

В. А. Рубин, С. П. Жогаль
(ГГУ им. Ф. Скорины, Гомель)

МУЛЬТИМОДАЛЬНЫЕ АНСАМБЛЕВЫЕ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ ЗАПОМИНАЕМОСТИ ВИДЕО

Моделирование запоминаемости медиа всегда было проблемой в области машинного обучения. В основном, для определения того,

Материалы XXIV Республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях», Гомель, 22–24 марта 2021 г.

насколько запоминаемым будет видео, анализируется несколько источников: статические изображения, входящие в состав видео, звук, и, непосредственно, отрезки видео с переменной длительностью, полученные из исходного видео. Также важным является разделение предсказания оценки запоминаемости в краткосрочной перспективе (через несколько минут после просмотра) и в долгосрочной (24–72 часа после просмотра).

В таком прогнозировании одним из наиболее эффективных подходов является использование функций на основе изображений, полученных из предобученных сверточных сетей (CNN). Если применить этот подход в комбинации с моделями семантического встраивания субтитров и моделями захвата изображений, можно получить достаточно точные результаты в предсказании показателей запоминаемости. Также, при исследовании запоминаемости видео, стоит обратить внимание на функции на основе видео, такие как C3D и I3D. И, в конце концов, наиболее производительные модели в предсказании запоминаемости видео можно получить, скомбинировав все упомянутые выше подходы. Обучающая выборка состоит из 590 коротких видеороликов, длительностью от 1 до 8 секунд с 2–5 аннотациями каждый. Каждое видео имеет соответствующие краткосрочные и долгосрочные оценки запоминаемости. Модели были обучены в соотношении обучения и проверки, как 80% к 20%. Поскольку конкатенация нескольких мультимодальных функций приводила к появлению очень многомерных векторов функций, которые были сложнообучаемы, мы обучили разные модели (Support Vector Regressor, Bayesian Ridge Regressor, Linear Models) для каждой отдельной функции независимо.

Наш основной вклад в демонстрации того, что модели на основе звука хорошо работают для прогнозирования краткосрочной запоминаемости и гораздо легче обобщают результат.