

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНО ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра математического анализа

Методы обнаружения дыма на видеоизображении

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 219 группы

направления 01.04.02 Прикладная математика и информатика

механико-математический факультет

Борисова Владимира Александровича

Научный руководитель

доцент, к.ф.-м.н. _____ Ю.В. Матвеева

Зав. кафедрой

д.ф.-м.н., профессор _____ Д.В. Прохоров

Саратов 2018

Введение. Большая часть систем обнаружения дыма основана на ультрафиолетовых камерах, многие другие работают по принципу анализа частиц, температуры, химического состава и прозрачности воздуха.

Минусы таких систем в том, что обнаружение происходит когда частицы задымленного воздуха уже в непосредственной близости от датчика. Также, такие системы не могут дать дополнительную информацию о расположении очага пожара, степени пожара. Поэтому системы, основанные на анализе видео, гораздо более эффективны при обнаружении дыма, особенно для больших и открытых зон. Также, такие системы значительно более просты в интеграции, за счет возможности использования уже установленных камер видеонаблюдения под систему пожарной безопасности. Стоит отметить, что установка новых камер видеонаблюдения проще, чем установка дымовых датчиков. Все это позволяет избежать многих временных и материальных издержек при организации системы пожарной безопасности.

Задача нахождения дыма на видео является важным и востребованным исследованием. На данный момент существует много работ, посвященных этой теме. Но из-за произвольной формы дыма, различных свойств дыма в зависимости от горящего материала, турбулентности движения огня и других факторов задача разработки системы обнаружения дыма по видео по-прежнему остается сложной и нерешенной задачей.

Базовую постановку задачи можно сформулировать следующим образом. Дан видеопоток. Под видеопотоком будем понимать конечную последовательность изображений. В общем случае, каждое изображение цветное и представляется в виде матрицы A размера $m \times n$. Каждый элемент матрицы — трехкомпонентный вектор из целых значений от 0 до 255.

$$a_{i,j} = (r, g, b \mid r, g, b \in [0, 255])$$

Результатом анализа видеопотока должен быть выделенный регион содержащий дым.

С точки зрения задач компьютерного зрения — это задача сегментации.

Целью данной работы являлось изучение существующих методов распознавания дыма, оценка и сравнение их эффективности и разработка компьютерной системы распознавания дыма на видео.

Работа разделена на главы по следующему принципу: В первом разделе «Данные» описаны формат данных, для которых будет решаться задача распознавания, необходимые преобразования над данными или предобработка перед работой Алгоритма, процесс и результат разметки данных, необходимый для оптимизации и оценки качества алгоритмов. Во втором разделе задача формулируется в терминах задач машинного обучения и формулируются необходимые в дальнейшем понятия. В разделе «Метрики оценки качества» даны определения метрик оценки качества алгоритмов, применяемых к построенным в данной работе алгоритмам а также описан способ выбора тестового множества данных для оценки качества. В разделе «Признаковое описание объектов. Выбор признаков» описан процесс представления данных в векторном пространстве. Данное представление называется — признаковым описанием, а каждый элемент этого вектора — признаком. В пятом разделе «Алгоритм на основе решающих деревьев» описан алгоритм построение решающего дерева, композиции решающих деревьев и реализована система распознавания дыма на их основе. Также произведено сравнение признаков, выбранных для построения деревьев. В последнем разделе «Алгоритм на основе сверточных нейронных сетей» был описан процесс построения конкурирующего алгоритма, в котором признаки были сгенерированы слоями нейросети.

Методы исследования. Методологическую основу работы составляют методы цифровой обработки сигналов, методы машинного обучения, численные методы приближения функций, аппарат дифференциального исчисления, матричная алгебра. Численные эксперименты произведены с помощью библиотек `opencv` для методов обработки изображений, `numpy` и `sklearn` для матричных операций и методов машинного обучения, `xgboost` для построения градиентного бустинга, `tensorflow` и `keras` для построения графа вычислений и нейросети.

Актуальность работы. В работах по теме обнаружения дыма было предложено множества способов описать дым, на основе которых были построены различные линейные классификаторы. Алгоритм на основе градиентного бу-

стинга используют большинство эффективных на данный момент признаков, и является композицией алгоритмов, способной давать более высокие результаты. Сравнительное сравнение признаков с помощью бустинга над решающими деревьями ранее не проводилась. Подход с применением сверточных нейронных сетей для распознавания дыма слабо изучен, не смотря на то, что в других задачах именно этот класс алгоритмов часто показывает наиболее высокие результаты.

Основная часть. Было принято понимать под видео конечную последовательность изображений. В общем случае, каждое изображение цветное и представляется в виде матрицы A размера $m \times n$. Каждый элемент матрицы — трехкомпонентный вектор из целых значений от 0 до 255.

$$a_{i,j} = (r, g, b \mid r, g, b \in [0, 255])$$

Результатом анализа видео должны быть координаты областей на кадрах видео, в которых присутствует дым. В первом разделе описаны этапы подготовки данных. В качестве предобработки размер каждого кадра приводится к 512 пикселям по ширине и происходит разбиение на блоки. Часть кадров размечаются вручную с помощью адаптированного инструмента `labelme`¹, адаптированного вручную для данной задачи. В ходе работы было вручную размечено 20736 блока.

Во втором разделе вводятся определения и формулировки для постановки задачи. Будем обозначать через x объект, для которого выполняется предсказание. В данном случае, объектами являются блоки изображения. Пространство всех объектов обозначим через X . Через $y = y(x)$ обозначим правильный ответ для объекта x . Все возможные ответы образуют пространство ответов Y . В задачах бинарной классификации это множество $Y = 0, 1$.

Определение 1. *Обучающей выборкой будем называть набор из l пар объект-ответ и будем обозначать его*

$$X = (x_i, y_i)_{i=1}^l$$

¹<https://github.com/wkentaro/labelme>

Определение 2. Алгоритмом или моделью будем называть функцию $a(x) : X \rightarrow Y$. Модель выбирается из параметрического семейства функций

$$A = \{a(x) = g(x, \theta) | \theta \in \Theta\} \quad (1)$$

где $g : X \times \Theta \rightarrow Y$, Θ — множество допустимых значений параметра θ .

Определение 3. Эмпирический риск — это функционал качества, характеризующий среднюю ошибку алгоритма a на обучающей выборке X^l :

$$Q(a, X^l) = \frac{1}{l} \sum_{i=1}^l L(a, x_i, y_i) \quad (2)$$

К каждому определению даны аналогии в контексте задачи. Задача сводится к задаче минимизации эмпирического риска.

$$a = \arg \min_{a \in A} Q(a, X^l) \quad (3)$$

В третьем разделе описаны метрики, позволяющие оценить качество найденного алгоритма a . Дано описание разбиения выборки на обучающую и тестовую с помощью кросс валидации.

Определение 4. Доля правильных ответов или ассюрасу вычисляется как

$$\text{accuracy}(a, X) = \frac{1}{l} \sum_{i=1}^l [a(x_i) = y_i] \quad (4)$$

Для определения других метрик введено понятие матрицы ошибок, имеющей вид:

	$y = 1$ (дым)	$y = 0$ (не дым)
$a(x) = 1$ (дым)	Верное срабатывание (TP)	Ложное срабатывание (FP)
$a(x) = 0$ (не дым)	Ложный пропуск (FN)	Верный пропуск (TN)

Определение 5. Метрика точность или precision определяется как

$$\text{precision}(a, X) = \frac{TP}{TP + FP} \quad (5)$$

Определение 6. Полнота или *recall* вычисляется как

$$recall(a, X) = \frac{TP}{TP + FN} \quad (6)$$

Определение 7. Обобщенная F - мера задается как

$$F_\beta = (1 + \beta^2) \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (7)$$

При $\beta < 1$ полнота имеет приоритет над точностью.

В четвертом разделе под названием «Признаковое описание объектов. Выбор признаков» определены функционалы в пространстве объектов, результатом которых являются некие уникальные числовые характеристики.

Определение 8. Каждому объекту x поставим в соответствие вектор $x = (f_1(x), f_2(x), f_3(x), \dots, f_d(x))$, где $f_i(x)$ — числовые признаки данного объекта, представляющие собой разные преобразования объекта x . Этот вектор будем называть признаковым описанием объекта x . Преобразование производится с помощью функционалов $f(x)$ вида:

$$f_i(x) = X \rightarrow D_j, \quad j = 1, \dots, n \quad (8)$$

где D_j — пространство признаков, в данной задаче $D_j = R$ — количественный признак.

В разделе описаны признаки, зарекомендовавшие себя в работах других авторов.

Признаки можно разделить на следующие группы:

1. Цветовые признаки
2. Текстурные признаки
3. Признаки формы и границ
4. Динамические признаки

Далее более детально рассмотрены признаки каждой из групп. Отдельно рассмотрены динамические признаки. На их основе построен алгоритм, основанный на одной из статей Chen-Yu Lee. По результатам была подтверждена

эффективность описанных динамических признаков. При экспериментально подобранном пороге для признаков дым был распознан на 6 из 8 видео.

В пятом разделе описан алгоритм на основе решающих деревьев и градиентного бустинга.

Определение 9. *Бинарное решающее дерево — алгоритм классификации $a(x)$, задающийся бинарным деревом: 1) $\forall v \in V \rightarrow$ предикат $\beta_v : X \rightarrow \{0, 1\}, \beta_v \in B$, 2) $\forall v \in V \rightarrow$ имя класса $c_v \in Y$, где B — множество бинарных признаков или предикатов (например, вида $\beta(x) = [x_j \leq \theta_j], x^j \in R$)*

Были описаны алгоритмы построения решающих деревьев ID3 и CART. Далее описан алгоритм построения ансамбля решающих деревьев — градиентный бустинг. С помощью этого алгоритма получены главные результаты этой работы — непосредственно сам алгоритм с результатами по описанным метрикам

precision	recall	f1-score
0.72	0.92	0.81

и график ранжированных по важности признаков.

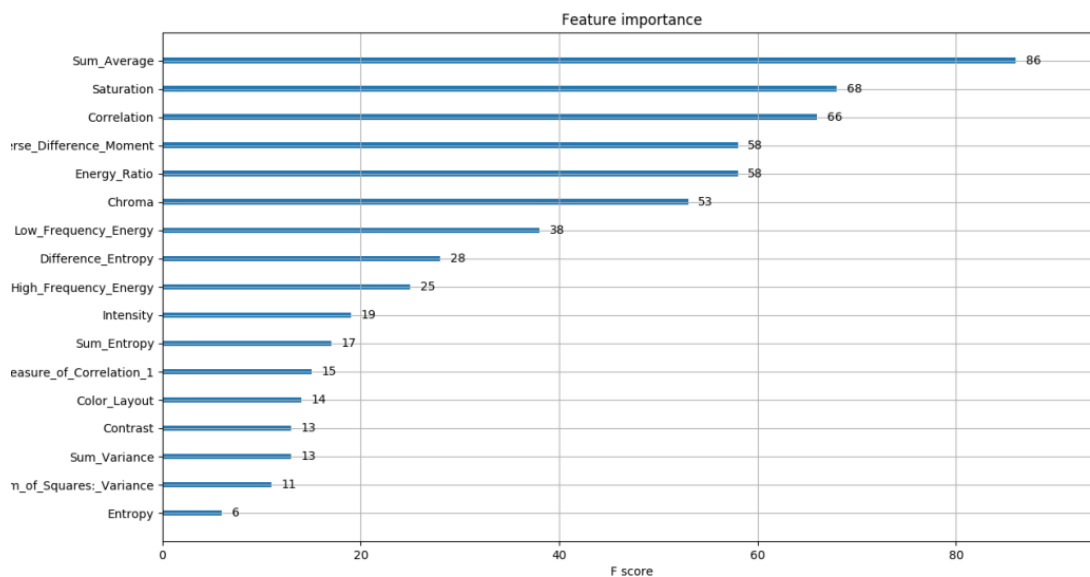


Рисунок 1 — Важность признаков

В разделе под номером 6 представлен алгоритм на основе сверточных нейронных сетей. Предварительно даны определения нейрона

Определение 10. Пусть $f_j : X \rightarrow R$, $j = 1, \dots, n$ — числовые признаки, $x_j = f_j(x)$. Функция вида

$$a(x, w) = \sigma(\langle w, x \rangle) = \sigma \left(\sum_{j=1}^n w_j f_j(x) \right) \quad (9)$$

где $w_0, w_1, \dots, w_n \in R$ — веса признаков, $\sigma(s)$ — произвольная функция активации называется линейной моделью нейрона МакКаллока-Питтса [?].

описан вид двухслойной нейронной сети

$$a_m(x, w) = \sigma_m \left(\sum_{h=0}^H w_{hm} \sigma_h \left(\sum_{j=0}^n w_{jh} f_j(x) \right) \right) \quad (10)$$

и ее аппроксимационные свойства. В качестве теоретического обоснования алгоритма представлены без доказательства следующие теоремы:

Теорема 11 (Колмогоров, 1957). Любая непрерывная функция n аргументов на единичном кубе $[0, 1]^n$ представима в виде суперпозиции непрерывных функций одного аргумента и операции сложения:

$$f(x^1, x^2, \dots, x^n) = \sum_{k=1}^{2n+1} h_k \left(\sum_{i=1}^n \varphi_{ik}(x^i) \right) \quad (11)$$

где h_k, φ_{ik} — непрерывные функции и φ_{ik} не зависят от f .

Теорема 12 (Горбань, 1998). Пусть X — компактное пространство, $C(X)$ — алгебра, непрерывных на X вещественных функций, F — линейное подпространство в $C(X)$, замкнутое относительно нелинейной непрерывной функции φ , содержащее константу ($1 \in F$) и разделяющее точки множества X . Тогда F плотно в $C(X)$.

Теорема 13 (Универсальная теорема аппроксимации (Hornik et al., 1989; Cybenko, 1989)). Сеть прямого распространения с линейным выходным сло-

ем и, по крайней мере, одним скрытым слоев с произвольной функцией активации может аппроксимировать любую измеримую по Борелю функцию, отображающую одно конечномерное пространство в другое с любую точностью, при условии, что в сети достаточно скрытых слоев.

Для изображений применяются сверточные нейронные сети. В работы описаны их уникальные отличия и слои архитектуры.

В рамках дипломной работы был реализован алгоритм с помощью сверточной нейронной сети. Использована архитектура resnet50, применялись библиотеки tensorflow и keras.

Эффективность этого алгоритма по значению f -меры равно 89.

Заключение. В данной работе были описаны и проанализированы различные признаки, уже использовавшиеся в других работах. На основе этих признаков был подстроен алгоритм градиентного бустинга, и продемонстрирована эффективность. Результаты по метрике f -меры составляет 81. Также этот алгоритм является легко интерпретируемым. В процессе построения алгоритма были выявлены признаки, оказавшие наибольшее влияние на результат, что в дальнейшем может быть использовано для оптимизации алгоритма, путем отброса части не сильно значимых признаков. В рамках работы был также построен алгоритм на основе динамических признаков и алгоритма вычитания фона MOG. Его точность по количеству распознанных видео 0.75. Третьим алгоритмом, построенным в данной работе — алгоритм на основе сверточных нейронных сетей дал наибольший результат в 89 по метрике f -меры. Данный алгоритм возможно частично интерпретировать с помощью применения других алгоритмов, но данное исследование не входит в рамки текущей работы.