

В общем-то немного о том, что я сделал. Я написал некую базовую модель, с помощью которой можно анализировать эффективность работы написанного алгоритма. Для этого нужны видео и файл с субтитрами к нему. В качестве первого видео я решил взять некое не очень длинное видео с онлайн-лекции на coursera.org длительностью 5 минут и субтитры к нему.

Собственно сам алгоритм, по которому определялось, говорит человек или нет, использовалась разность вертикальных точек и горизонтальных точек, а именно, расширение по вертикали до определенного размера либо сужение по горизонтали до определенного размера.

Теперь немного о том, как анализировал эффективность алгоритма. Я начинал прогонять свой алгоритм и выдавать текущее соотношение эффективности отрезками примерно по 5 секунд. На каждом из таких отрезков я смотрел, когда человек говорит на самом деле (по субтитрам) и когда человек говорит согласно написанному алгоритму. Далее я брал объединение всех отрезков с 2 множеств и находил их суммарную длительность. После чего оценивал эффективность алгоритма, поделив общей длительностью на максимум из отрезков субтитров или речи из алгоритма.

Полный результат работы находится в приложении к данному отчету в файле *output.txt*.
Конечный результат получился следующим:

Subtitles : 260.061 seconds
Speech : 233.777 seconds
Union : 214.228 seconds
Match : 80.11 %

В качестве некого бонуса написал скрипт *run*, чтобы не делать постоянно сборку и чистку рабочей директории вручную. Скорее всего, он пока работает только в *MacOSX* (в нем есть стандартный *curl* для скачивания файлов, например), к тому же он больше настроен под меня, так как пока работает нормально только в *z — shell*, но это в общем-то поправимо. В принципе там более-менее понятно, как все это запускать (есть некое подобие *user-friendly* интерфейса). На этом вроде все.