

Многоклассовая модель формы со скрытыми переменными

А.Н. Кириллов¹, М.И. Гавриков¹, Е.М. Лобачева¹, А.А. Осокин¹ и Д.П. Ветров^{1, 2}

akirillov@cs.msu.ru, gavrmike@gmail.com, lobacheva.tjulja@gmail.com, osokin@bayesgroup.ru, vetrovd@yandex.ru

¹Московский Государственный Университет им. М.В. Ломоносова

²Национальный исследовательский университет Высшая школа экономики

27 июля 2014 г.

Аннотация

Модель формы Больцмана (SBM, [1]) и ее многоклассовый вариант (MSBM, [2]) являются генеративными моделями, позволяющими задавать форму объекта. Многоклассовая модель обладает большей выразительной способностью, чем бинарная, но для ее обучения требуются данные с полной (многоклассовой) разметкой. В данной работе предложен новый алгоритм обучения MSBM, требующий только бинарные маски объектов и семена, задающие приближенное положение их частей. Семена при этом могут быть получены автоматически на основе использования детектора частей объекта. На этапе экспериментов мы показали, что MSBM, обученная нашим алгоритмом, превосходит по качеству SBM в задачах, где доступна только бинарная маска, и сопоставима по качеству с MSBM, обученной на данных с полной разметкой.

1 Введение

Модели формы объекта используются в различных задачах компьютерного зрения: сегментации [3, 4], восстановлении части изображения [5], детектировании объекта [6]. Учитывая знания о форме объекта в задаче сегментации можно значительно улучшить качество решения [2].

Существует множество способов задавать форму объекта [4, 7, 8], однако большинство из них либо недостаточно гибкие, либо требуют сложной дополнительной разметки обучающей выборки. SBM позволяет задавать форму объекта с помощью глубокой сети определенной структуры [1]. В [2] предложена многоклассовая модель, в которой класс имеет смысл некоторой части объекта. Многоклассовая модель обладает большей выразительной способностью по сравнению с бинарной, однако для ее обучения требуются данные с полной разметкой.

В данной работе мы предлагаем метод обучения MSBM, использующий лишь бинарные маски объектов и семена их частей (точки на изображении, в окрестностях которых лежат пиксели заданного класса). Для обучения модели с такими входными данными нами предложена EM-подобная процеду-

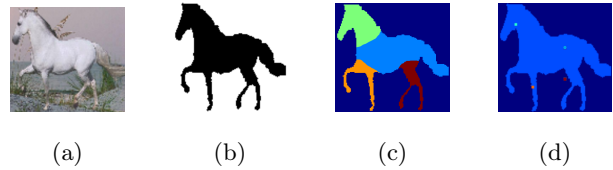


Рис. 1: (a) – изображение, (b) – бинарная разметка \mathbf{b} , (c) – многоклассовая разметка \mathbf{m} , (d) семена \mathbf{s} для 4 частей: головы, передних ноги, задних ноги, крупа.

ра. Семена частей могут размечаться как вручную, так и автоматически с помощью детектора частей объекта (part-based detector, [8]). Использование детектора частей объектов позволяет обучать MSBM на основе лишь исходных изображений и их бинарных разметок.

В разделе экспериментов мы показали, что MSBM, обученная нашим алгоритмом, превосходит SBM и сопоставима с MSBM, обученной стандартным способом. Также показано, что модель, обученная нашим алгоритмом, лучше MSBM, для которых многоклассовая разметка была получена из бинарной маски и семян на основе простых эвристических предположений.

2 Ноtация

Пусть нам дано D центрированных и масштабированных изображений с объектами одного типа (рис. 1а, тип: лошади в профиль). За P обозначим число частей объекта, а N – число пикселей на изображении. Пусть $B = \{\mathbf{b}^d\}_{d=1, \dots, D}$ – множество бинарных разметок изображений (рис. 1b), где $\mathbf{b}^d \in \{0, 1\}^N$, а $M = \{\mathbf{m}^d\}_{d=1, \dots, D}$ – множество многоклассовых разметок (рис. 1c), где $\mathbf{m}^d \in \{0, \dots, P\}^N$, 0 соответствует фону. Также будем использовать бинарные переменные $m_{ip}^d \in \{0, 1\}$ для обозначения принадлежности пикселя i на изображении d классу p и векторные переменные $\mathbf{m}_p^d = \{m_{ip}^d\}_{i=1, \dots, N}$. Введем функцию $f_{coord}(i)$, ставящую в соответствие номеру пикселя его координаты на изображении и множество семян $S = \{\mathbf{s}^d\}_{d=1, \dots, D}$, где \mathbf{s}^d – это упорядоченный набор семян для изображения d , каждое семя задается своими координатами на изображении.

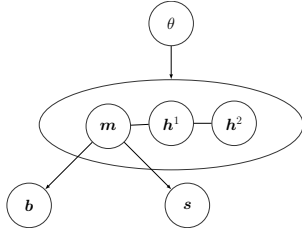


Рис. 2: Графическая модель $p(\mathbf{b}, \mathbf{s}, \mathbf{m}, \mathbf{H} | \boldsymbol{\theta})$.

3 Модели формы Больцмана

SBM — модель бинарной формы, представляющая собой трехслойную глубинную модель Больцмана (DBM, [9]) с дополнительными ограничениями. Наблюдаемый слой, соответствующий изображению, делится на 4 части с перекрытием и каждая часть соединяется только со своим подмножеством переменных на первом скрытом слое. Более того, веса для каждой из этих частей одинаковы. Такие ограничения позволяют существенно уменьшить число параметров модели, что позволяет избежать переобучения и ускорить процесс обучения. Будем считать, что эти ограничения неявно учтены в соответствующей матрице связей.

SBM с наблюдаемым слоем \mathbf{b} и скрытыми $\mathbf{H} = \{\mathbf{h}^1, \mathbf{h}^2\}$ задается с помощью распределения Гиббса:

$$p(\mathbf{b}, \mathbf{H} | \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp(-E(\mathbf{b}, \mathbf{H} | \boldsymbol{\theta})), \quad (1)$$

где $Z(\boldsymbol{\theta}) = \sum_{\mathbf{b}, \mathbf{H}} \exp(-E(\mathbf{b}, \mathbf{H} | \boldsymbol{\theta}))$ — нормировочная константа, $\boldsymbol{\theta} = \{\mathbf{a}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{W}^1, \mathbf{W}^2\}$ — набор параметров модели и энергия:

$$E(\mathbf{b}, \mathbf{H} | \boldsymbol{\theta}) = \mathbf{b}^T \mathbf{W}^1 \mathbf{h}^1 + \mathbf{h}^{1T} \mathbf{W}^2 \mathbf{h}^2 + \mathbf{h}^{1T} \mathbf{c}^1 + \mathbf{b}^T \mathbf{a} + \mathbf{h}^{2T} \mathbf{c}^2. \quad (2)$$

MSBM является обобщением SBM на многоклассовый случай. Она задается аналогичным распределением со следующей энергией:

$$E(\mathbf{m}, \mathbf{H} | \boldsymbol{\theta}) = \sum_{p=1}^P \mathbf{m}_p^T \mathbf{W}_p^1 \mathbf{h}^1 + \mathbf{h}^{1T} \mathbf{W}^2 \mathbf{h}^2 + \mathbf{h}^{1T} \mathbf{c}^1 + \sum_{p=1}^P \mathbf{m}_p^T \mathbf{a}_p + \mathbf{h}^{2T} \mathbf{c}^2 \quad (3)$$

Описание классического алгоритма обучения моделей Больцмана приведено в [10].

4 Совместная модель

Введем совместное распределение на переменные $\mathbf{b}, \mathbf{m}, \mathbf{s}, \mathbf{H}$ предполагая, что \mathbf{b} и \mathbf{s} условно независимы при известном \mathbf{m} :

$$p(\mathbf{b}, \mathbf{s}, \mathbf{m}, \mathbf{H} | \boldsymbol{\theta}) = p(\mathbf{b} | \mathbf{m}) p(\mathbf{s} | \mathbf{m}) p(\mathbf{m}, \mathbf{H} | \boldsymbol{\theta}). \quad (4)$$

Соответствующая графическая модель представлена на рис. 2. Под распределением $p(\mathbf{m}, \mathbf{H} | \boldsymbol{\theta})$ здесь понимается распределение Гиббса для MSBM, $p(\mathbf{b} | \mathbf{m})$ задается как:

$$p(\mathbf{b} | \mathbf{m}) = \prod_{i=1}^N p(b_i | m_i) = \prod_{i=1}^N [b_i = [m_i \neq 0]], \quad (5)$$

а вид $p(\mathbf{s} | \mathbf{m})$ основан на предположении, что пиксель i «тянет на себя» семя m_i :

$$p(\mathbf{s} | \mathbf{m}) \propto \prod_{i: m_i \neq 0} \mathcal{N}(s_{m_i} | f_{coord}(i), \sigma^2) \quad (6)$$

где σ — внешний параметр модели.

5 EM-алгоритм

Для обучения параметров $\boldsymbol{\theta}$ будем с помощью EM-алгоритма решать следующую задачу максимизации:

$$\log P(B, S | \boldsymbol{\theta}) = \sum_{d=1}^D \log p(\mathbf{b}^d, \mathbf{s}^d | \boldsymbol{\theta}) \rightarrow \max_{\boldsymbol{\theta}}. \quad (7)$$

При этом будем считать переменные \mathbf{b}, \mathbf{s} наблюдаемыми, а \mathbf{m}, \mathbf{H} — скрытыми.

На E-шаге будем находить вариационную оценку апостериорного распределения на скрытые переменные в семействе полностью факторизованных распределений $q^d(\mathbf{m}, \mathbf{H}) = \prod_{i=1}^I q_i^d(m_i) \prod_{j=1}^J q_j^d(h_j^1) \prod_{k=1}^K q_k^d(h_k^2)$ минимизируя дивергенцию Кульбака-Лейблера:

$$\min_{q^d} \text{KL}(q^d(\mathbf{m}, \mathbf{H}) \| p(\mathbf{m}, \mathbf{H} | \mathbf{b}^d, \mathbf{s}^d, \boldsymbol{\theta})). \quad (8)$$

Решение данной задачи сновано на [11, р. 466]. Оно представляет собой итеративную процедуру со следующими формулами пересчета:

$$q_i^d(m_i = p) \propto [b_i^d \cdot m_i \neq 0] \exp\left(-\frac{\|s_{m_i}^d - f_{coord}(i)\|_2^2}{2\sigma^2} + a_{i, m_i} + \sum_j W_{i, j, m_i}^1 q_j(h_j^1 = 1)\right) [b_i^d = 0] [m_i^d = 0] \quad (9)$$

$$q_j^d(h_j^1 = 1) = \sigma\left(c_j^1 + \sum_{i, p=1}^I q_i(m_i = p) W_{i, j, p}^1 + \sum_k W_{j, k}^2 q_k(h_k^2 = 1)\right), \quad (10)$$

$$q_k^d(h_k^2 = 1) = \sigma\left(c_k^2 + \sum_j q_j(h_j^1 = 1) W_{j, k}^2\right), \quad (11)$$

где $\sigma(x) = 1/(1 + e^{-x})$. Данная процедура также обобщается на случай, когда часть семян неизвестна.

На M-шаге обновляем параметры модели $\boldsymbol{\theta}$ решая следующую задачу:

$$\max_{\boldsymbol{\theta}} \sum_{d=1}^D \left[\sum_{\mathbf{m}, \mathbf{H}} q^d(\mathbf{m}, \mathbf{H}) \log p(\mathbf{b}^d, \mathbf{s}^d, \mathbf{m}, \mathbf{H} | \boldsymbol{\theta}) \right]$$

Учитывая полную факторизованность распределения $q(\mathbf{m}, \mathbf{H})$ получаем следующую задачу максимизации:

$$\max_{\boldsymbol{\theta}} \sum_{d=1}^D \left(\sum_{i=1}^I \sum_{p=1}^P a_{i, p} q_i(m_i^d = p) + \sum_{j=1}^J c_j^1 q_j(h_j^1 = 1) + \sum_{i=1}^I \sum_{j=1}^J \sum_{p=1}^P q_i(m_i^d = p) W_{i, j, p}^1 q_j(h_j^1 = 1) + \sum_{j=1}^J \sum_{k=1}^K q_j(h_j^1 = 1) W_{j, k}^2 q_k(h_k^2 = 1) + \sum_{k=1}^K c_k^2 q_k(h_k^2 = 1) - \log Z(\boldsymbol{\theta}) \right) \quad (12)$$

Данную максимизационную проблему можно решить используя стохастический градиентного метод.

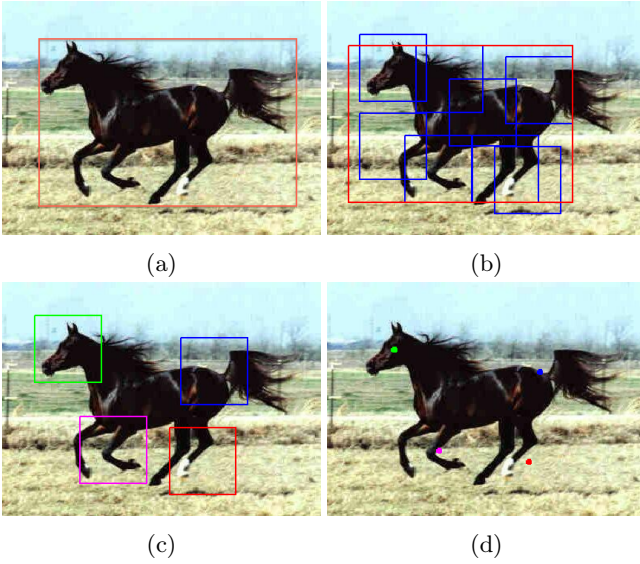


Рис. 3: (а) изображение с выделенным ограничивающим прямоугольником, (б) результат применения детектора, (в) выбранные части, (д) семена для выбранных частей.

6 Использование детектора частей

Семена на изображениях можно получать как посредством ручной разметки, так и автоматически с помощью детектора частей [8]. Для обучения такого детектора требуются изображения с размеченными ограничивающими прямоугольниками для объектов (рис. 3, а). Детектор сам выделяет части на изображении, имеющие сходную структуру на обучающей выборке. На этапе применения к тестовому изображению детектор выделяет ограничивающие прямоугольники для объекта и его частей (рис. 3, б). Далее выбираются только некоторые из частей (рис. 3, в), центры которых и принимаются за семена (рис. 3, г).

7 Эксперименты

Сравним MSBM, обученную предложенным алгоритмом, с SBM и с MSBM, обученными стандартным образом. Сравнение будем проводить на двух наборах данных: лошади ([3], 328 изображений) и мотоциклы ([12], 798 изображений). Предварительно обучим детектор и получим семена для этих выборок, масштабируем (для лошадей: 32 на 32 пикселя, для мотоциклов: 40 на 40) и центрируем бинарные маски объектов и разделим выборки на обучение и контроль (для лошадей: 278 и 50 изображений, для мотоциклов: 600 и 198). Для стандартного обучения MSBM выполним многоклассовую ручную разметку на наборе данных с лошадьми. Также получим многоклассовую разметку данных из бинарных масок и семян на основе двух эвристик: пиксель объекта принадлежит той части, к которой он ближе всего с точки зрения евклидова расстояния (Euc1), и с точки зрения длины кратчайшего пути, проходящего внут-

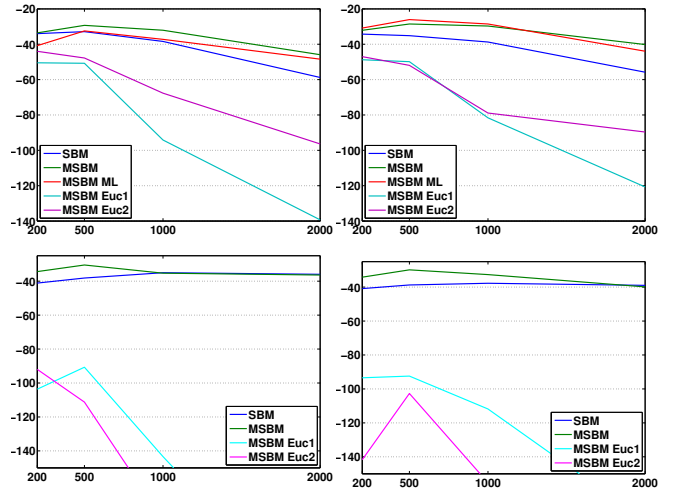


Рис. 4: Качество восстановления формы. Сверху – данные с лошадьми, снизу – с мотоциклами. Слева – $|h^2| = 100$, справа – $|h^2| = 200$. По оси абсцисс отложено $|h^1|$. Наша модель – зеленый график. Чем выше значение значения меры, тем лучше.

ри объекта (Euc2). Для сравнения моделей будем использовать две меры качества описанные ниже.

7.1 Восстановление формы

Первая мера оценивает качество восстановления части изображения по всему остальному изображению. Рассмотрим бинарную маску объекта с известными семенами. Разделим все пиксели этой маски на два класса случайных переменных: наблюдаемые b^o и скрытые b^u , семена при этом также делятся на два соответствующих класса. В случае с SBM можно оценить вероятность $p(b^u|b^o) \approx \sum_{r=1}^R p(b^u|H_r)/R$, а в случае MSBM – вероятность $p(b^u|b^o, s^o) \approx \sum_{r=1}^R p(b^u|H_r, s^o)/R$, где H_r , $r = \overline{1, R}$ – результаты R итераций семплирования Гиббса из распределений $p(H|b^o)$ и $p(H|b^o, s^o)$ соответственно.

Будем вычислять данные оценки для всех изображений тестовой выборки. Деление пикселей на классы зададим следующим образом: разделим изображение на 9 равных частей сеткой 3 на 3 и будем поочередно брать в качестве скрытых переменных один из полученных сегментов. Итоговая мера качества представляет собой усредненный по разбиениям пикселей и тестовым изображениям логарифм оцененных вероятностей. Результаты приведены на рис. 4.

7.2 Порождение формы из семян

Вторая мера оценивает качество порождения формы из семян. Пусть даны семена s и соответствующие им пиксели b^s : $f_{coord}(b_p^s) = s_p \forall p = \overline{1, P}$ принадлежат объекту. Сгенерируем значения скрытых переменных из распределений $p(H|b^s)$ для SBM и из $p(H|b^s, s)$ для MSBM. Далее сгенерируем форму b или m соответственно. Повторим эти две операции 100 раз и получим итоговые формы объекта. Приме-

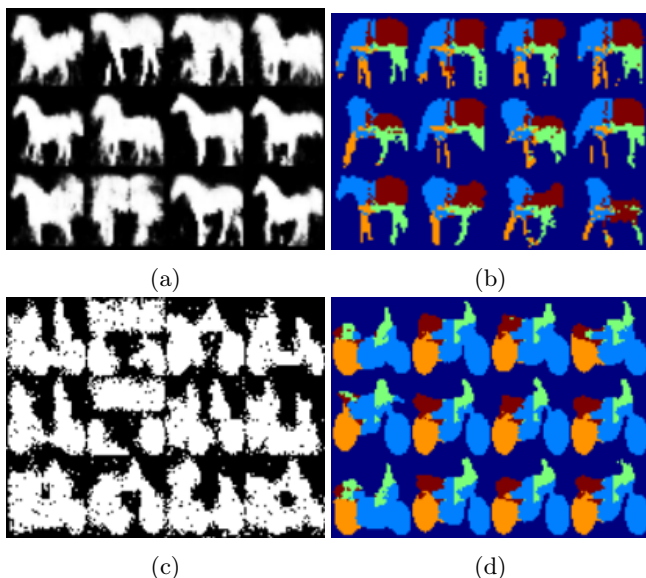


Рис. 5: Порождение формы из семян. Слева – SBM, справа – MSBM, обученная нашим алгоритмом.

ры получаемых форм представлены на рис. 5. Итоговая мера качества представляет собой усредненное по тестовой выборке расстояние Хэмминга между получившейся бинарной формой (в многоклассовом случае также бинаризуем результат) и исходной, для которой брались семена. Результаты приведены на рис. 6.

8 Заключение

В данной работе был предложен новый алгоритм обучения MSBM, требующий лишь бинарные маски объектов и семена для частей объектов. Также был рассмотрен метод автоматического получения семян с помощью детектора частей. На этапе экспериментов показано, что MSBM, обученная предлагаемым алгоритмом, превосходит по качеству SBM и сопоставима с MSBM, обученной стандартным способом. Также показано, что предлагаемый метод обучения дает лучшие результаты, чем эвристическое получение полной разметки из бинарной и семян и последующее классическое обучение MSBM.

Список литературы

- [1] Eslami, S.M.A., Heess, N., Williams, C.K.I., Winn, J.: The shape boltzmann machine: a strong model of object shape. In: International Journal of Computer Vision. (2013)
- [2] Eslami, S.M.A., Williams, C.: A generative model for parts-based object segmentation. In Bartlett, P., Pereira, F., Burges, C., Bottou, L., Weinberger, K., eds.: Advances in Neural Information Processing Systems 25. (2012) 100–107
- [3] Borenstein, E.: Combining top-down and bottom-up segmentation. In: In Proceedings IEEE workshop on Perceptual Organization in Computer Vision, CVPR. (2004) 46

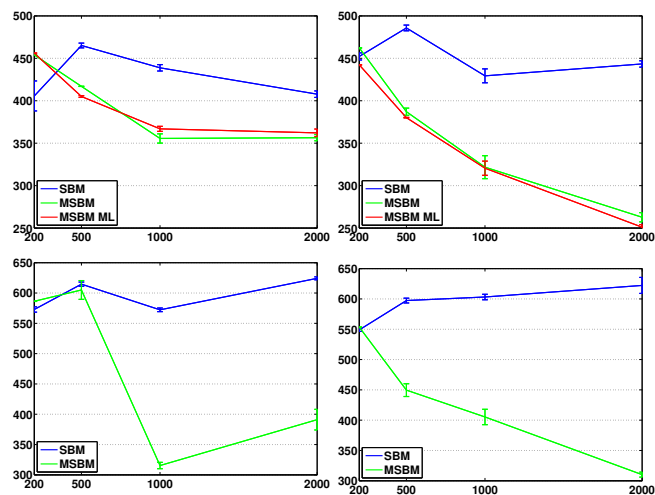


Рис. 6: Порождение формы из семян. Сверху – данные с лошадыми, снизу – с мотоциклами. Слева – $|h^2| = 100$, справа – $|h^2| = 200$. По оси абсцисс отложено $|h^1|$. Наша модель – зеленый график. Чем ниже значение меры, тем лучше.

- [4] Van Ginneken, B., Frangi, A.F., Staal, J.J., ter Haar Romeny, B.M., Viergever, M.A.: Active shape model segmentation with optimal features. IEEE Transactions on Medical Imaging **21**(8) (2002) 924–933
- [5] Chan, T.F., Shen, J.: Non-texture inpainting by curvature-driven diffusions (cdd). J. Visual Comm. Image Rep **12** (2001) 436–449
- [6] Ferrari, V., Jurie, F., Schmid, C.: From images to shape models for object detection. International Journal of Computer Vision **87**(3) (2010) 284–303
- [7] Yangel, B., Vetrov, D.: Image segmentation with a shape prior based on simplified skeleton. In: Energy Minimization Methods in Computer Vision and Pattern Recognition. Volume 6819 of Lecture Notes in Computer Science. (2011) 247–260
- [8] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(9) (2010) 1627–1645
- [9] Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Computation **18**(7) (2006) 1527–1554
- [10] Salakhutdinov, R., Hinton, G.: An efficient learning procedure for deep boltzmann machines. Neural Computation **24**(8) (2012) 1967–2006
- [11] Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)
- [12] Fei-Fei, L., Fergus, R., Perona, P.: One-shot learning of object categories. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(4) (2006) 594–611