Динамическое программирование с построчным комбинированием переменных для обработки изображений

Копылов А. В.
kopylov@uic.tula.ru
Тула, Тульский технический университет

Множество задач анализа изображений, таких как сглаживание, локальный текстурный анализ, сегментация, совмещение изображений в стереовидении или распознавании образов, и т. д., могут быть представлены как задачи преобразования исходного изображения $Y = (y_{\mathbf{t}},\ \mathbf{t} \in \mathbf{T})$, которое обычно принимает значения на непрерывной или дискретной оси уровня яркости $y_{\mathbf{t}} \in \mathbf{Y}$ и определено на подмножестве $\mathbf{T} = \big\{\mathbf{t} = (t_1, t_2)\colon t_1 = 1, \ldots, N_1,\ t_2 = 1, \ldots, N_2\big\}$ двумерного пространства (плоскости изображения), во вторичный массив $X = (x_{\mathbf{t}}, \mathbf{t} \in \mathbf{T})$. Данный массив определен на том же множестве аргументов $\mathbf{t} \in \mathbf{T}$ и принимает значения $x \in \mathbf{X}$ из множества, специфичного для каждой задачи.

Основная идея оптимизационного подхода к анализу изображений состоит в том, что алгоритм анализа данных строится как алгоритм оптимизации, условно минимизации, подходящей целевой функции, определенной на множестве всех возможных вариантов вторичного массива данных. Такая целевая функция практически всегда может быть выбрана в так называемой парно-сепарабельной форме как сумма элементарных функций двух видов, а именно узловых функций и функций связи.

$$J\,(X\,|Y\,) = \sum_{\mathbf{t}\in\mathbf{T}} \psi_{\mathbf{t}}(x_{\mathbf{t}}|Y_{\mathbf{t}}) + \sum_{(\mathbf{t}',\mathbf{t}'')\in G} \gamma_{\mathbf{t}',\mathbf{t}''}(x_{\mathbf{t}'}, x_{\mathbf{t}''}). \tag{1}$$

Узловые функции $\psi_{\mathbf{t}}(x_{\mathbf{t}}|Y_{\mathbf{t}})$, несущие информацию о данных, играют роль меры несходства между искомым локальным значением $x_{\mathbf{t}}$ и некоторой окрестностью $Y_{\mathbf{t}}$ точки $\mathbf{t}$ обрабатываемого массива. Каждая функция связи $\gamma_{\mathbf{t}',\mathbf{t}''}(x_{\mathbf{t}'}, x_{\mathbf{t}''})$, основанная на модельных представлениях, накладывает штраф на различие значений пары соседних переменных на соответствующем ребре неориентированного графа G, который образован множеством пар $G \subset T \times T$ соседних элементов изображения.

В случае, когда целевые переменные принимают значения из конечного множества $x \in \mathbf{X} = \{1, \ldots, m\}$, данная оптимизационная задача является одним из частных случаев так называемых задач разметки, который известен как (max, +) или (min,+) задачи.

Для произвольного графа соседства G задача оптимизации парно-сепарабельной целевой функции (1) является NP-полной, но в частном случае (min, +) задачи, когда соседство переменных определяется при помощи произвольного ациклического графа, принцип последовательного исключения переменных приводит к оптимизационной процедуре [1], имеющей много общего с принципом динамического программирования Беллмана. Этот принцип состоит в разложении исходной задачи на последовательность подзадач с уменьшающимся числом переменных, вплоть до единственной переменной на последнем шаге.

В том случае, когда граф соседства имеет вид решетки, принцип динамического программирования Беллмана не может быть непосредственно применен, тем не менее представляется очень заманчивым сохранить вычислительные преимущества процедуры динамического программирования, даже ценой некоторых эвристических компромиссов.

Прежде всего мы можем провести разбиение исходного графа на совокупность поддеревьев. Затем, следуя итерационному методу Гаусса-Зайделя, можно оптимизировать целевую функцию путем поиска на каждом шаге глобального минимума частичных целевых функций, связанных с поддеревьями, при помощи процедуры древовидного динамического программирования [1]. Для того, чтобы достичь более глубокого минимума, можно изменять разбиение графа на каждом шаге процедуры. Возможно также построение неитерационного приближенного алгоритма на основе удаления некоторых ребер графа соседства [2], [3].

Тем не менее, существует класс задач разметки, для которых граф смежности переменных не удается разложить на поддеревья, для которых частичная оптимальная разметка совпадает с глобально оптимальной. Такие задачи не могут быть решены подобными методами.

В данной работе рассматривается совершенно другой подход к задаче парно-сепарабельной оптимизации с графом смежности переменных в виде решетки. Основная идея данного подхода состоит в представлении исходной целевой функции с решетчатой смежностью переменных (1) как функции, переменные которой представляют уже не отдельные узлы решетки, а целые строки этих узлов. При этом строки решетки естественным образом выстраиваются в вертикальную цепочку, образуя тем самым цепочечный граф смежности для нового типа агрегированных переменных, каждая из которых представляет собой всю совокупность элементарных переменных в соответствующей строке.

Оптимизация функции с цепочечным графом смежности переменных формально является классической задачей динамического программирования, и может быть выполнена за число шагов, равное числу строк решетки N_1. Однако, являясь линейной относительно числа строк, вычислительная сложность процедуры оптимизации оказывается экспоненциальной относительно числа целевых переменных. Для преодоления этой трудности в данной работе рассматривается приближенная процедура оптимизации парно-сепарабельных целевых функций с решетчатым графом смежности, основанная на эвристической замене каждой функции Беллмана относительно группы переменных в отдельной строке подходящей парно-сепарабельной функцией, что позволяет снизить вычислительную сложность процедуры до линейной относительно числа элементов массива $N_1 N_2$.

Для экспериментальной проверки точности разработанных алгоритмов были использованы некоторые классы (min, +) задач разметки, для которых известно точное решение [4]. Первый класс задач представлен субмодулярными задачами, возникающими при построении трехмерной модели человеческого лица. Решение этих задач получено при помощи алгоритма минимального сечения графа [5]. Второй класс составляют задачи сегментации, для которых существуют эквивалентные тривиальные задачи, и которые решены при помощи алгоритма линейной релаксации [6]. Третий класс задач составляют модельные задачи, полученные из тривиальных путем случайных эквивалентных преобразований [6].

Для сравнения применялись алгоритмы на основе древовидной декомпозиции и принципа Гаусса-Зайделя [3], который использует решение, полученное алгоритмом с аппроксимацией решетчатого графа соседства последовательностью деревьев [1], в качестве начального приближения. Эксперименты показали, что алгоритмы с ациклической оптимизацией имеют достаточную точность для субмодулярных задач, так же как и для задач остальных двух классов.

Работа выполнена при поддержке РФФИ, проекты №06-01-08042 и №06-01-00412.

Литература

[1] *Mottl V., Blinov A., Kopylov A., Kostin A.* Optimization techniques on pixel neighborhood graphs for image processing // Graph-Based Representations in Pattern Recognition. — Springer–Verlag/Wien, 1998. — Pp. 135–145.

[2] *Veksler. O.* Stereo Correspondence by Dynamic Programming on a Tree // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). — 2005. — Vol. 2. — Pp. 384–390.

[3] *A. Kopylov.* Acyclic pair-wise separable optimization for image processing // Proceedings of the Ninth International Conference on Pattern Recognition and Information Processing. Vol 1, 2007. — Pp. 203–207.

[4] *Schlesinger M. I., Flach B.* Some solvable subclasses of structural recognition problems // Proc. of Czech Patt. Recogn. Workshop, Praha, 2000. — Pp. 55–62.

[5] *Boykov Yu., Veksler O., Zabih R.* Fast approximate energy minimization via graph cuts // PAMI, 23(11), November 2001. — Pp. 1222–1239.

[6] *Schlesinger M. I.* Syntactic analysis of two-dimensional visual signals in noisy conditions // Kibernetika, 4, 1976. — Pp. 113–130.