2013. — 192 с.: ил.

5 Панков А.Р., Платонов Е.Н. Практикум по математической статистике: Учебное пособие. — М.: Изд-во МАИ, 2006.

6 Володин И.Н. Лекции по теории вероятностей и математической статистике [Текст]/И.Н. Володин – Казань: [б.и.], 2006. - 271с.

7 Берестнева О.Г. Прикладная математическая статистика: учебное пособие / Берестнева О.Г., Марухина О.В., Шевелёв Г.Е. – Томск: Изд-во Томского политехнического университета, 2012. – 200 с.

8 L. Wyse. Audio Spectrogram Representations for Processing with Convolutional Neural Networks // Proceedings of the First International Workshop on Deep Learning and Music joint with IJCNN. - 2017. - Vol.1. - No.1. - p. 37-41.

9 Ахмад Х. М. Введение в цифровую обработку речевых сигналов: учеб. пособие / Х. М. Ахмад, В. Ф. Жирков ; Владим. гос. ун-т. – Владимир : Изд-во Владим. гос. ун-та, 2007. – 192 с.

10 Гладин П.Е. Технологии машинного обучения[Текст]/ Гладин П.Е., Боченина К.О. – СПб:Университет ИТМО, 2020. -75с.

11 I. Melekhov. Siamese network features for image matching/ I. Melekhov, J. Kannala, E. Rahtu// XXIII International Conference on Pattern Recognition. -2016. - p. 378-383.

12 G. Koch. Siamese neural networksfor one-shot image recognition/ G. Koch, R. Zemel, and R. Salakhutdinov// ICML Deep Learning Workshop. -2015. –Vol 2.

13 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.

14 H. Almuallim. An Efﬁcient Algorithm for Optimal Pruning of Decision Trees//Artiﬁcial Intelligence - 1996. - p. 347-362.