Алгоритмы минимизации энергии на основе разрезов графов

Потоки в сетях и разрезы графов

Рассмотрим ориентированный граф $\bar{\mathcal{G}} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$, где $\bar{\mathcal{V}}$ – множество вершин, $\bar{\mathcal{E}}$ – множество ребер. Каждому ориентированному ребру (дуге) $(i,j) \in \bar{\mathcal{E}}$ соответствует неотрицательное число $c(i,j)$ – ёмкость (capacity). Пусть в графе есть две выделенные вершины: s – *исток*, t – *сток*. Граф $\bar{\mathcal{G}}$ также называют *транспортной сетью*.

Потоком в сети назовем функцию $f : \bar{\mathcal{E}} \to \mathbb{R}$, такую что

1. $\forall (i,j) \in \bar{\mathcal{E}}, \quad f(i,j) \le c(i,j)$;
2. $\forall i \in \bar{\mathcal{V}} \setminus \{s,t\}, \quad \sum\limits_{j:(i,j)\in\bar{\mathcal{E}}} f(i,j) - \sum\limits_{j:(j,i)\in\bar{\mathcal{E}}} f(j,i) = 0.$

Величиной потока f назовем число $\sum\limits_{i:(s,i)\in\bar{\mathcal{E}}} f(s,i) = \sum\limits_{i:(i,t)\in\bar{\mathcal{E}}} f(i,t)$.

Задача о максимальном потоке в графе состоит в поиске потока f, обладающего максимальной величиной.

st-разрезом графа называется разбиение множества $\bar{\mathcal{V}}$ на два множества $\mathcal{S}$ и $\mathcal{T}$ ($\mathcal{S} \cup \mathcal{T} = \bar{\mathcal{V}}$, $\mathcal{S} \cap \mathcal{T} = \varnothing$), такое что $s \in \mathcal{S}$, $t \in \mathcal{T}$.

Величиной st-разреза называется сумма ёмкостей всех ребер, ведущих из $\mathcal{S}$ в $\mathcal{T}$: $\sum\limits_{\substack{(i,j)\in\bar{\mathcal{E}} \\ j\in\mathcal{S},\, j\in\mathcal{T}}} c(i,j)$.

Задача о минимальном st-разрезе в графе состоит в поиске разреза $(\mathcal{S}, \mathcal{T})$, обладающего минимальной величиной.

Теорема Форда-Фалкерсона гласит, что величина максимального потока равна величине минимального st-разреза.

Известно, что задачи о максимальном потоке и о минимальном st-разрезе можно сформулировать как двойственные задачи линейного программирования. В таких формулировках теорема Форда-Фалкерсона передоказывает сильную двойственность этих двух задач линейного программирования.

Отметим, что если нам известен максимальный поток, то найти минимальный разрез очень легко: например, можно к области истока $\mathcal{S}$ отнести все вершины, до которых существует путь по ненасыщенным ребрам сети ($f(i,j) < c(i,j)$). Построить максимальный поток, зная минимальный разрез – сложная задача.

Для решения задачи о максимальном потоке существует большое число алгоритмов. Классические алгоритмы Форда-Фалкерсона, проталкивания предпотока (push-relabel) описаны в Википедии и в книге “Алгоритмы: построение и анализ” (Т. Кормен и др.) [1]. Существует несколько специализированных алгоритмов, наиболее эффективных для задач, возникающих в

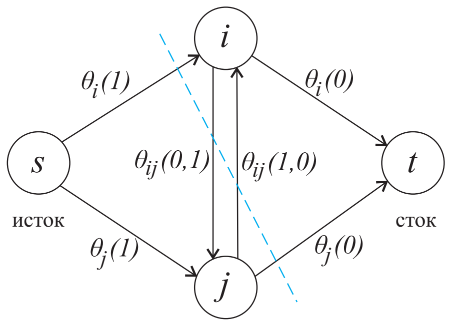

Рис. 1: Граф, построенный для минимизации энергии от двух переменных x_i, x_j. Разрез, отображенный пунктирной линией соответствует присваиванию $x_i = 1$, $x_j = 0$. Величина разреза составляет $\theta_i(1) + \theta_j(0) + \theta_{ij}(1, 0)$.

компьютерном зрении: Бойкова-Колмогорова [2][1], IBFS [3][2].

Алгоритм IBFS в худшем случае выполняет $O(|\mathcal{V}|^2|\mathcal{E}|)$ операций (так же, как алгоритм проталкивания предпотока). На практике алгоритмы Бойкова-Колмогорова и IBFS для графов, возникающих в задачах зрения, работают практически за время, линейной по числу вершин.

Задача минимизации парно-сепарабельной энергии от бинарных переменных

Рассмотрим неориентированный граф $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Каждой вершине $i \in \mathcal{V}$ поставим в соответствие бинарную переменную $x_i \in \{0, 1\}$. Рассмотрим следующую *энергию* E (функцию $E : \{0, 1\}^{|\mathcal{V}|} \to \mathbb{R}$):

$$E(\boldsymbol{x}) = \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{\{i,j\} \in \mathcal{E}} \theta_{ij}(x_i, x_j) + \theta_0. \tag{1}$$

Здесь $\theta_i(x_i)$ – *унарные потенциалы*, $\theta_{ij}(x_i, x_j)$ – *парные потенциалы*, θ_0 – константа. Энергия, состоящая из унарных, парных, и константного потенциалов называется *парно-сепарабельной*.

Задача минимизации энергии E состоит в поиске значения переменных $\boldsymbol{x} = \{x_i\}_{i \in \mathcal{V}}$, на котором энергия $E(\boldsymbol{x})$ принимает минимальное значение. Отметим, что значение константы θ_0 не влияет на конфигурацию, на которой достигается минимум энергии, а влияет лишь на значение в точке минимума.

Задача минимизации энергии (1) с потенциалами общего вида является NP-трудной.

Для определённости будем всегда считать, что все вершины $\mathcal{V}$ упорядочены каким-либо способом, и в неориентированном ребре $(i, j) \in \mathcal{V}$ всегда $i < j$.

Соответствие минимизации энергии разрезу графа

Рассмотрим энергию (1), в которой потенциалы удовлетворяют следующим ограничениям:

- унарные потенциалы неотрицательны: $\forall i \in \mathcal{V} \quad \theta_i(0) \geq 0,\ \theta_i(1) \geq 0$;
- парные потенциалы неотрицательны и равны 0 при равенстве связанных переменных: $\forall \{i, j\} {\in} \mathcal{E} \quad \theta_{ij}(0, 0) = \theta_{ij}(1, 1) = 0, ; \theta_{ij}(0, 1) \geq 0,\ \theta_{ij}(1, 0) \geq 0$.

[1] http://pub.ist.ac.at/~vnk/software/maxflow-v3.02.src.tar.gz

[2] http://www.cs.tau.ac.il/~sagihed/ibfs/

В этом случае энергию (1) можно записать в следующем виде:

$$E(\boldsymbol{x}) = \sum_{i\in\mathcal{V}} (x_i\theta_i(1) + (1-x_i)\theta_i(0)) + \sum_{\{i,j\}\in\mathcal{E}} (x_i(1-x_j)\theta_{ij}(1,0) + x_j(1-x_i)\theta_{ij}(0,1)) + \theta_0. \quad (2)$$

По энергии (2) построим ориентированный граф $\bar{\mathcal{G}} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$, разрез которого будет соответствовать присвоению значений переменным $\boldsymbol{x}$:

- Множество вершин $\bar{\mathcal{V}} = \mathcal{V} \cup \{s,t\}$, где s и t – две дополнительные вершины.
- Множество дуг $\bar{\mathcal{E}}$ строится следующим образом: каждому ребру $\{i,j\} \in \mathcal{E}$ ставим в соответствие две дуги (i,j) и (j,i) (*нетерминальные дуги*), каждой вершине $i \in \mathcal{V}$ ставим в соответствие дуги (s,i) и (i,t) (*терминальные дуги*).
- Будем считать, что множество истока $\mathcal{S}$ соответствует значению переменных 0, множество стока $\mathcal{T}$ – 1.
- Веса терминальных дуг: $c(s,i) = \theta_i(1)$, $c(i,t) = \theta_i(0)$, где $i \in \mathcal{V}$.
- Веса нетерминальных дуг: $c(i,j) = \theta_{ij}(0,1)$, $c(j,i) = \theta_{ij}(1,0)$, где $(i,j) \in \mathcal{E}$, и, по договоренности, $i < j$.
- Любой st-разрез такого графа соответствует разбиению множества $\bar{\mathcal{V}}$ на множества истока $\mathcal{S}$ и стока $\mathcal{T}$. Присваивание значений переменных строится следующим образом: $i \in \mathcal{S} \Rightarrow x_i = 0$, $i \in \mathcal{T} \Rightarrow x_i = 1$. Величина разреза совпадает со значением энергии (2) при таких значениях переменных $\boldsymbol{x}$. Таким образом, задача минимизации энергии вида (2) эквивалента задача поиска минимального разреза в графе $\mathcal{G}$.

Пример графа, построенного для энергии от 2-х переменных, и его разреза приведен на рис. 1.

Репараметризация

В данном разделе рассматривается вопрос о том, какие еще энергии можно минимизировать при помощи разрезов графов (кроме (2)).

На энергию $E(\boldsymbol{x})$ можно смотреть как на функцию. У одной и той же функции может быть несколько различных записей вида (1). Назовем преобразование записи энергии $E(\boldsymbol{x})$, не меняющее энергию, как функцию, *репараметризацией*. Рассмотрим несколько видов репараметризаций:

- Вычитание константы из унарного потенциала: $\theta_i(0) := \theta_i(0) - \delta$, $\theta_i(1) := \theta_i(1) - \delta$, $\theta_0 := \theta_0 + \delta$. Здесь $i \in \mathcal{V}$, $\delta \in \mathbb{R}$.
- Изменение парных потенциалов: $\theta_{ij}(p,0) := \theta_{ij}(p,0) - \delta$, $\theta_{ij}(p,1) := \theta_{ij}(p,1) - \delta$, $\theta_i(p) := \theta_i(p) + \delta$. Аналогично $\theta_{ij}(0,p) := \theta_{ij}(0,p) - \delta$, $\theta_{ij}(1,p) := \theta_{ij}(1,p) - \delta$, $\theta_j(p) := \theta_j(p) + \delta$. Здесь $\{i,j\} \in \mathcal{E}$, $\delta \in \mathbb{R}$, $p \in \{0,1\}$.

Легко показать, что описанные преобразования не меняют энергию $E(\boldsymbol{x})$ как функцию.

Рассмотрим какие парно-сепарабельные энергии общего вида (1) при помощи описанных преобразований-репараметризаций можно свести к виду (2). Заметим, что применяя репараметризацию унарных потенциалов с $\delta = \min(\theta_i(0), \theta_i(1))$ унарные потенциалы всегда можно сделать неотрицательными. Таким образом, можно снять все ограничения на унарные потенциалы энергии (2).

Рассмотрим, что можно сделать при помощи репараметризаций парных потенциалов. Пусть потенциал ребра $\{i,j\} \in \mathcal{E}$ принимает следующие значения: $\theta_{ij}(0,0) = a$, $\theta_{ij}(1,1) = b$, $\theta_{ij}(0,1) = c$, $\theta_{ij}(1,0) = d$. Применим следующую последовательность преобразований:

1. $\theta_{ij}(0,0) := \theta_{ij}(0,0) - a$, $\theta_{ij}(0,1) := \theta_{ij}(0,1) - a$, $\theta_i(0) := \theta_i(0) + a$;
2. $\theta_{ij}(0,1) := \theta_{ij}(0,1) - c + a$, $\theta_{ij}(1,1) := \theta_{ij}(1,1) - c + a$, $\theta_j(1) := \theta_j(1) + c - a$;

3. $\theta_{ij}(1,1) := \theta_{ij}(1,1) - b + c - a$, $\theta_{ij}(1,0) := \theta_{ij}(1,1) - b + c - a$, $\theta_i(1) := \theta_j(1) + b - c + a$;
Можно убедиться, что данные преобразования приводят к следующим значениям парного потенциала $\{i,j\}$: $\theta_{ij}(0,0) = \theta_{ij}(1,1) = \theta_{ij}(0,1) = 0$, $\theta_{ij}(1,0) = d + c - a - b$. Если после этого при помощи описанных ранее действий сделать все унарные потенциалы неотрицательными, то энергия будет иметь вид (2), когда $d + c - a - b \geq 0$. Условие

$$\theta_{ij}(0,0) + \theta_{ij}(1,1) \leq \theta_{ij}(0,1) + \theta_{ij}(1,0) \tag{3}$$

называется условием *субмодулярности* и является необходимым и достаточным условием применимости алгоритмов разрезов графов для минимизации парно-сепарабельной энергии от бинарных переменных (1). Данное условие вызвано условием неотрицательности ёмкостей дуг графа для полиномиальной разрешимости задач о максимальном потоке и минимальном разрезе.

Задача минимизации парно-сепарабельной энергии от K-значных переменных

Рассмотрим неориентированный граф $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Каждой вершине $i \in \mathcal{V}$ поставим в соответствие дискретную переменную $y_i \in \{1, \ldots, K\} = \mathcal{K}$. Рассмотрим следующую *энергию* E_M (функцию $E_M : \mathcal{K}^{|\mathcal{V}|} \to \mathbb{R}$):

$$E_M(\boldsymbol{y}) = \sum_{i \in \mathcal{V}} \psi_i(y_i) + \sum_{\{i,j\} \in \mathcal{E}} \psi_{ij}(y_i, y_j) + \psi_0. \tag{4}$$

Здесь $\psi_i(y_i)$ – *унарные потенциалы*, $\psi_{ij}(y_i, y_j)$ – *парные потенциалы*, ψ_0 – константа. Энергия, состоящая из унарных, парных, и константного потенциалов называется *парно-сепарабельной*.

Задача минимизации энергии E_M состоит в поиске значения переменных $\boldsymbol{y} = \{y_i\}_{i \in \mathcal{V}}$, на котором энергия $E_M(\boldsymbol{y})$ принимает минимальное значение.

Задача минимизации энергии (4) является NP-трудной даже при $K = 3$ и парных потенциалах Поттса: $\psi_{ij}(y_i, y_j) = c_{ij}[y_i \neq y_j], c_{ij} \geq 0$ [4].

В работе [4] предложены приближённые алгоритмы минимизации энергии (4): α-расширение (α-expansion) и $\alpha\beta$-замена ($\alpha\beta$-swap).

Алгоритм α-расширение

Алгоритм α-расширение минимизирует энергию (4) при помощи выполнения шагов между разметками $\boldsymbol{y}$, каждый из которых гарантированно не увеличивает значение энергии. Каждый шаг представляет собой задачу минимизации энергии бинарных переменных вида (1). Неформально каждый шаг позволяет каждой переменной из $\boldsymbol{y}$ либо присвоить выбранное значение α, либо оставить текущее значение (расширение метки α). Пример таких шагов на задаче стерео Tsukuba приведен на рис. (2).

Приведем формальное описание каждого шага алгоритма α-расширение. Пусть есть текущая разметка $\boldsymbol{y}^0$ и выбрана “расширяемая” метка $\alpha \in \mathcal{K}$. Обозначим новую K-значную разметку $\boldsymbol{y}^*$. Будем строить энергию бинарных переменных $E(\boldsymbol{x})$ (1) по энергии $E_M(\boldsymbol{y})$ (4).

- Граф энергий $E(\boldsymbol{x})$ и $E(\boldsymbol{y})$ одинаков: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
- Значение бинарной переменной $x_i = 0$ соответствует $y_i^* = y_i^0$ (значение переменной y_i не меняется); $x_i = 1$ соответствует $y_i^* = \alpha$.

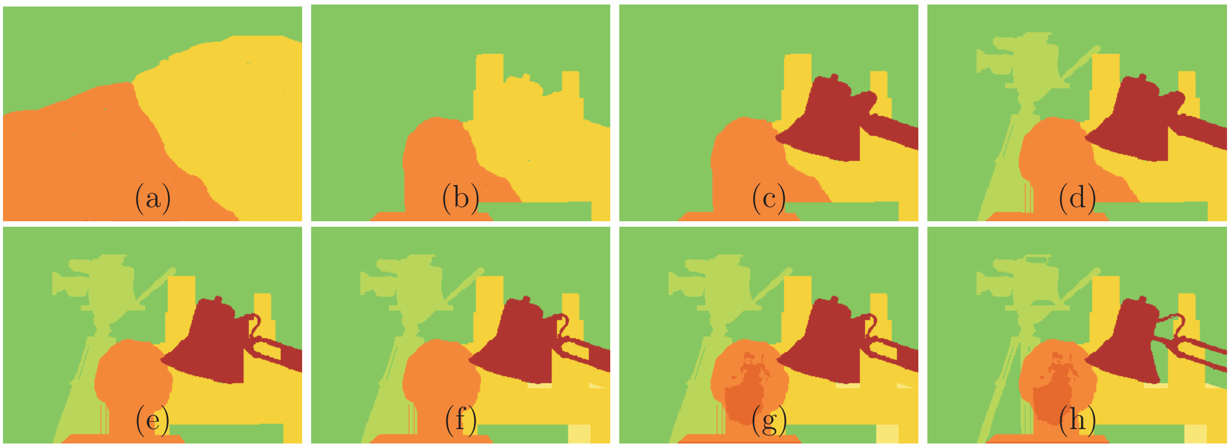

Рис. 2: Пример работы алгоритма α-расширения для задачи выровненного стерео. (a) – начальная разметка; (b)-(h) – последовательные расширения различных меток.

- Унарные потенциалы: $\theta_i(0) = \psi_i(y_i^0)$, $\theta_i(1) = \psi_i(\alpha)$.
- Парные потенциалы: $\theta_{ij}(0,0) = \psi_{ij}(y_i^0, y_j^0)$, $\theta_{ij}(1,1) = \psi_{ij}(\alpha,\alpha)$, $\theta_{ij}(0,1) = \psi_{ij}(y_i^0,\alpha)$, $\theta_{ij}(1,0) = \psi_{ij}(\alpha, y_j^0)$.

Условием применимости алгоритма разреза графа является условие субмодулярности парных потенциалов (3). В терминах K-значных переменных условие выглядит следующим образом:

$$\psi_{ij}(\alpha,\alpha) + \psi_{ij}(\beta,\gamma) \leq \psi_{ij}(\beta,\alpha) + \psi_{ij}(\alpha,\gamma), \tag{5}$$

где $\alpha, \beta, \gamma \in \mathcal{K}$. Часто предполагают, что $\psi_{ij}(\alpha,\alpha) = 0$ и $\psi_{ij}(\beta,\gamma) = \psi_{ij}(\gamma,\beta)$. В этом случае условие (5) становится неравенством треугольника, а все парные потенциалы ψ_{ij} – метриками.

Эффективная реализация алгоритмов α-расширение и $\alpha\beta$-замена доступна на странице авторов[3].

Условие (5) некорректно называть субмодулярностью. Определение субмодулярности для К-значных энергий из работы [5]:

$$\psi_{ij}(\min(y_i^1, y_i^2), \min(y_j^1, y_j^2)) + \psi_{ij}(\max(y_i^1, y_i^2), \max(y_j^1, y_j^2)) \leq \psi_{ij}(y_i^1, y_j^1) + \psi_{ij}(y_i^2, y_j^2),$$

где $y_i^1, y_i^2, y_j^1, y_j^2$ – произвольные метки. Согласно этому определению потенциалы Поттса при $K \geq 3$ не являются субмодулярными. Энергия с субмодулярными потенциалами может быть точно минимизирована за полиномиальное время при помощи алгоритмов разрезов графов (см. [5]).

Алгоритмы разрезов графов для минимизации энергий с потенциалами высоких порядков

Рассмотрим гиперграф $\mathcal{G} = (\mathcal{V}, \mathcal{C})$. Здесь $\mathcal{C} \subseteq 2^{\mathcal{V}}$ – множество гиперребер (подмножеств вершин). Каждой вершине $i \in \mathcal{V}$ поставим в соответствие бинарную переменную $x_i \in \{0,1\}$. Рассмотрим следующую энергию E (функцию $E : \{0,1\}^{|\mathcal{V}|} \to \mathbb{R}$):

$$E(\boldsymbol{x}) = \sum_{i\in\mathcal{V}} \theta_i(x_i) + \sum_{C\in\mathcal{C}} \theta(\boldsymbol{x}_C). \tag{6}$$

[3] http://vision.csd.uwo.ca/code/gco-v3.0.zip

Здесь $\theta_i(x_i)$ – унарные потенциалы, $\theta_C(\boldsymbol{x}_C)$ – потенциалы высоких порядков, $\boldsymbol{x}_C$ – переменные $\boldsymbol{x}$, входящие в множество C.

Задача минимизации энергии вида (6) является очень сложной и может быть эффективно решена лишь в некоторых частных случаях. Наиболее популярным способом минимизации энергий вида (6) является сведение их к парно-сепарабельным энергиям. Часто используется следующее тождество:

$$-\prod_{i\in C} x_i = \min_{z\in\{0,1\}} \left(z(|C|-1) - \sum_{i\in C} z x_i \right).$$

На каждое слагаемое энергии, при котором стоит неположительный коэффициент, вводится дополнительная переменная z так, чтобы относительно переменных $\boldsymbol{x}$ и $\boldsymbol{z}$ функция была парно-сепарабельной и субмодулярной. При этом, минимум энергии по расширенному множеству переменных ($\boldsymbol{x}$ и $\boldsymbol{z}$) совпадает с минимумом энергии по переменным $\boldsymbol{x}$. Подробнее о сведении энергий высоких порядков к парно-сепарабельным можно почитать в работе [6].

Список литературы

[1] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction To Algorithms*, 3rd ed. MIT Press, 2009.

[2] Y. Boykov and V. Kolmogorov, "An experimental comparison of Min-Cut/Max-Flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 26, no. 9, pp. 1124–1137, 2004.

[3] A. V. Goldberg, S. Hed, H. Kaplan, R. E. Tarjan, and R. F. Werneck, "Maximum flows by incremental breadth-first search," in *19th European Symposium on Algorithms (ESA 2011)*, 2011.

[4] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23, no. 11, pp. 1222–1239, 2001.

[5] J. Darbon, "Global optimization for first order Markov random fields with submodular priors," *Discrete Applied Mathematics*, vol. 157, no. 16, pp. 3412–3423, 2009.

[6] H. Ishikawa, "Transformation of general binary MRF minimization to the first order case," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, no. 6, pp. 1234–1249, 2011.