

УДК 004.93

## ОФФЛАЙН ВАЛИДАЦИЯ ЦИФРОВЫХ ПОДПИСЕЙ НА ОСНОВЕ КОМБИНИРОВАННОГО АЛГОРИТМА С ИСПОЛЬЗОВАНИЕМ СТАТИСТИЧЕСКИХ ДАННЫХ И МАШИННОГО ОБУЧЕНИЯ

Бушев К.Р. (Университет ИТМО)

Научный руководитель – кандидат технических наук, доцент Муромцев Д.И.  
(Университет ИТМО)

Аннотация:

В данной статье рассматривается метод валидации цифровых подписей основанный на комбинированном подходе с использованием статистических данных извлеченных из изображения и дальнейшей их обработке алгоритмами машинного обучения. Проведено сравнение итоговых метрик, а так же даны рекомендации к использованию алгоритма с указанием сильных и слабых сторон подхода.

**Введение.**

Оффлайн алгоритмы валидации подписей работающие с высокой точностью, позволяющие определить поддельную подпись имеют высокую вычислительную сложность. В промышленных системах, при обработке десятков тысяч документов в час, существует необходимость отсекаания заранее поддельных или корректных подписей, для уменьшения количества запусков высоконагруженных процессов. В данной статье рассматривается метод позволяющий отсекаать чужие или очевидные поддельные подписи, что позволяет применять сложные алгоритмы реже.

**Основная часть.**

Так как в реальной системе изображения с подписями, как правило извлекаются из цифровых копий документов, необходимо произвести предобработку для уменьшения влияния внешних факторов на качество распознавания подписей. Процесс предобработки разделяется на несколько основных этапов:

- Очистка изображения от шумов
- Перевод изображения в бинарный формат
- Обрезка подписи по ее границе
- Выравнивание угла наклона подписи
- Изменение масштаба изображения
- Изменение толщины подписи

Так как хранение обучение алгоритма верификации подписи конкретного человека является  $n$ -мерной задачей, что накладывает ограничения на ее реализацию и последующее расширение при добавлении новых подписей. Так же данный вид классификации может применяться далеко не во всех системах, так как внутри него хранится информация о персональных данных человека, поэтому алгоритм представленный в данной работе является бинарным классификатором, который на основе двух подписей определяет принадлежат ли они одному и тому же человеку или написаны разными людьми.

Для верификации подписей используется входной вектор состоящий из объединения статистических данных двух векторов извлеченных из изображений подписей. В качестве признаков, находящихся в данных векторах выбраны следующие признаки:

- Масса подписи - количество черных пикселей во всей подписи
- Координаты центра масс подписи
- Направляющая угла наклона подписи
- Границы разбиения сегментов подписи - подпись разделяется на 10 равно массных отрезков по вертикали после чего проводится анализ каждого из них в отдельности
- Центр масс сегментов подписи
- Масса сегмента подписи

- Верхняя и нижняя граница сегментов подписи
- Направляющие угла наклона сегментов подписи

В итоге каждая из подписей представляется в виде вектора состоящего из 54 признаков, а итоговый вектор содержит 108 признаков, объединенные из двух подписей в шахматном порядке. Для последующей классификации подписей используется алгоритм Random Forest. Для его обучения сформирован датасет из 4754 комбинаций подписей извлеченных из реальных документов, принадлежащих 188 различным людям. Из них 1416 подписей использовались в качестве тестового датасета, остальные 3324 были подготовлены для обучения. В качестве метода обучения использовался метод кросс валидации с разделением данных на 40 блоков с динамической маркировкой валидационного блока в процессе обучения. По результатам обучения проводился сбор метрик и сравнение их с результатами на тестовом датасете. Тестовый датасет формировался из подписей в соотношении 50 на 50 для получения наиболее репрезентативной выборки для оценки качества распознавания алгоритма.

### **Выводы.**

По результатам апробации валидация подписей проходит успешно, однако в рамках данного алгоритма возможно представить способы генерации изображений пользователей, которые могут привести к возникновению ошибок первого и второго рода. Одним из основных преимуществ алгоритма является быстроедействие, что может быть использовано для предварительной валидации подписи и дальнейшем принятии решения о запуске более точного, но медленного алгоритма.

Бушуев К.Р. (автор)

Подпись

Муромцев Д.И. (научный руководитель)

Подпись