

Лабораторная работа №7. Классификация на основе признаков, анализ профилей

Каждый студент выполняет все задания на выбранном алфавите и использует наработки двух предыдущих лабораторных работах.

1. Реализовать расчёт меры близости изображений символов на основе признаков. Использовать евклидово расстояние в n -мерном пространстве нормализованных признаков: масса, координаты центра тяжести, осевые моменты инерции. Нулевое расстояние между изображениями соответствует единичной мере близости.
2. Для каждого обнаруженного символа в строке рассчитать меру близости со всеми символами выбранного алфавита. Таким образом, для M символов распознаваемого изображения получится по N гипотез с оценками близости. Гипотезы следует отсортировать в порядке убывания меры близости.
3. Вывести результаты в файл, где в i -ой строке будут записаны гипотезы для i -ого символа распознаваемого текста. При этом гипотезы должны быть отсортированы по убыванию меры близости, так чтобы в первой колонке читались наиболее достоверные гипотезы. Например, в python в виде List[Tuple]:

```
1: [ ("а", 0.99), ("о", 0.87), ("с", 0.78), ... ]
2: [ ("п", 1.0), ("т", 0.77), ("р", 0.68), ... ]
...
```
4. Вывести лучшие гипотезы (из первого столбца) в виде строки и сравнить с распознаваемой строкой.
5. Вычислить количество ошибок и долю (%) верно распознанных символов.
6. Провести эксперимент: Сгенерировать изображение исходной строки с размером шрифта, отличающемся от исходного на несколько пунктов. Распознать её. Сравнить результаты.
7. Магистрам: Дополнить меру близости сравнением профилей с помощью метрики Левенштейна.