

На правах рукописи

Шаповалов Роман Викторович

**Методы структурного обучения
в задачах совместной разметки**

Специальность 01.01.09 —
дискретная математика и математическая кибернетика

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2014

Работа выполнена на кафедре математических методов прогнозирования факультета вычислительной математики и кибернетики Федерального государственного бюджетного учреждения высшего профессионального образования «Московский государственный университет имени М. В. Ломоносова»

Научный руководитель: **Ветров Дмитрий Петрович**,
кандидат физико-математических наук, доцент кафедры математических методов прогнозирования факультета вычислительной математики и кибернетики Федерального государственного бюджетного учреждения высшего профессионального образования «Московский государственный университет имени М. В. Ломоносова»

Официальные оппоненты: **Урмаев Олег Станиславович**,
доктор технических наук, главный эксперт по концептуальному инжинирингу — начальник департамента геологии и разработки новых активов ООО «Газпромнефть научно-технический центр»
Лемпицкий Виктор Сергеевич,
кандидат физико-математических наук, старший преподаватель Автономной некоммерческой образовательной организации высшего профессионального образования «Сколковский институт науки и технологий»

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Московский физико-технический институт (государственный университет)»

Защита состоится 27 марта 2014 г. в 11⁰⁰ часов на заседании диссертационного совета Д 501.001.44 при Московском государственном университете имени М. В. Ломоносова по адресу: 119991, г. Москва, ГСП-1, Ленинские горы, д. 1, стр. 52, 2-й учебный корпус, аудитория 685.

С диссертацией можно ознакомиться в Научной библиотеке Московского государственного университета имени М. В. Ломоносова по адресу: 119992, г. Москва, Ломоносовский проспект, д. 27 или на сайте <http://cs.msu.ru/theses/>.

Автореферат разослан « » _____ 2014 года.

Учёный секретарь
диссертационного совета Д 501.001.44,
доктор физико-математических наук, доцент

О. В. Шестаков

Общая характеристика работы

Актуальность темы. Задачей машинного обучения с учителем является восстановление функциональной зависимости между случайными величинами X и Y по обучающей выборке $\{(\mathbf{x}^j, y^j)\}_{j=1}^J$. В классической постановке задачи Y является скалярной случайной величиной, а пары (\mathbf{x}^j, y^j) получаются *независимой* выборкой из генеральной совокупности. Это позволяет прогнозировать значение y лишь по соответствующему значению \mathbf{x} . Однако во многих практических задачах это предположение о независимости не выполняется. В этом случае моделирование зависимости между переменными y^j позволяет повысить качество предсказания. Диссертация посвящена решению задач *совместной разметки* — классификации объектов с учётом зависимостей между ними.

Различные прикладные задачи могут быть сформулированы как задачи совместной разметки: семантическая сегментация изображений в компьютерном зрении, распознавание символов и определение частей речи в вычислительной лингвистике, определение структуры белка в биоинформатике, и многие другие. В этих задачах восстановление совместной разметки (в противоположность поэлементной) позволяет повысить качество прогнозирования.

Формализуем задачу совместной разметки. Пусть теперь вектор признаков \mathbf{x} объекта соответствует V векторам \mathbf{x} . Он может быть равен их конкатенации, но может также включать признаки взаимодействия отдельных элементов. По признакам \mathbf{x} необходимо определить вектор-разметку $\mathbf{y} \in \mathcal{K}^V$, где $\mathcal{K} \equiv \{1, \dots, K\}$ — множество меток отдельных компонент разметки, соответствующих переменным y , имеющих индексы $\mathcal{V} \equiv \{1, \dots, V\}$. Будем считать, что пары (\mathbf{x}, \mathbf{y}) получаются *независимой* выборкой из совместного распределения на случайные векторы X и Y . Прогнозирование разметки осуществляется с помощью вывода модальной разметки, на которой функция апостериорного распределения достигает максимума: $\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y}} P(\mathbf{y} | \mathbf{x})$. В реальных задачах для этого приходится рассматривать факторизации распределения $P(\mathbf{y} | \mathbf{x})$, делая предположения об условных независимостях между случайными величинами. Эти факторизации удобно описывать с помощью *марковских сетей*. Марковская сеть определяет *энергию* $E(\mathbf{y}; \mathbf{x})$ каждой из разметок \mathbf{y} , которой однозначно соответствует вероятность $P(\mathbf{y} | \mathbf{x}) \propto \exp(-E(\mathbf{y}; \mathbf{x}))$, при этом предполагается

следующий вид энергии:

$$E(\mathbf{y}; \mathbf{x}) = \sum_{f=1}^F \phi_f(\mathbf{y}_{C_f}), \quad (1)$$

где \mathbf{y}_{C_f} — проекция вектора \mathbf{y} на компоненты C_f , то есть вектор $[y_v]_{v \in C_f}$, а индекс f перечисляет т.н. *факторы* марковской сети. Фактор определяется соответствующим подмножеством переменных $C_f \subset \mathcal{V}$ и *потенциальной функцией* $\phi_f(\cdot)$. Количество соответствующих фактору переменных $|C_f|$ называют *порядком фактора*.

В некоторых задачах, таких как оценка оптического потока и стереосопоставление изображений, удаётся задать значение потенциальных функций из эвристических соображений. В других задачах необходимо восстанавливать функцию энергии $E_{\mathbf{w}}(\mathbf{y}; \mathbf{x})$ в параметрическом виде по обучающей выборке $\{(\mathbf{x}^j, \mathbf{y}^j)\}_{j=1}^J$. При этом вектор параметров \mathbf{w} находится как оптимум некоторой целевой функции. В *структурном методе опорных векторов* (структурном SVM) минимизируется следующая целевая функция:

$$L(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{j=1}^J \left[\max_{\bar{\mathbf{y}}} \{ \Delta(\bar{\mathbf{y}}; \mathbf{y}^j) - E_{\mathbf{w}}(\bar{\mathbf{y}}; \mathbf{x}^j) \} + E_{\mathbf{w}}(\mathbf{y}^j; \mathbf{x}^j) \right]. \quad (2)$$

Здесь C — гиперпараметр, контролирующий силу регуляризации, а $\Delta(\bar{\mathbf{y}}; \mathbf{y})$ — эмпирическая функция потерь, которая измеряет отклонение произвольной разметки $\bar{\mathbf{y}}$ от верной разметки \mathbf{y} . При использовании такой целевой функции параметры \mathbf{w} настраиваются так, что разница между энергией верной разметки и энергией любой другой разметки должна быть тем больше, чем больше они отличаются друг от друга.

Альтернативу использованию марковских сетей для учёта зависимостей между случайными величинами при решении задач совместной разметки представляет метод *последовательной классификации*. Решающее правило в этом случае сводится к последовательному применению классификаторов, множество которых разделено на *слои*. Классификаторы каждого следующего слоя принимают на вход выходы классификаторов предыдущего слоя, таким образом, разметка постепенно уточняется с учётом обновлённой предварительной разметки. При обучении может использоваться жадный алгоритм: классификаторы

обучаются по слоям, используя несмещённые оценки выходов классификаторов предыдущего слоя. Достоинством данного метода является концептуальная простота, как следствие, более высокая скорость прогнозирования, чем в марковских сетях.

Методы вывода модальной разметки с помощью марковских сетей и обучения структурного SVM требуют большого количества вычислительных ресурсов, однако в ходе обучения находится минимум верхней оценки функции потерь на обучающей выборке. Методы на основе последовательной классификации требуют меньше ресурсов, однако не обеспечивают оценок точности. При использовании больших обучающих выборок, доступных в настоящее время, оказывается выгодно использовать гибкие (в том числе нелинейные) модели, которые могут повысить точность прогнозирования. Подобные модели для задач совместной разметки в настоящее время исследованы мало. Недостатком обоих методов также является требовательность к аннотации обучающей выборки: для каждого объекта x^j должна быть известна разметка y^j , получение которой может быть трудозатратно. В работе рассматриваются слабые аннотации объектов z^j , представляющие собой в общем случае некоторые статистики от разметки y^j , такие что по аннотации z^j можно однозначно определить множество *совместных с ней разметок* $L(z^j)$. Предполагается, что задание такой аннотации пользователем, в отличие от полной разметки, занимает пренебрежимо малое время. Поэтому задача обучения по слабоаннотированным данным является актуальной.

Целью данной работы является сокращение требований, предъявляемых к аннотации обучающей выборки, повышение точности и скорости работы методов машинного обучения для решения задач совместной разметки.

Для достижения поставленной цели были решены следующие **задачи**:

1. Исследована модификация структурного SVM, при которой часть обучающей выборки размечена не полностью, а известна лишь слабая аннотация объектов. Предложена общая схема построения функций потерь для объектов, полная разметка которых недоступна, описаны несколько специальных функций потерь для конкретных видов слабой аннотации, а также методика их комбинирования в рамках одной оптимизационной задачи. Для этих

специализаций предложены алгоритмы оптимизации, используемые при настройке параметров модели.

2. Исследованы модификации структурного SVM, позволяющие обучать более гибкую модель марковской сети. В частности, рассмотрен аппарат *неассоциативных марковских сетей*, позволяющий учитывать отрицательные корреляции между переменными, который ранее редко использовался из-за трудностей при минимизации энергии. Исследована возможность ядерного перехода в структурном SVM и применение аналога гауссова ядра, а также предложена модификация функции потерь, позволяющая обучаться на данных с выраженным дисбалансом категорий.
3. Исследованы методы последовательной классификации для решения задач совместной разметки. Предложен метод обучения последовательной классификации, позволяющий учитывать априорные знания о структуре пространственных зависимостей между метками при анализе двумерных и трёхмерных изображений.

Практическая значимость диссертационной работы определяется широтой задач, которые могут быть классифицированы как задачи совместной разметки. Обязательным условием применимости описываемых методов является наличие аннотированной обучающей выборки, позволяющей избежать ручного проектирования алгоритма разметки, невозможного во многих случаях из-за сложной природы зависимостей разметок от признаков. Для экспериментальной апробации были выбраны задачи, для которых это характерно: семантическая сегментация двумерных изображений и трёхмерных облаков точек, а также категоризация текстовых документов. Задача семантической сегментации заключается в назначении каждому из элементов сцены (пикселей изображения или точек облака) метки одной из семантических категорий, а задача категоризации документов — в определении подмножества категорий (тегов), соответствующих текстовому документу. Для всех экспериментов использовались реальные наборы данных. Предложенные методы позволяют увеличить качество автоматической разметки, либо сократить трудозатраты на аннотирование обучающей выборки: оба улучшения имеют положительный экономический эффект.

Метод обучения структурного SVM по слабоаннотированным данным позволяет учитывать различные типы аннотации данных, отличные от полной разметки. Например, в задаче семантической сегментации изображений знание примерных расположений объектов и списка категорий, присутствующих на изображении, позволяет обучить модель, которая практически не уступает по качеству модели, обученной по полной разметке, хотя получение такой слабой аннотации значительно менее трудозатратно.

Метод обучения последовательной классификации, учитывающий пространственное взаиморасположение объектов, позволяет повысить качество сегментации изображений сцен, в которых присутствует некоторая структура расположения объектов, таких как комнатные сцены. Задача сегментации трёхмерных облаков точек, полученных внутри помещений, возникает в робототехнике, при этом важна вычислительная сложность используемых алгоритмов, поэтому последовательная классификация оказывается полезной.

Методы обучения нелинейных моделей (такие как структурный метод опорных векторов с гауссовыми ядрами и предложенный вариант последовательной классификации) практически полезны, так как позволяют сократить усилия разработчиков по предобработке признакового описания, при этом также позволяют достичь высокого качества разметки.

Научная новизна. В предложенном методе обучения по слабоаннотированным данным новой является общая схема построения функций потерь для разметки относительно известной слабой аннотации, предлагающая оценивать значение расстояния Хэмминга до множества разметок, совместных с данной слабой аннотацией. Эта схема применена при решении задач семантической сегментации изображений и категоризации документов для построения функций потерь, ранее не описанных в литературе. Задача оптимизации полученной целевой функции структурного метода опорных векторов с такими функциями потерь ранее не была исследована; стандартные приёмы для оптимизации возникающих при её решении дискретных оптимизационных подзадач неприменимы, поэтому в работе используются специализированные методы дискретной оптимизации, которые ранее не применялись при обучении структурного SVM.

Неассоциативная марковская сеть была впервые применена к задаче сегментации трёхмерных облаков точек, полученных лазерным сканированием. Ис-

пользован метод восстановления нелинейной зависимости функции энергии от параметров модели, описана гауссова ядровая функция для структурного SVM.

Модифицирована стандартная схема применения последовательной классификации. Введено понятие *типов факторов*, позволяющее комбинировать в схеме последовательной классификации различные типы взаимодействий между элементами сцены. Главным практическим эффектом является возможность определять *пространственные* типы факторов, позволяющие настраивать классификаторы так, чтобы учитывалось пространственное взаиморасположение объектов, присутствующих на изображении.

Достоверность и апробация работы. Достоверность изложенных в работе результатов обусловлена строгостью доказательств утверждений и экспериментальной апробацией методов на реальных данных. Основные результаты работы докладывались на конференции по фотограмметрическому компьютерному зрению и анализу изображений «PCV 2010» (г. Париж, Франция), на конференции по трёхмерному моделированию и обработке, визуализации и передаче трёхмерных изображений «IEEE 3DIMPVT 2011» (г. Ханчжоу, Китай), на конференциях по интеллектуализации обработки информации «ИОИ 2012» (г. Будва, Черногория) и математическим методам распознавания образов «ММРО 2013» (г. Казань), на конференции по компьютерному зрению и распознаванию образов «IEEE CVPR 2013» (г. Портленд, Орегон, США). Основные результаты по теме диссертации изложены в 7 научных публикациях, из них 4 — в изданиях, рекомендованных ВАК РФ. Личный вклад диссертанта в каждой из работ состоит в получении основных теоретических результатов, в частности, в доказательстве всех теорем в [1], и в экспериментальной апробации предложенных методов обучения структурного SVM в [1–3] и последовательной классификации в [4].

Объем и структура работы. Диссертация состоит из введения, четырех глав и заключения. Полный объем диссертации составляет 119 страниц с 19 рисунками, 8 таблицами и 5 листингами алгоритмов. Список литературы содержит 92 наименования.

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи,

сформулированы научная новизна и практическая значимость представляемой работы, а также кратко описана структура работы и введена базовая нотация.

Формально описан класс задач *совместной разметки*. Приведены примеры прикладных задач, принадлежащих ему, из различных областей: компьютерного зрения, обработки изображений, биоинформатики, вычислительной лингвистики. Приведён краткий обзор методов машинного обучения для решения задач совместной разметки: максимизация правдоподобия, методы максимизации отступа, методы последовательной классификации. Отмечены достоинства методов, в частности, теоретическая обоснованность первых двух, высокая скорость работы последних двух, и гибкость методов последовательной классификации. Отмечены проблемы существующих методов, такие как трудность учёта контекстуальной информации и получения размеченной обучающей выборки.

Тем самым обоснован выбор сокращения требований, предъявляемых к аннотации обучающей выборки, повышения точности и скорости работы методов машинного обучения для решения задач совместной разметки в качестве цели диссертационного исследования.

Первая глава носит обзорный характер. В ней приведены теоретические результаты и методы, на которые опирается исследование. В начале главы формально определено понятие разметки, а также описаны ненаправленные графические вероятностные модели (марковские сети): рассмотрена факторизация распределения, представимая марковскими сетями, а также задачи вероятностного вывода, связанные с ними.

Далее приводятся методы нахождения моды распределения, представленного марковской сетью, большинство из которых являются приближёнными, так как задача точного поиска моды в общем виде является NP-трудной. Она является частным случаем задачи целочисленного линейного программирования, которую можно решать приближённо с помощью LP-релаксации. Описаны методы передачи сообщений, которые выдают точный ответ для ациклических марковских сетей, однако в общем случае не гарантирована даже их сходимость. Описан метод двойственного разложения, оптимизирующий двойственную задачу к задаче поиска моды, который позволяет, в частности, учитывать факторы высоких порядков. Описаны методы нахождения моды на основе решения задачи построения минимального разреза в графе, которые дают точный ответ для

широкого класса бинарных задач (в которых число меток $K = 2$), а также служат основой аппроксимаций для задач с несколькими метками и факторами высоких порядков специального вида.

Формально определена задача *структурного обучения* как задача восстановления зависимости между случайными величинами \mathbf{X} и \mathbf{Y} по обучающей выборке $\{(\mathbf{x}^j, \mathbf{y}^j)\}_{j=1}^J$, где пространство правильных ответов состоит из сложных комбинаторных объектов (в данной работе — разметок). Для восстановления этой зависимости функцию апостериорного распределения часто представляют в параметрическом виде. Параметры находятся как оптимальные значения некоторой целевой функции. Рассматривается *логлинейная зависимость* апостериорной вероятности от параметров \mathbf{w} :

$$P(\mathbf{y} | \mathbf{x}, \mathbf{w}) = \frac{1}{Z(\mathbf{x}, \mathbf{w})} \exp \left(\sum_{f=1}^F \mathbf{w}^\top \boldsymbol{\psi}^{t(f)}(\mathbf{y}_{\mathcal{C}_f}; \mathbf{x}_f) \right) = \frac{1}{Z(\mathbf{x}, \mathbf{w})} \exp(\mathbf{w}^\top \boldsymbol{\psi}(\mathbf{y}; \mathbf{x})), \quad (3)$$

где Z — нормировочная константа, необходимая, для того, чтобы распределение суммировалось к 1, \mathbf{x}_f — часть вектора признаков, имеющая отношение к фактору f , $\boldsymbol{\psi}^{t(f)}(\mathbf{y}_{\mathcal{C}_f}; \mathbf{x}_f)$ — функция, возвращающая вектор *обобщённых признаков* фактора f , длина которого равна длине вектора параметров \mathbf{w} , а $\boldsymbol{\psi}(\mathbf{y}; \mathbf{x}) = \sum_{f=1}^F \boldsymbol{\psi}^{t(f)}(\mathbf{y}_{\mathcal{C}_f}; \mathbf{x}_f)$. Тип фактора $t(f)$ определяет тип зависимости, моделируемой вектором обобщённых признаков; например, он может разделять унарные и парные потенциалы (порядка 1 и 2, соответственно). Функция $\boldsymbol{\psi}^{t(f)}(\mathbf{y}_{\mathcal{C}_f}; \mathbf{x}_f)$ «копирует» вектор \mathbf{x}_f в компоненты, соответствующие параметрам, отвечающим за данный тип факторов и конфигурацию переменных \mathcal{C}_f , а остальные компоненты оставляет нулевыми. Таким образом, размерность вектора параметров \mathbf{w} равна $\sum_{t \in \mathcal{T}} K^{C_t} D_t$, где \mathcal{T} — множество всех типов факторов, $C_t = |\mathcal{C}_f|$ — порядок любого факторов f типа t , D_t — размерность векторов признаков, соответствующих факторам типа t .

Рассмотрены различные варианты целевых функций в задачах структурного обучения и методов оптимизации. В частности, определена функция правдоподобия параметров \mathbf{w} при известной обучающей выборке, и описан метод максимизации правдоподобия. На практике эта максимизация часто оказывается невозможной: вычисление функции правдоподобия и её градиентов требует оценки нормировочной константы Z при данном значении параметров. Она яв-

ляется суммой, число слагаемых в которой экспоненциально зависит от числа переменных в разметке V . В марковских сетях с циклами не существует эффективного метода её вычисления.

Структурный метод опорных векторов (структурный SVM) не требует подсчёта нормировочной константы. Поскольку при выводе разметки интерес представляет не распределение $P(y | x, w)$ само по себе, а лишь его мода, подбор параметров можно проводить из следующих соображений: объекты обучающей выборки должны иметь вероятность, максимизирующую функцию распределения, причём эта вероятность должна иметь как можно больший отступ от второй по вероятности разметки. При использовании такого критерия игнорируются значения функции распределения во всех точках, кроме этих двух. Другой важной особенностью метода является использование функции потерь $\Delta(\bar{y}; y)$, которая измеряет отклонение произвольной разметки \bar{y} от верной разметки y . Минимизируемая целевая функция выглядит следующим образом:

$$L_{MM}(w) = \frac{1}{2} w^T w + C \sum_{j=1}^J \left[\max_{\bar{y}} \{ \Delta(\bar{y}; y^j) + w^T \psi(\bar{y}; x^j) \} - w^T \psi(y^j; x^j) \right]. \quad (4)$$

Здесь C — гиперпараметр, контролирующий силу регуляризации.

В конце главы также рассмотрены два метода настройки *нелинейных параметризаций* марковской сети: функциональный градиентный бустинг и поле решающих деревьев.

Во **второй главе** разработан метод структурного обучения для задач разметки, позволяющий учитывать различные виды *слабой аннотации* обучающей выборки. Необходимость в этом возникает на практике, так как аннотация достаточного количества обучающих объектов часто представляет собой наиболее затратную часть создания системы совместной разметки. Слабая аннотация представляет собой некоторую статистику полной разметки, для получения которой требуется меньше человеческих усилий. Например, при решении задачи семантического разбора предложений слабая аннотация обучающей выборки может быть представлена разметкой частей речи, а в ряде задач анализа видеопоследовательностей разметка может быть дана только для ключевых кадров.

В этой главе целевыми приложениями являются задачи семантической сегментации изображений и категоризации документов. Примерами слабых ан-



Рис. 1: Различные типы аннотаций для изображения из набора данных MSRC

нотаций в первой служат метки изображения, которые отражают присутствие или отсутствие категорий; метки площади, которые задают число пикселей каждой категории на изображении; набор плотных рамок для объектов, присутствующих в разметке; а также набор зёрен — подмножеств координат пикселей, принадлежащих объектам (рис. 1). В задаче категоризации документов разметка текстового документа представляет собой подмножество тегов (категорий) некоторого допустимого множества. При получении такой разметки легко пропустить некоторые категории. Таким образом, слабой аннотацией документа может являться некоторое подмножество реального множества категорий.

Формально, *слабой аннотацией* экземпляра обучающей выборки называется любой объект \mathbf{z} , для которого однозначно определяется множество разметок $L(\mathbf{z}) \subseteq \mathcal{Y}$, *совместных со слабой аннотацией*. Рассмотрен случай, когда помимо J полностью размеченных объектов, обучающая выборка содержит I слабо аннотированных: $\{(\mathbf{x}^i, \mathbf{z}^i)\}_{i=J+1}^{J+I}$. Предложено следующее обобщение целевой функции (4) в методе максимизации отступа на случай присутствия в обучающей выборке полностью размеченных и слабоаннотированных данных.

Оптимизационная задача 1.

$$\min_{\mathbf{w}, \xi, \eta} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{C}{J+I} \left(\sum_{j=1}^J \xi_j + \alpha \sum_{i=J+1}^{J+I} \eta_i \right), \quad (5)$$

$$\text{п. у.} \quad \mathbf{w}^\top \boldsymbol{\psi}(\mathbf{y}^j; \mathbf{x}^j) \geq \max_{\bar{\mathbf{y}} \in \mathcal{Y}} \{ \mathbf{w}^\top \boldsymbol{\psi}(\bar{\mathbf{y}}; \mathbf{x}^j) + \Delta(\bar{\mathbf{y}}; \mathbf{y}^j) \} - \xi_j, \quad \forall j, \quad (6)$$

$$\max_{\mathbf{y} \in L(\mathbf{z}^i)} \mathbf{w}^\top \boldsymbol{\psi}(\mathbf{y}; \mathbf{x}^i) \geq \max_{\bar{\mathbf{y}} \in \mathcal{Y}} \{ \mathbf{w}^\top \boldsymbol{\psi}(\bar{\mathbf{y}}; \mathbf{x}^i) + K(\bar{\mathbf{y}}; \mathbf{z}^i) \} - \eta_i, \quad \forall i. \quad (7)$$

Здесь $K(\bar{y}, \mathbf{z})$ — *слабая функция потерь*, задающая степень несогласованности некоторого ответа $\bar{y} \in \mathcal{Y}$ со слабой аннотацией \mathbf{z} , ξ_j и η_i — минимизируемые нарушения ограничений, α — коэффициент, балансирующий вклад сильно- и слабоаннотированных объектов.

Задача 1 является обобщением метода максимизации отступа с латентными переменными. Оптимизацию в ней тоже можно выполнять с помощью выпукло-вогнутой процедуры. Поскольку задача не является выпуклой, гарантировано лишь нахождение некоторого локального минимума, и выбор алгоритма оптимизации может сильно влиять на результат. Установлена связь задачи с максимизацией неполного правдоподобия следующего вида (для краткости будем игнорировать полностью размеченную часть выборки):

$$L(\mathbf{w}) = \prod_{j=1}^J \sum_{\bar{y} \in \mathcal{Y}} P(\bar{y}, \mathbf{z}^j \mid \mathbf{x}^j, \mathbf{w}) = \prod_{j=1}^J \sum_{\bar{y} \in \mathcal{Y}} P(\bar{y} \mid \mathbf{x}^j, \mathbf{w}) P(\mathbf{z}^j \mid \bar{y}, \mathbf{x}^j, \mathbf{w}) \quad (8)$$

при традиционной параметризации апостериорной вероятности: $P(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) \propto \exp(\mathbf{w}^\top \boldsymbol{\psi}(\mathbf{y}; \mathbf{x}))$ и следующем виде правдоподобия разметки при известной аннотации: $P(\mathbf{z} \mid \mathbf{y}, \mathbf{x}, \mathbf{w}) = P(\mathbf{z} \mid \mathbf{y}) = \mathbb{I}[\mathbf{y} \in L(\mathbf{z})]$. Доказана следующая теорема.

Теорема 1. Пусть слабая функция потерь неотрицательна: $K(\mathbf{y}; \mathbf{z}) \geq 0$. Тогда, при условии равенства начальных приближений \mathbf{w}^0 , выпукло-вогнутая процедура минимизации целевой функции обобщённого SSVМ (5) сходится к тому же вектору \mathbf{w}^* , что и EM-алгоритм для максимизации апостериорного распределения, соответствующего правдоподобию (8), со следующими модификациями:

- на E-шаге оценка математического ожидания производится не по действительному распределению на латентные переменные, а по его точечной MAP-оценке;
- на M-шаге максимизируется не полученная на E-шаге оценка математического ожидания, а её нижняя оценка, где логарифмы нормировочных констант распределений на латентные разметки слабоаннотированных объектов оцениваются согласно оценке

$$-\log \sum_{\tilde{y} \in \mathcal{Y}} \exp(\mathbf{w}^\top \boldsymbol{\psi}(\tilde{y}; \mathbf{x})) \geq -\max_{\tilde{y} \in \mathcal{Y}} \{\mathbf{w}^\top \boldsymbol{\psi}(\tilde{y}; \mathbf{x}) + K(\tilde{y}; \mathbf{z})\} + \text{const.} \quad (9)$$

Далее формально определяются три типа слабой аннотации в задаче семантической сегментации изображений: аннотации метками изображе-

ния (рис. 1e), плотными рамками (рис. 1c) и зёрнами объектов (рис. 1d). Для использования каждого из этих видов слабой аннотации при обучении определены функции потерь K , специфичные для типов аннотаций. Описаны эффективные алгоритмы вывода, дополненного функций потерь (максимизация в правой части ограничения (7)) и вывода, согласованного с аннотацией (максимизация в левой части ограничения). Также показано, что некоторые типы аннотаций могут комбинироваться в рамках одного изображения. Для каждой из введённых функций потерь доказана *теорема о несмещённости оценки* её параметров.

В задаче **категоризации документов** разметка представляет собой множество категорий (тегов), характерных для документа. Слабая аннотация *частичной разметкой* определена так, что для каждого документа про некоторые категории известно, характерны они или нет для данного документа, а про остальные информация отсутствует. Определена функция потерь для такого типа аннотации. Доказано, что она является *несмещённой оценкой функции потерь по полной разметке*. Описаны эффективные алгоритмы для вывода, дополненного функцией потерь и вывода, согласованного с аннотацией.

Экспериментальная валидация методов семантической сегментации проведена на двух наборах данных: *MSRCv2* и *SIFT-flow*, в которых выделены 21 и 29 категорий соответственно. Обучающая выборка в первом состоит из 276 фотографий, во втором — из 2488. Первый эксперимент заключается в том, чтобы для части изображений обучающей выборки оставить трудоёмкую полную разметку, а для остальных сгенерировать аннотацию метками изображения. Исследована точность сегментации в зависимости от доли полностью размеченных изображений. На наборе *MSRC* достаточно полностью разметить лишь седьмую часть обучающей выборки, чтобы получить точность, близкую к обучению на полной выборке: она уменьшается лишь на 4 процентных пункта, в то время как при отсутствии слабоаннотированных объектов она уменьшается на 11 п. п. На наборе *SIFT-flow* обучение с разметкой лишь 10% выборки уступает обучению с полностью размеченной выборкой всего 4 п. п. В следующем эксперименте исследуется добавление аннотаций плотными рамками и зёрнами. Показано, что они позволяют повысить качество сегментации объектных категорий. На наборе *MSRC* при аннотации всей выборки метками изображений и плотными рамками (без полной разметки вообще) проигрыш в качестве по срав-

нению с обучением по полной разметке составляет всего 5 п. п. Зёрна также улучшают точность, но в меньшей степени, однако они проще в получении, чем плотные рамки.

Для экспериментов по категоризации документов использовалась база данных юридических документов *EUR-lex*, состоящая из 17413 обучающих и 1935 тестовых документов, каждый из которых описан 5000 признаками на основе частот встречаемости терминов. Каждому документу назначено подмножество множества из 201 категории. Проведено обучение на выборке, для части которой известна полная разметка, а для остальных — часть категорий скрыты. При использовании лишь 10 % полной разметки потеря точности составляет лишь 2 п. п.

В третьей главе предложена параметризация *неассоциативной* парно-сепарабельной марковской сети, а также предложен алгоритм обучения её параметров на основе структурного SVM. Кроме того, предложена модификация функции потерь на основе Хэммингова расстояния, позволяющая настраивать параметры в случае, когда разные категории представлены в данных несбалансированно. Также показано, как обучать нелинейный структурный SVM с гауссовскими ядрами (англ. *gaussian radial basis function, RBF*). Эксперименты на облаках точек, полученных лазерным сканированием с движущихся транспортных средств, показывают, что эти нововведения позволяют улучшить качество сегментации.

Большинство прикладных задач разметки используют *ассоциативные* марковские сети. В этом случае поощряется назначение одной и той же метки переменным из одного фактора. Популярность ассоциативных марковских сетей объясняется тем, что такой вид энергии допускает эффективные алгоритмы минимизации на основе разрезов на графах. Ассоциативность может также служить формой регуляризации. Если известно, что связанные парными потенциалами переменные не могут быть отрицательно коррелированы, то модель не будет настроена на такие шумовые зависимости в данных. Однако это предположение не всегда выполняется. Например, в задачах семантической сегментации естественных сцен марковская сеть может включать связи между удалёнными регионами, про которые известно, что они с большой вероятностью принадлежат разным категориям. Если связь соответствует суперпикселям, находящимся

друг над другом вверху и внизу изображения, то они вероятнее принадлежат категориям ‘небо’ (верхний) и ‘трава’ (нижний), чем оба одновременно к любой из этих категорий.

Определены **неассоциативные** парные потенциалы в марковской сети с логлинейной параметризацией (3) и потенциалами, имеющими порядок не более двух:

$$\mathbf{w}^\top \psi_{vu}(y_v, y_u; \mathbf{x}) = \sum_{k \in \mathcal{K}} \sum_{l \in \mathcal{K}} \mathbb{I}[y_v = k] \mathbb{I}[y_u = l] (\mathbf{x}_{vu}^\top \mathbf{w}_{kl}^e) = \mathbf{x}_{vu}^\top \mathbf{w}_{y_v y_u}^e. \quad (10)$$

В отличие от используемых ранее, в этом потенциале не накладываются ограничений на знаки, и он не является метрикой относительно (y_v, y_u) . Поэтому методы оптимизации на основе разрезов графов неприменимы. Для вывода, в том числе дополненного функцией потерь, используется алгоритм передачи сообщений на деревьях с перевзвешиванием (TRW), возвращающий приближённое решение задачи. Обучение структурного SVM с таким выводом эквивалентно ослаблению части ограничений, поэтому такой процесс называется оптимизацией на расширенном множестве.

Далее решается проблема **несбалансированности обучающей выборки** по категориям. Фоновые категории, такие как ‘земля’, обычно содержат гораздо больше суперпикселей, чем объектные категории, такие как ‘автомобиль’ или ‘столб’. Введена следующая эмпирическая функция потерь, позволяющая учитывать штрафы за ошибки на отдельных категориях:

$$\Delta(\bar{\mathbf{y}}; \mathbf{y}) = \sum_{v \in \mathcal{V}} c_v r_{y_v} \mathbb{I}[\bar{y}_v \neq y_v]. \quad (11)$$

Здесь \mathbf{y} — верная разметка, $\bar{\mathbf{y}}$ — произвольная разметка, c_v — количество точек в суперпикселе, а r_k — штраф за неправильную классификацию суперпикселя категории k . Доказана лемма об эквивалентности функции потерь (11) средней по классам полноте при назначении штрафов r_k пропорциональными обратной частоте встречаемости точек данной категории в выборке.

Далее описана модификация структурного SVM, позволяющая учитывать **нелинейные зависимости**. Для этого приведена его двойственная формулировка. Векторы обобщённых признаков участвуют в ней только как множители скалярных произведений с другими векторами обобщённых признаков. Та-

ким образом, возможно выполнить ядровой переход, используя вместо скалярного произведения так называемую *ядровую функцию*. Приведён пример такой функции, имеющей практическое значение. *Гауссовской ядровой функцией* (англ. *gaussian radial basis function, RBF*) называется функция следующего вида:

$$q(\mathbf{y}', \mathbf{x}', \mathbf{y}'', \mathbf{x}'') = \sum_{v' \in \mathcal{V}'} \sum_{v'' \in \mathcal{V}''} \exp(-\gamma \|\mathbf{x}'_{v'} - \mathbf{x}''_{v''}\|^2) \mathbb{I}[y'_{v'} = y''_{v''}] + \sum_{\substack{(v', u') \\ \in \mathcal{E}'}} \sum_{\substack{(v'', u'') \\ \in \mathcal{E}''}} \exp(-\gamma \|\mathbf{x}'_{v'u'} - \mathbf{x}''_{v''u''}\|^2) \mathbb{I}[y'_{v'} = y''_{v''}] \mathbb{I}[y'_{u'} = y''_{u''}]. \quad (12)$$

Здесь под \mathcal{V}' и \mathcal{V}'' понимаются множества вершин марковских сетей, образующих \mathbf{x}' и \mathbf{x}'' , соответственно, а под \mathcal{E}' и \mathcal{E}'' — множества их рёбер. Параметр γ отражает ширину ядра, он полагается равным 1.

Гауссовская ядровая функция измеряет расстояние между разметками с учётом близости признаков. Рассматриваются все пары унарных и все пары парных потенциалов, и чем ближе соответствующие признаки, тем больший вклад в расстояние дают неодинаковые значения переменных в соответствующих компонентах разметок.

Экспериментальная валидация проводилась на двух наборах данных, полученных лазерным сканированием с движущихся самолёта и автомобиля, соответственно. В первом присутствуют три категории, а обучающая и тестовая выборки содержат по 100 000 точек. Во втором присутствуют четыре категории, обучающая выборка содержит 0.4 миллиона точек, тестовая — 1 миллион. На первом наборе данных предложенный метод с гауссовым ядром показал среднегеометрическую полноту выше, чем метод с линейным ядром, и чем метод с функцией потерь, не сбалансированной по категориям. Также сравнение осуществлялось с другими используемыми в этой задаче методами: независимой классификацией ансамблем решающих деревьев и нелинейной ассоциативной марковской сетью, обученной функциональным градиентным бустингом. Предложенный метод показал более точный результат, подтвердив полезность неассоциативных потенциалов. На втором наборе данных предложенный метод качественно сегментировал три из четырёх категорий, но не смог найти точки категории ‘столбы’, сильно недопредставленной в обучающей выборке.

Четвёртая глава посвящена исследованию методов *последовательной классификации* для решения задач разметки. В отличие от графических моделей, решающее правило сводится к последовательному применению классификаторов, которые обучаются с помощью жадного алгоритма. Классификаторы каждого следующего слоя зависят от выходов классификаторов предыдущего слоя, таким образом, разметка последовательно уточняется с учётом обновлённой предварительной разметки. Достоинствами данного метода являются концептуальная простота и, как следствие, более высокая скорость вывода, чем в графических моделях. Гибкость метода позволяет моделировать произвольные зависимости между метками, причём не только локальные, как в графических моделях.

В главе описана модель последовательной классификации — **пространственная машина вывода**, которая позволяет учитывать априорные знания о структуре задачи. Основным инструментом для этого является *д-фактор*, который может относиться к одному из *типов факторов*. Д-фактором называется пара $p = (d_f, \mathcal{S}_f)$, состоящая из приёмника — переменной $d_f \in \mathcal{V}$ и передатчика — множества переменных $\mathcal{S}_f \subset \mathcal{V}$. В ходе вывода разметки итеративно вычисляется *функция-предиктор* сообщения $\mathbf{g}_{t(f)}^n(\cdot)$, которая на n -й итерации для типа факторов $t(f)$ имеет следующий вид:

$$\mu_{\mathcal{S}_f \rightarrow d_f}^n = \mathbf{g}_{t(f)}^n(\mathbf{b}_{d_f}^{n-1}, \mathbf{x}_{d_f}, \mathbf{x}_f, \mathbf{x}_{\mathcal{S}_f}, \frac{1}{|\mathcal{S}_f|} \sum_{v \in \mathcal{S}_f} \mathbf{b}_v^{n-1}). \quad (13)$$

Она подбирается в семействе ансамблей решающих деревьев с помощью алгоритма random forest, однако может использоваться любой другой классификатор, допускающий вероятностный выход. В качестве аргументов предиктора используются признаки д-фактора \mathbf{x}_f и усреднённые убеждения о метках в множестве-передатчике \mathbf{b}_v^{n-1} , $v \in \mathcal{S}_f$, а также признаки приёмника \mathbf{x}_{d_f} и убеждения о его метке с предыдущей итерации $\mathbf{b}_{d_f}^{n-1}$. Первые позволяют получить качественную классификацию по локальным признакам уже на первой итерации, а последние позволяют получить тождественную функцию (по отношению к убеждениям приёмника) на последних итерациях, когда остальные параметры малоинформативны. Кроме того, аргумент может включать некоторые признаки передатчика $\mathbf{x}_{\mathcal{S}_f}$. Конкатенация всех аргументов функции-предиктора называется *расширенными признаками*.

Переменной v могут соответствовать несколько д-факторов, имеющих эту переменную приёмником: в этом случае вероятностные выходы предикторов необходимо агрегировать. На n -й итерации вектор убеждений относительно метки y_v определяется как нормированное взвешенное произведение сообщений д-факторов из \mathcal{S}_f в v :

$$b_v^n(y) \propto \prod_{f:d_f=v} \left(\mu_{\mathcal{S}_f \rightarrow v}^n(y) \right)^{\alpha_{t(f)}^n}, \quad \forall y \in \{1, \dots, K\}, \quad (14)$$

где $\alpha_{t(f)}^n$ — настраиваемый параметр, соответствующий вкладу типа факторов $t(f)$.

Таким образом, алгоритм вывода работает следующим образом: начальные убеждения \mathbf{b}_v^0 инициализируются равномерными распределениями, затем поочерёдно вычисляются (13) и (14) в течение фиксированного количества итераций N . Аргмаксимумы финальных убеждений \mathbf{b}_v^N принимаются за значения соответствующих переменных разметки.

Алгоритм **обучения** также выполняется итеративно. На каждой итерации настраиваются соответствующие предикторы, см. алгоритм 1. В качестве правильных ответов берутся переопределённые представления меток, заданных для объектов обучающей выборки: $\Upsilon_v(k) \equiv \llbracket y_v = k \rrbracket$, $\forall v \in \mathcal{V}, \forall k \in \mathcal{K}$. Среди аргументов функций-предикторов сообщений присутствуют убеждения с предыдущей итерации, поэтому для их обучения необходимо получить несмещённую оценку убеждений, предсказываемых на этапе вывода, это выполняется в строках 5–12 алгоритма. Также имеет смысл настраивать весовые коэффициенты агрегации сообщений α^n (строка 16) с использованием этих оценок.

Практическая значимость модели продемонстрирована на примере задачи семантической сегментации. Предполагается, что модель определяется в некоторой двумерной или трёхмерной визуальной сцене, состоящей из *элементов* — пикселей, вокселей, точек или суперпикселей, соответствующих переменным в задаче разметки. Элемент v характеризуется координатами $\mathbf{p}_v = (x, y)$ в двумерном пространстве или $\mathbf{p}_v = (x, y, z)$ в трёхмерном. Определены два **семейства типов факторов**: пространственные и структурные.

Пространственные д-факторы — семейство типов факторов, моделирующих пространственное взаиморасположение элементов сцены. Тип факторов t

Алгоритм 1 Обучение пространственной машины вывода

- 1: **Вход:** размеченная выборка (\mathbf{x}, \mathbf{y}) , множество д-факторов обучающей выборки \mathcal{F} , разделённое на части \mathbf{f} , множество типов факторов \mathcal{T} , число итераций вывода N .
 - 2: **Выход:** набор пар функций-предикторов сообщений и весов $\{(g_{n,t}(\cdot), \alpha_t^n)\}_{t \in \mathcal{T}, n \in \{1, \dots, N\}}$
 - 3: инициализировать $\mathbf{b}_v^0 = \frac{1}{K}, \forall v \in \mathcal{V}$
 - 4: **for** $n = 1$ to N **do**
 - 5: **for all** $\mathbf{f} \in \mathcal{F}$ **do**
 - 6: **for all** $t \in \mathcal{T}$ **do**
 - 7: обучить предиктор $\mathbf{g}_t^{imp}(\cdot)$ так, чтобы $\Upsilon_{d_f} \approx \mathbf{g}_t^{imp}(\langle \text{расширенные признаки } f \rangle)$
на выборке д-факторов $\{f \in \bigcup_{\mathbf{f}' \in \mathcal{F} \setminus \{\mathbf{f}\}} \mathbf{f}' \mid t(f) = t\}$
 - 8: **end for**
 - 9: **for all** $f \in \mathbf{f}$ **do**
 - 10: $\mu_{S_f \rightarrow d_f}^{imp} \leftarrow \mathbf{g}_{t(f)}^{imp}(\mathbf{b}_{d_f}^{n-1}, \mathbf{x}_{d_f}, \mathbf{x}_f, \mathbf{x}_{S_f}, \frac{1}{|S_f|} \sum_{v \in S_f} \mathbf{b}_v^{n-1})$
 - 11: **end for**
 - 12: **end for**
 - 13: **for all** $t \in \mathcal{T}$ **do**
 - 14: обучить предиктор $\mathbf{g}_t^n(\cdot)$ так, чтобы $\Upsilon_{d_f} \approx \mathbf{g}_t^n(\langle \text{расширенные признаки } f \rangle)$
на выборке д-факторов $\{f \in \bigcup_{\mathbf{f}' \in \mathcal{F}} \mathbf{f}' \mid t(f) = t\}$
 - 15: **end for**
 - 16: задать веса типов факторов α^n , например $\alpha^n = \mathbf{1}$ или настроить по выборке
 - 17: **if** $n < N$ **then**
 - 18: **for all** $v \in \mathcal{V}$ **do**
 - 19: вычислить убеждения \mathbf{b}_v^n по сообщениям $\mu_{S_f \rightarrow v}^{imp}$ согласно (14)
 - 20: **end for**
 - 21: **end if**
 - 22: **end for**
-

однозначно задаётся регионом координатного пространства \mathcal{P}_t . Для элемента пространства, соответствующего переменной v с координатами \mathbf{p}_v , порождается д-фактор (v, \mathcal{S}) , где в \mathcal{S} входят переменные, соответствующие всем элементам из региона $\mathcal{P}_t + \mathbf{p}_v = \{\mathbf{p}_t + \mathbf{p}_v \mid \mathbf{p}_t \in \mathcal{P}_t\}$. Пространственные типы факторов позволяют моделировать произвольные дальнедействующие контекстуальные зависимости между переменными. На рис. 2 показаны четыре пространственных д-фактора. *Структурные* д-факторы — тип факторов, моделирующих локальные зависимости. Для элемента сцены, соответствующего переменной v с координатами \mathbf{p}_v , порождаются д-факторы $(v, \{u\})$ для каждой из переменных u , таких что \mathbf{p}_u принадлежит некоторой окрестности \mathbf{p}_v (например, входит в её k ближайших соседей). Всем структурным д-факторам соответствуют одинаковые предикторы \mathbf{g}_t^n и весовые коэффициенты α_t^n . На рис. 2 каждая пара соседних пикселей связана структурными д-факторами.

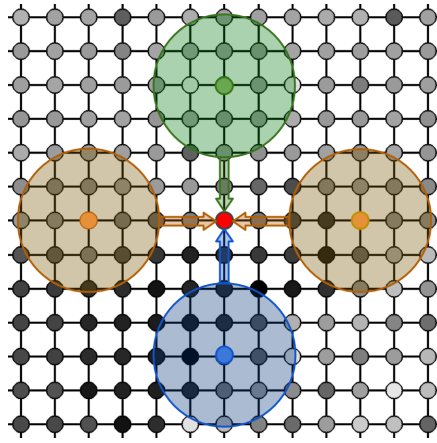


Рис. 2: Иллюстрация определения структурных и пространственных д-факторов для фрагмента изображения; переменные модели показаны серыми кругами и соответствуют пикселям изображения

Глава завершается описанием **экспериментальной валидации** разработанного метода. Для этого используются два набора данных, представляющих собой зарегистрированные карты глубины и RGB-изображения, полученные датчиком Kinect. Осуществляется семантическая сегментация облаков точек на 17 категорий. Сравниваются три вариации предлагаемого метода: модель с только структурными д-факторами, а также две модели с пространственными типами факторов: с настройкой весов α и без неё. Результаты показывают, что добавление пространственных типов факторов позволяет повысить точность сегментации, но только при условии настройки весов; в противном случае точность падает из-за учёта шумовых взаимодействий. Вектор оптимальных весов пространственных типов факторов оказался разреженным; выявлены наиболее значимые зависимости в данных, соответствующие ненулевым весам. Также установлено, что метод превосходит парно-сепарабельную марковскую сеть, обученную методом максимизации отступа, как по точности на скользящем контроле, так и по скорости вывода.

В **заключении** приведены основные результаты, полученные в работе.

Основные положения, выносимые на защиту:

- методы, обобщающие структурный SVM для обучения нелинейной неассоциативной марковской сети по слабоаннотированным данным, а также метод, позволяющий учитывать дальнедействующие пространственные зависимости при последовательной классификации;

- методика назначения функций потерь структурного SVM, учитывающих особенности обучающей выборки;
- экспериментальная апробация предложенных методов, сравнение точности и скорости работы с существующими методами.

Публикации автора в изданиях из списка ВАК

1. Обучение алгоритма семантической сегментации изображений на выборке с разнообразными типами аннотаций / Роман Шаповалов, Дмитрий Ветров, Антон Осокин [и др.] // Интеллектуальные системы. 2014. Т. 18, № 3. с. 81–107.
2. Семантическая сегментация данных лазерного сканирования / Роман Шаповалов, Александр Велижев, Ольга Барина [и др.] // Программные продукты и системы. 2012. № 1. С. 47–52.
3. Shapovalov Roman, Velizhev Alexander. Cutting-Plane Training of Non-associative Markov Network for 3D Point Cloud Segmentation // IEEE International Conference on 3D Imaging, Modeling, Processing, Visualisation and Transmittion. Hangzhou, China: 2011. С. 1–8. URL: <http://shapovalov.ro/papers/Shapovalov-Velizhev-3dimpvt2011.pdf>.
4. Shapovalov Roman, Vetrov Dmitry, Kohli Pushmeet. Spatial Inference Machines // IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR: 2013. С. 2985–2992. URL: <http://shapovalov.ro/papers/SIM-Shapovalov-et-al-CVPR2013.pdf>.

Прочие публикации автора

1. Шаповалов Роман, Ветров Дмитрий, Коли Пушмит. Учёт дальнодействующих зависимостей в задаче семантической сегментации трёхмерных облаков точек с помощью последовательной классификации // Доклады 16-й Всероссийской конференции «Математические методы распознавания образов» (ММРО-16). Казань, Россия: 2013. с. 51.
2. Шаповалов Роман, Осокин Антон, Ветров Дмитрий. Обучение структурного метода опорных векторов со слабым учителем в задачах сегментации изображений // Доклады 9-й Международной конференции «Интеллектуализация обработки информации» (ИОИ-2012). Будва, Черногория: 2012. С. 452–455.
3. Shapovalov Roman, Velizhev Alexander, Barinova Olga. Non-associative Markov networks for 3D point cloud classification // Photogrammetric Computer Vision and Image Analysis. IAPRS, vol. 38, part 3A. Paris, France: 2010. С. 103–108. URL: <http://shapovalov.ro/papers/Shapovalov-et-al-PCV2010.pdf>.