

# Методы классификации с обучением по прецедентам в задаче распознавания объектов на изображениях

А.П. Вежнев

Лаборатория Компьютерной Графики и Мультимедиа факультета ВМиК,  
Московский государственный университет им. М.В.Ломоносова, Москва, Россия  
avezhnevets@graphics.cs.msu.ru

## Аннотация

Одной из важных частных задач машинного зрения является задача распознавания объектов на изображениях. Задача состоит в определении, является ли предъявленное изображение изображением интересующего нас объекта. К задаче распознавания сводятся задачи обнаружения и локализации объектов: обнаружение раковых клеток на снимках с микроскопа, обнаружение пешеходов или автомашин в системах видео наблюдения, и, уже ставшая классической, задача поиска человеческих лиц на изображениях. Последние исследования показали эффективность обучаемых методов для решения данной задачи. В статье рассматриваются различные методы классификации с обучением по прецедентам, применяющиеся к задаче распознавания объектов на изображениях. Проводится сравнение, и анализ причин тех или иных показателей работы рассмотренных методов.

**Ключевые слова:** распознавание объектов на изображениях, классификация, обучение по прецедентам, машина опорных векторов, усиление слабых классификаторов, *SNoW*.

## 1. ВВЕДЕНИЕ

Одной из важных частных задач машинного зрения является задача распознавания объектов на изображениях. Задача состоит в определении, является ли предъявленное изображение изображением интересующего нас объекта. К задаче распознавания сводятся задачи обнаружения и локализации объектов: обнаружение раковых клеток на снимках с микроскопа, обнаружение пешеходов или автомашин в системах видео наблюдения, и уже ставшая классической, задача поиска человеческих лиц на изображениях. Распознавание объектов является, в общем случае, некорректно поставленной задачей (легко представить два объекта разной природы, одинаково выглядящих на изображении). Несмотря на то, что человек зачастую с лёгкостью справляется с этой задачей, для вычислительной системы такая задача представляет сложность.

Можно выделить три основных подхода [1] ниже к ее решению:

### 1. Эвристические методы [2][3]:

- a. **Полная эвристическая модель.** Экспертом составляется набор правил, описывающих изображение объекта (строится модель), согласно которым производится обнаружение;
- b. **Поиск характерных инвариантных признаков.** Эвристически описываются не изображение искомого

объекта в целом, а его характерные признаки, инвариантные относительно возможных искажений (изменение освещения, поворот, масштабирование);

2. **Метод сравнения с шаблоном.** Составляется шаблон для изображения всего объекта или его характерных признаков. Также вводится функция проверки соответствия;

3. **Методы с обучением по прецедентам.** Модель автоматически строится на основе набора изображений объекта, составленных заранее из возможных входных данных системы [6][12][19][26].

### 1.1 Эвристические методы

Данные методы являются исторически самыми ранними, именно они применялись в самых первых системах машинного зрения. Системы, основанные на эвристике, были достаточно просты в реализации и работали с высокой скоростью, однако жёстко запрограммированные правила лишали систему гибкости и устойчивости. Как правило, эвристические системы ориентированы на относительно узкий класс задач. Например, какое-то характерное освещение, ракурс, ограниченный класс изображений не являющихся объектом и т.д. На этом классе задач эвристические системы могут достигать высокого уровня точности и низкой частоты сбоев, но при этом очень плохо поддаются обобщению и практически не применимы для объектов иной природы, чем изначальные. Инвариантность к пространственным преобразованиям и освещению зависит от качества подобранной эвристики. Данный подход допускает реализацию в реальном времени [30][31][32].

### 1.2 Метод сравнения с шаблоном

Это более универсальный подход, в смысле обобщения на другие виды объекта, однако требует наличие весьма точного шаблона изображения объекта. Шаблон может быть сложной структурой и допускать различные деформации и преобразования, таким образом, способствуя инвариантности системы к пространственным искажениям изображения объекта и изменениям освещённости. Системы, основанные на сравнении с шаблоном, зачастую используются для решения задач отслеживания объектов в видео с инициализацией на первом кадре – до начала работы системы существует общая модель шаблона, при инициализации она уточняется и корректируется во время работы системы.

При хорошо заданном шаблоне достигается высокая точность и очень низкий уровень сбоев. Инвариантность к пространственным искажениям и изменению освещения зависит от сложности шаблона. Данный подход допускает реализацию в реальном времени.

### 1.3 Методы с обучением по прецедентам

Это наиболее общий подход. Задача распознавания объектов на изображение сводится к задаче классификации и для неё применяется хорошо разработанный математический аппарат построения модели (обучения) по прецедентам. Модель строится автоматически по заранее собранному набору прецедентов - изображений, для которых известно, являются ли они изображениями объекта или нет. Наблюдением, в данном случае, является некоторый «вектор признаков», полученный из исходного изображения некоторым преобразованием, отображающим изображения в пространство действительных векторов. Гипотеза, подлежащая проверке - принадлежность изображения к классу изображений искомого объекта.

Таким образом, система распадается на два модуля: модуль преобразования изображения в вектор признаков и модуль классификации. Задачей модуля преобразования является наиболее полное и информативное представление изображения в виде числового вектора. Задачей модуля классификации является проверка гипотезы принадлежности изображения классу изображений объекта на основании наблюдения, которым является вектор признаков.

Модуль преобразования и модуль классификации тесно связаны. Главная цель модуля преобразования – представить изображение в форме наиболее удобной для модуля классификации. Основные требования, предъявляемые к модулю преобразования: скорость, наиболее полное и информативное представление данных, масштабируемость (преобразование корректно работает с изображением разных размеров, размерность вектора признаков не меняется). Также является желательной инвариантность модуля преобразования к деформациям и пространственным искажениям объекта. Несмотря на важность метода преобразования, его подробное рассмотрение выходит за рамки этой работы.

## 2. МЕТОДЫ КЛАССИФИКАЦИИ С ОБУЧЕНИЕМ ПО ПРЕЦЕДЕНТАМ

Целью обзора является проанализировать существующие методы классификации с обучением по прецедентам, применимые к задаче распознавания объектов на изображениях. Данные методы выбраны к рассмотрению, ввиду их общности и многообещающих результатов в реальных системах.

Вначале будет приведена математическая постановка задачи, определены базовые понятия и введена нотация. Далее будут сделаны некоторые общие замечания, соотносящие математическую постановку задачи с практической реализацией системы. Формулировка критериев сравнения различных методов будет приведена в секции 2.3. Затем последует основная часть обзора.

### 2.1 Формальная постановка задачи, определения и нотация

Опишем формальную постановку задачи для модуля классификации и введём нотацию.

**Вектор признаков**  $x \in X \subseteq R^n$  - является описанием объекта, которое поставляется модулем преобразования.

**Классом** будем называть некоторое подмножество  $K_y = \{x \in X \mid y^*(x) = y\}$  множества  $X$ . Пусть  $y \in Y \subset Z$  - множество маркеров классов, будем рассматривать случай бинарной классификации, то есть  $Y = \{-1, +1\}$ . Применительно к задаче распознавания объектов подразумевается два варианта – изображение есть изображение искомого объекта или изображение есть изображение чего-либо другого.  $X \xrightarrow{y^*} Y$  - отображение, определённое для всех  $x \in X$ , задающее разбиение  $X$  на подмножества  $K_y$ . Множества  $K_y$ , вообще говоря, могут пересекаться. Причина возможности пересечения классов, заключается в том, что вектор признаков есть всего лишь описание объекта, а не сам объект – описание может быть неточным и для двух разных объектов описания могут совпадать.

**Обучающей выборкой** называется набор пар (набор прецедентов)  $T = (x_1, y_1), \dots, (x_l, y_l)$ , для которых  $y^*(x_i) = y_i, i = \overline{1, l}$ , то есть это известная информация об отображении  $X \xrightarrow{y^*} Y$ .

Для применения алгоритмов классификации и распознавания образов принимается следующая гипотеза:

**Гипотеза:** Множество  $X \times Y$  является вероятностным пространством с вероятностной мерой  $P$ . Прецеденты  $(x_1, y_1), \dots, (x_l, y_l)$  появляются случайно и независимо в соответствии с распределением  $P$ .

**Задача классификации** – построить функцию  $F(x)$ , классификатор, приближающую отображение  $y^*$ , основываясь на обучающей выборке  $(x_1, y_1), \dots, (x_l, y_l)$ .

Введём несколько дополнительных определений:

**Эмпирическим риском** будем называть  $P(F(x) \neq y \mid (x, y) \in T)$ , то есть вероятность неверной работы классификатора для векторов признаков из обучающей выборки.

**Общим риском** будем называть  $P(F(x) \neq y \mid x \in X)$ , то есть вероятность того, что классификатор ошибётся на данных, не входивших в обучающую выборку. Понятно, что основной целью при построении классификатора является минимизация именно общего риска. Так как на прямую вычислить величину общего риска невозможно, для проверки качества классификатора используется оценка ошибки на **контрольной выборке**, которая состоит из прецедентов, не входящих в обучающую выборку. Говорят, что классификатор обладает хорошей **обобщающей способностью**, если при обучении классификатор эффективно уменьшает общий риск, оценённый на контрольной выборке.

**Обучение или тренировка** – процесс построения классификатора по тренировочной выборке  $T$ .

Говорят, что алгоритм построения склонен к **переобучению**, если при минимизации эмпирического риска, общий риск начинает возрастать. Этот феномен связан с тем, что классификатор начинает обобщать признаки свойственные не данным в целом, конкретно прецедентам из обучающей выборки. Переобучению способствует невыполнение принятой гипотезы, шум, высокая сложность классифицирующей функции (высокая размерность Вапника-Червоненкиса [16]).

## 2.2 Общие замечания

Для дальнейшего изложения, требуется сделать несколько замечаний, связывающих практическую задачу распознавания объектов на изображениях и теоретическую постановку задачи обучения по прецедентам.

Обучающая выборка, основа будущего классификатора, составляется разработчиками системы. В неё должны войти примеры изображений, с которыми будет работать система. Важно, чтобы выборка максимально отвечала гипотезе о случайности и независимости прецедентов.

Обучение в реальных системах может занимать несколько десятков, а иногда и сотен, часов. Таким образом, в случае склонности классификатора к переобучению, тонкая настройка параметров обучения может сильно замедлить процесс разработки системы.

За составление вектора признаков отвечает модуль преобразования, подробное рассмотрение которого выходит за рамки работы, однако требуется отметить некоторые свойства, присущие всем преобразованиям, используемым на практике:

- Важное свойство вектора признаков – его размерность зачастую много больше, чем размер обучающий выборки  $l \ll n$ , то есть задача плохо обусловлена;
- Входное изображение может быть зашумлено и, как следствие, будут зашумлены векторы признаков, однако, характер шума в векторах признаков будет неизвестен, т.к. шум тоже пройдет через модуль преобразования;
- Несмотря на то, что вектор признаков составляется таким образом, чтобы максимально полно описать объект, могут возникать случаи, когда для различных объектов векторы признаков будут очень близки или вообще идентичны.

## 2.3 Критерии сравнения алгоритмов классификации

Для дальнейшего изложения требуется ввести четкие критерии сравнения методов классификации. Сравнение предлагается проводить по следующим критериям:

1. Ошибки первого и второго рода на контрольных выборках в задачах распознавания объектов на изображениях;
2. Устойчивость к шуму;
3. Склонность к переобучению;
4. Вычислительная сложность и скорость реальных систем.

Численные данные, которые будут приводиться ниже, взяты из публикаций авторов соответствующих систем и независимых обзоров. Все системы ориентированы конкретно на задачу распознавания лиц, однако, допускают

применение к задачам распознавания любых других объектов.

## 2.4 Байесовские методы классификации

**Принцип максимума апостериорной вероятности** [5] основывается на трёх гипотезах:

1. Множество  $X \times Y$  является вероятностным пространством с вероятностной мерой  $P$ . Прецеденты  $(x_1, y_1), \dots, (x_l, y_l)$  появляются случайно и независимо в соответствии с распределением  $P$ ;
2. Известны плотности распределения классов  $p_y(x) = p(x | K_y), y \in Y$ , называемые функциями правдоподобия;
3. Известны вероятности появления объектов каждого из классов  $P_y = P(K_y), y \in Y$ , называемые априорными вероятностями.

При принятии данных гипотез решающее правило записывается в аналитическом виде:

$$F(x) = \arg \max_{y \in Y} P(K_y | x) = \arg \max_{y \in Y} p_y(x) P_y$$

**Формула 2.1.** Принцип максимума апостериорной вероятности.

Доказано [5], что такой выбор решающего правила является оптимальным с точки зрения минимизации общего риска. Разделяющая поверхность в данном случае не имеет какой-либо четкой геометрической структуры.

Основная проблема, заключается в том, что на практике гипотезы 2 и 3 почти никогда не выполняются. Попытки оценить эти функции распределения по обучающей выборке могли бы привести к некоторому результату, если бы не плохая обусловленность задачи, которая приводит к вырожденным решениям.

Существуют системы обнаружения объектов на изображениях основанные на «наивном» Байесовском методе [33]. Данный метод основывается на построении эмпирической плотности распределения вероятностей классов по обучающей выборке в предположении о независимости компонентов вектора признаков.

Проблема слабой обусловленности решается за счет выбора компактного представления вектора признаков. Данный подход является аппроксимацией метода максимума апостериорной вероятности и поэтому не обладает оптимальностью с точки зрения общего риска. Метод очень сильно зависит от выбора модуля преобразования, обычно признаками являются достаточно сложные характеристики изображения объекта интереса, подбираемые эвристически.

Практические результаты следующие:

1. В задаче распознавания лиц примерно 91% верных обнаружений и порядка 0.2% ложных срабатываний [1];

2. Алгоритм построения «наивного» Байесовского классификатора склонен к переобучению;
3. Алгоритм построения «наивного» Байесовского классификатора чувствителен к шуму, т.к. основывается на эмпирических функциях плотности распределения;
4. Скорость работы самого классификатора крайне высока, основное время может занимать вычисление вектора признаков.

**Байесовские сети (сети доверия)** [7] – это подход к классификации, основанный на совмещении Байесовского подхода и теории графов. Строится граф, каждая вершина которого соответствует какой-либо компоненте вектора признаков, дуги обозначают причинно-следственную связь. Построение сети может быть осуществлено автоматически за счет анализа корреляции компонент вектора признаков.

Такой подход не требует столь сильных предположений, как принцип максимума апостериорной вероятности, однако и не обладает теоретической привлекательностью, то есть при отсутствии априорных данных построенная сеть не будет доставлять минимум общему риску.

Плохая обусловленность, тем не менее, является проблемой и для байесовской сети, т.к. большая размерность вектора признаков делает граф связей очень сложным для построения и анализа. Также сильно возрастает вычислительная сложность. Одним из вариантов решения данной проблемы является сокращение размерности вектора признаков, что приводит к ухудшению обобщающей способности.

Применительно к задаче распознавания объектов на изображениях байесовские сети дают следующие результаты [6]:

1. 90% правильных обнаружений при 0.4% ошибок второго рода для задачи распознавания лиц [1];
2. Утверждается, что байесовские сети достаточно чувствительны к шуму[8], однако более глубоких исследований этого вопроса не проводилось;
3. Утверждается, что метод устойчив к переобучению[7];
4. Вычислительная сложность линейно зависит от размерности вектора признаков. Данные по скорости реальных систем не опубликованы.

## 2.5 Классические нейронные сети

Под термином нейронные сети скрывается целый класс алгоритмов [9]. Основная идея лежащая в основе нейронных сетей - это последовательное преобразование сигнала, параллельно работающими элементарными функциональными элементами, нейронами. Основной принцип настройки нейронной сети заключается в применении оптимизационных методов к минимизации среднеквадратичной ошибки, как следствие – склонность к переобучению. Главное преимущество нейронных сетей – гибкость. Возможности эвристических модификаций нейросетей почти безграничны[10]. Нейросети относятся к классическим методам машинного зрения, хотя в последнее время стали терять популярность.

Геометрически, разделяющая классы поверхность представляет собой множество гиперплоскостей. Каждая из

областей, на которые гиперплоскости разбивают пространство признаков  $X$  относится к одному из классов. Существует множество методов обучения нейросетей, однако все они сводятся к минимизации среднеквадратичной ошибки. Важно отметить, что найденный минимум, будет локальным. Также следует отметить, что верно классифицированные прецеденты не вносят никаких изменений в оптимизируемый функционал. Таким образом, найденная разделяющая поверхность не будет являться ни единственным, ни оптимальным решением.

Системы распознавания объектов на изображениях [11][12] основанные на нейронных сетях используют иерархическую архитектуру. Вначале вектор признаков обрабатывается грубой сетью с высоким уровнем ошибок второго рода, далее, если вектор не был классифицирован как не объект, решение корректируется более точной и более медленной сетью. Такой подход даёт вполне приемлемые результаты:

1. 92% правильных обнаружений при 1.3% ошибок второго рода [12] для задачи распознавания лиц;
2. Утверждается [13], что нейронные сети в целом склонны к переобучению, хотя и существуют некоторые методы, которые в частном случае могут решить эту проблему;
3. Устойчивость к шуму сильно зависит от конкретной архитектуры сети. В общем случае, нейросеть чувствительна к шуму;
4. Вычислительная сложность квадратично зависит от числа нейронов в скрытом слое. Каждый нейрон требует вычисления функции активации. Для задач распознавания объектов на изображениях скорость в реальных системах недостаточна для применения систем в реальном времени.

## 2.6 Метод опорных векторов (SVM)

Оригинальный метод опорных векторов был предложен В. Вапником в 1963 году [14], как метод для построения оптимального линейного классификатора. Хотя, предположение о линейной разделимости классов менее сильное, чем, например, предположения принципа максимальной апостериорной вероятности, на практике оно редко выполняется. В 1992 был предложен способ обобщения метода опорных векторов на достаточно широкий класс нелинейных задач [15].

Классический алгоритм заключается в построении линейной разделяющей поверхности (гиперплоскости), равноудалённой от выпуклых оболочек классов, выпуклая оболочка строится по прецедентам. Утверждается, что такая разделяющая гиперплоскость будет оптимальна, с точки зрения общего риска, относительно любых других возможных гиперплоскостей. Если такая гиперплоскость не существует (классы линейно не разделимы) то для осуществления нелинейной классификации применяется ядровое преобразование, проецирующие исходное пространство  $X$  в пространство ещё большей, возможно бесконечной, размерности. Линейная разделяющая поверхность в индуцированном ядровым преобразованием пространстве является нелинейной в исходном. Таким образом, частично решается проблема нелинейной классификации.

Сам алгоритм построения классификатора сводится к задаче квадратичного программирования и решается известными

оптимизационными методами [16]. Следует отметить, что решение задачи квадратичного программирования будет единственным и найденный экстремум глобальным.

Метод опорных векторов был успешно применён для задачи распознавания объектов на изображениях [17]. Подход даёт следующие результаты:

1. При порядка 72% верных обнаружений порядка 0.6% ложных срабатываний, для задачи распознавания лиц [1].
2. Очень высокая устойчивость к переобучению
3. Чувствительность к шуму может регулироваться, за счет уменьшения точности. Обычно стараются найти компромисс между устойчивостью и точностью.
4. Вычислительная сложность линейна от сложности вычисления скалярного произведения векторов признаков в индуцированном ядерном преобразованием пространстве. В системах распознавания объектов на изображениях метод даёт ускорение в несколько раз, по сравнению с нейронными сетями.

## 2.7 SNoW – разреженная просеивающая сеть

SNoW (Sparse network of Winnows) – особый вид нейронной сети [18]. Вектор признаков полагается бинарным. Сеть состоит из двух (по числу возможных классов) линейных нейронов, связанных с компонентами вектора признаков. Классификация проходит по принципу победитель забирает всё [10].

Геометрически, SNoW представляет собой две гиперплоскости в пространстве векторов признаков. Вектор относится к тому классу, соответствующая гиперплоскость которого ближе всего. Таким образом, результирующая разделяющая поверхность является гиперплоскостью в исходном пространстве  $X$ . Одно из достоинств данной архитектуры – возможность «прореживать» вектор признаков, на основе обучающей выборки – компоненты вектора, не несущие информации, отбрасываются.

SNoW считается достаточно эффективным методом для решения задачи обнаружения объектов на изображениях [19]. Согласно некоторым исследованиям, [20] SNoW превосходит по своим параметрам метод опорных векторов. На практике метод обладает следующими свойствами:

1. Порядка 94% верных обнаружений, при порядка 0.12% ложных срабатываний [1], для задачи распознавания лиц;
2. Склонность к переобучению не исследована. Следует отметить, что за счет простой классифицирующей функции (две гиперплоскости), согласно теории Вапника можно ожидать хорошей устойчивости к переобучению;
3. Устойчивость к шуму не исследована, но теоретически должна быть достаточно высокой;
4. За счёт просеивания компонент вектора признаков достигается высокая скорость – сложность линейна относительно количества эффективных компонент вектора признаков. Скорость реальной системы выше, чем у систем на основе метода опорных векторов.

## 2.8 Метод усиления слабых классификаторов (classifier boosting)

Усиление слабых классификаторов - это подход к решению задачи классификации, путём комбинирования примитивных классификаторов в один более сильный комитет. Основная идея метода заключается в итеративной минимизации выпуклого функционала ошибки классификации, путем добавления в комитет очередного слабого классификатора.

Фактически, готовый комитет осуществляет линейную классификацию в пространстве выходных значений слабых классификаторов. Можно сказать, что boosting пытается спроецировать исходные данные в пространство особого вида, где классы будут линейно разделены. В методе опорных векторов для этой цели использовалось ядровое преобразование, здесь же используются слабые классификаторы.

Существует множество разновидностей базового алгоритма [25][27][28]. Важными вехами в развитии метода были обобщение на случай вещественнозначного слабого классификатора и получение верхней оценки общего риска [28], доказательство связи подхода с логистической регрессией [25] и регуляризованной многомерной оптимизацией [29].

Отметим, что boosting использует жадный алгоритм минимизации функционала ошибки, таким образом, найденный минимум не будет глобальным, а решение единственным.

Для прикладных систем распознавания объектов на изображениях подход был успешно применён в 2001 году Viola, Jones [26]. Ими был построен каскад и комитетов слабых классификаторов, работавший по принципу последовательных приближений.

Каскад состоит из  $K$  ступеней, каждая ступень – комитет простых классификаторов построенный алгоритмом AdaBoost (в более поздних вариантах использовалась его модификация). Если ступень принимает решение о том, что вектор признаков относится к классу объекта принимается, только если все ступени каскада это подтвердили, иначе вектор признаков классифицируется, как не искомый объект.

Каскадный подход был специально создан для задачи распознавания объектов на изображениях. Его показатели [26] превосходят все остальные системы:

1. 95% верных обнаружений, порядка  $1 \cdot 10^{-3}$ % ложных обнаружений для комитета, построенного методом AdaBoost. 94% верных обнаружений, порядка  $1 \cdot 10^{-4}$ % ложных обнаружений для каскада [26], для задачи распознавания лиц;
2. Метод AdaBoost в некоторых случаях склонен к переобучению [25], для каскада исследований не проводилось, но нет причин полагать, что склонность к переобучению ступеней не приведёт к склонности к переобучению каскада в целом;
3. Метод AdaBoost чувствителен к шуму, для каскада исследований не проводилось, но нет причин полагать, что чувствительность к шуму ступеней не приведёт к неустойчивости каскада в целом;

4. Вычислительная сложность линейна по количеству итераций обучения (не зависит от размерности вектора признаков). Самая быстрая существующая система - скорость в реальном времени.

### 3. ОБСУЖДЕНИЕ

По показателям работы в реальных системах распознавания объектов на изображениях наиболее удачными оказались алгоритмы boosting (усиления слабых классификаторов) и SNoW. Оба подхода обеспечивают высокую скорость, высокий уровень обнаружений и низкий уровень ошибок второго рода.

Таблица 1. – сводные данные систем

Методы	Процент верных обнаружений	Ошибка второго рода
«Наивный» Байес	~91%	~0.2%
Байесовская сеть	~90%	~0.4%
Нейронные сети	~92%	~1.3%
Метод опорных векторов	~72%	~0.6%
SNoW	~94%	~0.12%
Усиление слабых классификаторов	~94%	~0.0001%

Достаточно интересные результаты у системы, основанной на «наивном» Байесовском обучении. Единственным серьёзным недостатком системы является сильная зависимость от качества модуля преобразования. Фактически подход требует пересмотра представления изображения объекта интереса в векторе признаков для каждой конкретной задачи.

Нейронные сети и метод опорных векторов достаточно сильно уступают по показателям вышеперечисленным подходам. Главным недостатком нейросетей является низкая скорость и большое число ошибок второго рода. Метод опорных векторов выгодно отличается скоростью и низким уровнем ошибок второго рода, но имеет очень низкий процент верных обнаружений.

Для того чтобы проанализировать получившиеся результаты сформулируем методы, основанные на нейронных сетях, метод опорных векторов, SNoW и boosting в некотором общем виде. Мы опустили Байесовские методы, так как их принцип работы довольно сильно отличается от остальных, но скажем о них чуть ниже. Итак, все четыре метода основываются на оптимизации некоторого функционала ошибки  $F(T, w)$ , который, в свою очередь, зависит от обучающей выборки и  $T$  параметров классификатора  $W$ . Так как обучающая выборка фиксирована, параметр  $T$  можно опустить и, фактически, задача сводится к оптимизации  $F(w)$  по  $w$ . Таким образом, решением, готовым классификатором, будет являться  $\hat{w} = \arg \min_w F(T, w)$ . Для нейросети  $W$  является набором её весов, для SNoW и метода опорных векторов

$W$  представляет направляющий вектор разделяющей гиперплоскости. Для boosting  $W$  является вектором весов для всех возможных простых классификаторов, то есть, если какой-либо простой классификатор не вошёл в итоговый комитет, его вес полагается равным нулю. Важно заметить, что для boosting и некоторых видов метода опорных векторов (зависит от используемого ядерного преобразования)  $w \in R^\infty$ .

Секрет успеха методов, показавших наилучшие результаты, может заключаться в одном их общем свойстве. Дело в том, что SNoW и boosting производят достаточно «разреженные решения», то есть  $W$  содержит множество нулей. В то же время, нейронные сети и метод опорных векторов получают достаточно «плотные» решения – большая часть компонентов  $W$  не равна нулю. Математически это означает, что наилучшее обобщение дают классификаторы с наименьшей  $l_1$  нормой. Это может означать, что данные в задачах распознавания имеют особую структуру, наилучшим образом представимую «разреженной» моделью. На практике это означает, что, несмотря на высокую размерность вектора признаков, количество реально значимых компонент в нём не велико. В предположение о наличие шума «разреженные» методы так же могут иметь преимущество, так как наиболее зашумлённые компоненты вектора признаков будут иметь нулевой вес и не влиять на решение. Высокое количество фиктивных переменных является следствием общности подходов. Для конкретной задачи размерность может быть сокращена, что и делается неявно «разреженными» методами. Похожие выводы были сделаны при сравнении метода опорных векторов и SNoW [20].

Метод, основанный на «наивном» Байесовском обучении, даёт также очень хорошие результаты. На самом деле, метод представляет собой компромисс между эвристикой и обучением по прецедентам. Аккуратно сконструированный под конкретную задачу модуль преобразования вместе с простым, но эффективным обучаемым методом. Такой подход может широко использоваться при создании специфических систем.

Метод, основанный на Байесовских сетях доверия, не показал высоких результатов. Возможно, его структура плохо подходит к задаче распознавания объектов на изображениях. На данный момент этот подход не кажется перспективным.

### 4. ЗАКЛЮЧЕНИЕ

В статье были рассмотрены основные методы классификации с обучением по прецедентам, применимые к задаче распознавания объектов на изображениях. На основе результатов сравнения были сделаны выводы о перспективности подходов, дающих «разреженные» решения.

Так же отмечается хорошая эффективность метода основанного на «наивном» Байесовском обучении и аккуратном выборе представления векторов признаков для конкретной задачи.

### 5. БИБЛИОГРАФИЯ

[1]Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on*

- Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 1, pp. 34-58, 2002
- [2]G. Yang and T. S. Huang, "Human Face Detection in ComplexBackground," *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.
- [3]C. Kotropoulos and I. Pitas, "Rule-Based Face Detection in Frontal Views," *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, vol. 4, pp. 2537-2540, 1997.
- [4]Воронцов К. В. "Обзор современных исследований по проблеме качества обучения алгоритмов" *Таврический вестник информатики и математики.* — 2004. [www.ccas.ru/frc/papers/voron04twim.pdf](http://www.ccas.ru/frc/papers/voron04twim.pdf)
- [5]Воронцов К. В. "Байесовские алгоритмы классификации. Черновики лекций" <http://www.ccas.ru/voron/download/Bayes.pdf>
- [6]T. V. Pham, M. Worring, and A. W. M. Smeulders. "Face detection by aggregated bayesian network classifiers" *Pattern Recognition Letters*, 23(4):451-- 461, February 2002.
- [7]Heckerman, D. (1998) "A Tutorial on Learning with Bayesian Networks," In Jordan, M. (Ed.), *Learning in Graphical Models*, MIT Press.
- [8]Adam Zagorecki and Marek J. Druzdel. "An empirical study of probability elicitation under Noisy-OR assumption" In *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS-2004)*
- [9]Александр Вежневцев "Популярные нейросетевые архитектуры" электронный журнал "Графика и Мультимедиа" [http://cgm.graphicon.ru/metodyi/populyarnyie\\_neyrosetevyie\\_arhitekturyi.html](http://cgm.graphicon.ru/metodyi/populyarnyie_neyrosetevyie_arhitekturyi.html)
- [10]Александр Вежневцев "Нестандартные нейросетевые архитектуры" электронный журнал "Графика и Мультимедиа" [http://cgm.graphicon.ru/metodyi/nestandartnyie\\_neyrosetevyie\\_arhitekturyi.html](http://cgm.graphicon.ru/metodyi/nestandartnyie_neyrosetevyie_arhitekturyi.html)
- [11]H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, pp. 23--38, 1998.
- [12]Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Human face detection in visual scenes", *Technical Report CMU-CS-95-158R, School of Computer Science, Carnegie Mellon University, November 1995.*
- [13]S. Lawrence and C. Giles. *Overfitting and neural networks: Conjugate gradient and backpropagation.* In *Proceedings of the IEEE International Conference on Neural Networks*, pages 114--119. *IEEE Press*, 2000.
- [14]Vapnik, V. and Lerner, A. J., "Generalized portrait method for pattern recognition," *Automation and Remote Control*, vol. 24, no. 6, 1963.
- [15]Boser, B.; Guyon, I.; Vapnik, V. 1992. "A training algorithm for optimal margin classifiers" *Fifth Annual Workshop on Computational Learning Theory.* *ACM Press, Pittsburgh.*
- [16]C. J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition" *Knowledge Discovery and Data Mining*, 2(2), 1998.
- [17]E. Osuna, R. Freund & F. Girosi. "Training support vector machines:an application to face detection" In *Proceedings of Computer Vision and pattern Recognition'97*, pages 130--136, 1997.
- [18]D. Roth, «The SNoW Learning Architecture" *Technical Report UIUCDCS-R-99-2102, UIUC Computer Science Department, 1999.*
- [19]D. Roth, M.-H. Yang, and N. Ahuja, "A SNoW-based face detector" in *Advances in Neural Information Processing Systems 12 (NIPS 12)*, MIT Press, Cambridge, MA, pp. 855-861, 2000.
- [20]M. Yang, D. Roth, and N. Ahuja, "A Tale of Two Classifiers: SNoW vs. SVM in Visual Recognition" *Proc. of the European Conference on Computer Vision (ECCV) (2002)* pp. 685--700
- [21]L. Breiman, J. H. Friedman, R. A. Olshen & C.J. Stone. «Classification and Regression Trees" *Rapport technique, Wadsworth International, Monterey, CA, 1984.*
- [22]J.R. Quinlan, «C4.5 Programs for Machine Learning», *Morgan Kaufmann, San Mateo, CA, 1993.*
- [23]R. E. Schapire. "The Boosting Approach to Machine Learning. An Overview" *MSRI Workshop on Nonlinear Estimation and Classification, 2002*
- [24]Y Freund and R. E. Schapire. "Game theory, on-line prediction and boosting" In *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, pages 325--332, 1996.
- [25]Jerome Friedman, Trevor Hastie, and Robert Tibshirani. "Additive logistic regression: A statistical view of boosting" *The Annals of Statistics*, 38(2):337--374, April 2000.
- [26]P. Viola and M. Jones. "Robust Real-time Object Detection" In *Proc. 2nd Int'l Workshop on Statistical and Computational Theories of Vision -- Modeling, Learning, Computing and Sampling, Vancouver, Canada, July 2001.*
- [27]Alexander Vezhnevets, Vladimir Vezhnevets "Modest AdaBoost' - Teaching AdaBoost to Generalize Better". *Graphicon-2005, Novosibirsk Akademgorodok, Russia, 2005.*
- [28]R.E. Schapire and Y. Singer. "Improved boosting algorithms using confidence-rated predictions." *Machine Learning*, 37(3):297-336, December 1999.
- [29]Rosset, Zhu and Hastie. "Boosting as a Regularized Path to a Maximum Margin Classifier" *Journal of Machine Learning Research* 5 (2004) 941--973, 2004.
- [30]G. Yang and T. S. Huang, "Human Face Detection in ComplexBackground," *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.
- [31]T. Kanade, "Picture Processing by Computer Complex and Recognition of Human Faces," *PhD thesis, Kyoto Univ., 1973.*
- [32]H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating Faces and Facial Parts," *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 41-46, 1995.
- [33] Charles Elkan "Naïve Bayesian learnin" *Adapted from Technical Report No. CS97-557, Department of Computer Science and Engineering, University of California, San Diego, September 1997.*

## 6. ОБ АВТОРАХ

Вежнев Александр Петрович студент факультета вычислительной математики и кибернетики Московского Государственного Университета им. М.В. Ломоносова. К основным научным интересам относятся машинное зрение, обработка изображений, классификация, математическая статистика.

Адрес: Москва, 119899, Воробьевы горы, МГУ, 2-й учебный корпус, факультет ВМиК, лаборатория Машинной Графики и Мультимедиа E-mail: [vezhnick@gmail.com](mailto:vezhnick@gmail.com)

### Machine Learning Techniques in Task of Visual Recognition

#### Abstract

Task of visual recognition is an important subproblem of computer vision. Visual recognition is widely used in practical applications in medicine, human-computer interface etc. Recent developments show the effectiveness of machine learning techniques in application to visual recognition. In this paper we provide an overview of applicable machine learning techniques. We make some conclusions about described algorithms performance and properties.

**Keywords:** *visual recognition, classification, machine learning, Support Vector Machine, Classifier Boosting, SNoW.*

#### About the author(s)

Vezhnevets Alexander is a student at Moscow State University, Department of Computational Mathematics and Cybernetics. His contact email is [vezhnick@gmail.com](mailto:vezhnick@gmail.com) . Main scientific interests lie in the fields of computer vision, image processing, machine learning, classification and statistics.