Введение в марковские случайные поля. Часть 2.

Д. П. Ветров[1]

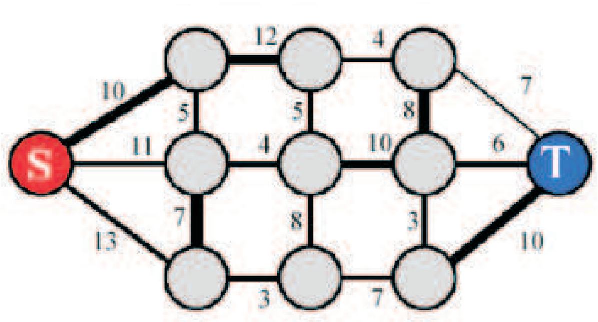
[1]МГУ, ВМиК, каф. ММП

ГрафиКон 2009

План

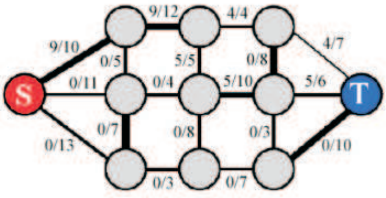
Потоки в сетях

- Рассмотрим неориентированный граф с двумя выделенными вершинами (стоком t и истоком s)
- Пусть с каждым ребром $(u, v) \in E$ ассоциировано некоторое неотрицательное число $c(u, v) \geq 0$ — пропускная способность



Потоки в сетях

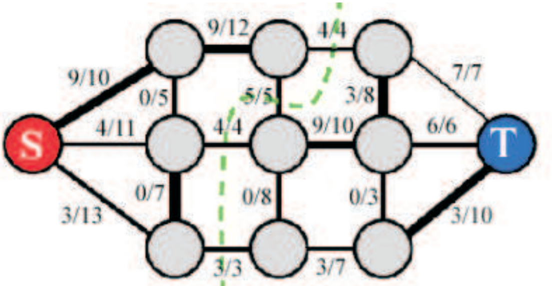
- Назовем потоком неотрицательную функцию $f(u, v)$, определенную на ребрах графа, такую что
$$f(u, v) \leq c(u, v)$$
$$\sum_{v:(u,v)\in E} f(u, v) - \sum_{v:(v,u)\in E} f(v, u) = 0, \quad \forall u \neq \{s, t\}$$
- Первое условие ограничивает поток через ребро его пропускной способностью, а второе гарантирует отсутствие источников и стоков вне выделенной пары вершин
- Задача поиска максимального потока состоит в максимизации величины
$$M(f) = \sum_{v:(s,v)\in E} f(s, v) \to \max_f$$
по всем допустимым потокам



Разрезы графов

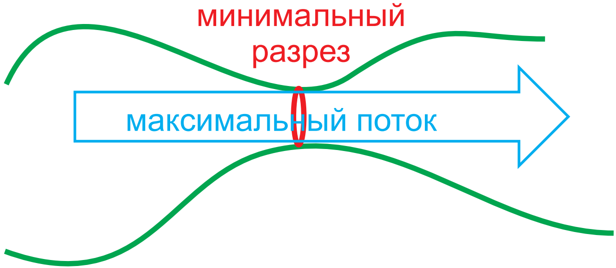
- $(s-t)$-разрезом графа называется разбиение вершин графа на два непересекающихся множества S и T, такие что $s \in S$, $t \in T$
- Величиной разреза называется сумма пропускных способностей всех ребер, один конец которых находится в множестве S, а другой — в множестве T

$$c(S,T) = \sum_{(u,v)\in E, u\in S, v\in T} c(u,v)$$



Теорема Форда-Фалкерсона

- Известная теорема Форда-Фалкерсона гласит, что максимальный поток в сети равен ее минимальному разрезу
- Существует эффективный (полиномиальной сложности) алгоритм решения задачи поиска минимального разреза в графе
- Задачи поиска максимального потока и минимального разреза являются двойственными

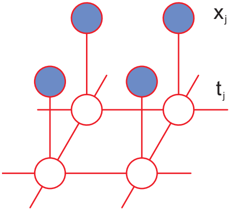


Марковские решетки

- В дальнейшем будем рассматривать т.н. марковские решетки, в которых размер максимальной клики не превосходит двух
- В этом случае совместное распределение переменных марковской сети выражается формулой

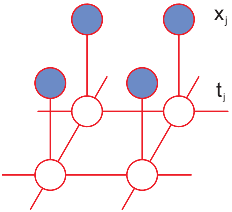
$$p(Y) = \frac{1}{Z} \prod_{(y_i, y_j) \in E} \psi_{ij}(y_i, y_j)$$

- Наиболее типичным примером таких сетей являются изображения



Марковские решетки с бинарными переменными

- В дальнейшем будем рассматривать марковские решетки такого вида

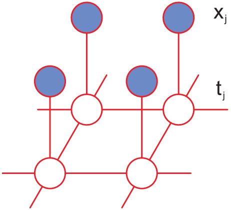


- Необходимо по наблюдаемым переменным X восстановить наиболее вероятные значения скрытых переменных T

$$T_{MP} = \arg\max_T P(T|X)$$

- Остановимся на важном частном случае, когда скрытые переменные бинарные $t \in \{0, 1\}$

Марковские решетки с бинарными переменными



- Распределение скрытых переменных марковской сети в этом случае выглядит так

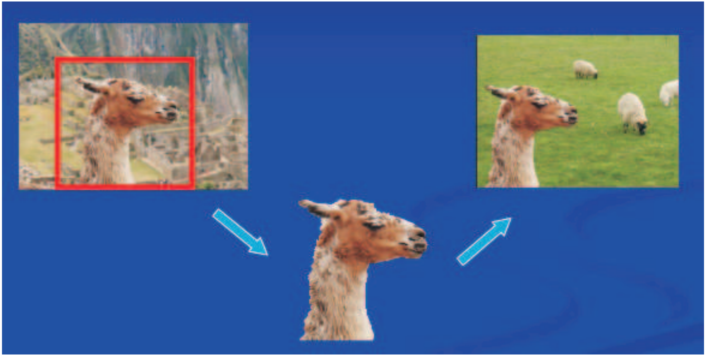
$$p(T|X) = \frac{p(X,T)}{p(X)} = \prod_{(i,j)\in E} \psi_{ij}(t_i,t_j) \prod_i \psi_i(x_i,t_i) \times \text{Const}$$

- В энергетической нотации задача максимизации этого распределения принимает вид минимизации энергии, что более удобно с вычислительной точки зрения

$$E(T|X) = \sum_{(i,j)\in E} \log \psi_{ij}(t_i,t_j) + \sum_i \log \psi_i(x_i,t_i) =$$

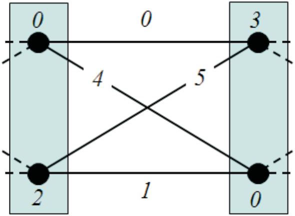
$$\sum_{(i,j)\in E} E_{ij}(t_i,t_j) + \sum_i E_i(x_i,t_i) \to \min_T \quad (1)$$

Пример задачи сегментации



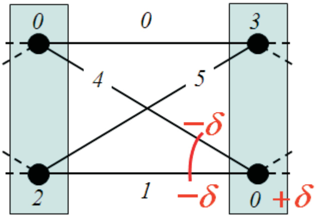
Значения скрытых переменных кодируют принадлежность каждого пикселя к объекту либо к фону. Использование графических моделей позволяет учесть, что соседние пиксели чаще всего относятся к одному классу

Репараметризация



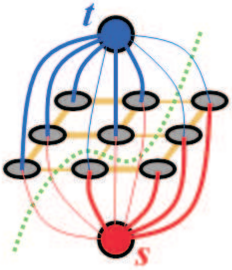
- Рассмотрим две соседние вершины t_i и t_j, каждая из которых может принимать два значения
- Тогда функция $E_i(x_i, t_i)$ задается двумя значениями (при известном x_i), а функция $E_{ij}(t_i, t_j)$ — четырьмя
- Для сведения к задаче о поиске минимального разреза нам понадобится выполнить т.н. репараметризацию, сделав «веса» горизонтальных ребер нулевыми

Репараметризация



- Вычитая одинаковые значения из двух ребер, сходящихся в одну вершину и прибавляя это же значение к весу самой вершины, мы получаем эквивалентный функционал энергии
- Значения такой энергии в каждой точке совпадает со значением исходной энергии
- Применение этой процедуры позволяет путем изменения функции E_i получить эквивалентный энергетический функционал, в котором $E_{ij}(0,0) = E_{ij}(1,1) = 0$ для всех $(i,j) \in E$

Сведение к разрезу графов



- Определим на следующем графе пропускную способность таким образом:

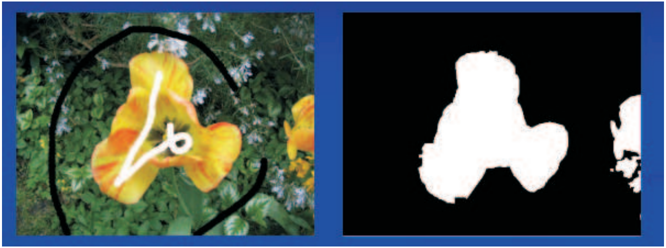
$$c(S, t_i) = E_i(x_i, 0), \quad c(T, t_i) = E_i(x_i, 1), \quad c(t_i, t_j) = E_{ij}(0, 1), \ \forall (i, j) \in E$$

- Поиск минимального разреза в таком графе отвечает минимизации энергии

$$E(T|X) = \sum_{(i,j) \in E} E_{ij}(t_i, t_j) + \sum_i E_i(x_i, t_i)$$

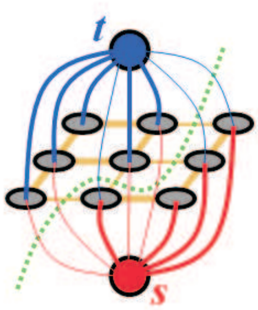
т.е. поиску наиболее вероятных значений T

Сегментация с семенами



- Часто для некоторых пикселей известно заранее, к какому классу они принадлежат
- Например, пользователь может задать фрагменты изображения и фона (семена)

Сведение к разрезу графов



- Пусть O — семена объекта ($t_i = 1$), а B — семена фона ($t_i = 0$)
- Тогда достаточно задать $c(S, t_i) = +\infty$, $\forall t_i \in O$ и $c(T, t_i) = +\infty$, $\forall t_i \in B$
- Этим мы запретим соответствующие разрезы, сделав невозможным отнесение семян объекта к фону и наоборот

Способы введения унарного слагаемого

По смыслу унарное (относительно скрытых переменных) слагаемое $E_i(x_i, t_i)$ задает насколько данный пиксель соответствует тому или иному классу. Оно может отражать следующую информацию:

- Цветовая модель — показывает насколько появление тех или иных цветов более вероятно в данном классе
- Позиционная модель — показывает априорные предположения о положении данного класса на изображении
- Текстурная модель — показывает насколько текстура окрестности пикселя вероятна для данного пикселя

Способы введения парного слагаемого

Парное слагаемое $E_{ij}(t_i, t_j)$ отражает степень взаимозависимостей классов соседних пикселей. Наиболее распространенными примерами являются

- Модель Поттса: $E_{ij}(t_i, t_j) = 1 - \delta(t_i, t_j)$ — штраф за несовпадение классов соседних пикселей
- Штраф за несовпадение классов с учетом контраста

$$E_{ij}(t_i, t_j) = \exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2}\right)(1 - \delta(t_i, t_j))$$

  Чем сильнее различаются цвета (интенсивности) пикселей x_i, тем меньше штраф за несовпадение классов
- Заметим, что во втором случае парное слагаемое зависит от наблюдаемых переменных x_i и x_j - такая конструкция называется условным случайным полем

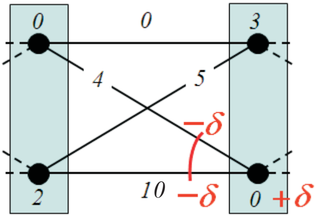
Субмодулярность

- Назовем энергию субмодулярной, если для всех ее парных слагаемых верно

$$E_{ij}(0,0)+E_{ij}(1,1)\leq E_{ij}(0,1)+E_{ij}(1,0)$$

- Условие субмодулярности является в некотором смысле аналогом выпуклости для функций бинарного переменного
- Унарное слагаемое при этом может быть произвольным
- Легко показать, что с помощью разрезов графов можно оптимизировать именно субмодулярную энергию

Доказательство необходимости

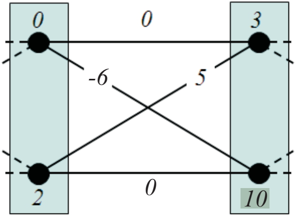


- Пусть имеется парное слагаемое, не являющееся субмодулярной функцией

$$E_{ij}(0,0) + E_{ij}(1,1) > E_{ij}(0,1) + E_{ij}(1,0)$$

- Тогда и только тогда в результате репараметризации нулевые веса горизонтальных связей приведут к возникновению отрицательных диагональных связей

Доказательство необходимости



- Но это означает, что пропускная способность некоторых ребер в графе, разрез которого мы будем минимизировать, станет отрицательной!
- В этой ситуации классические полиномиальные алгоритмы поиска минимального разреза в графе неприменимы
- Задача оптимизации энергии стала NP-трудной
- Существуют некоторые обобщения полиномиального алгоритма на случаи, когда энергия может быть сведена в субмодулярной путем замены части переменных на свои отрицания: $t_i \to (1 - t_i)$

Случай небинарных переменных

- До сих пор рассматривался случай, когда скрытые переменные бинарные
- Теперь рассмотрим ситуацию, когда скрытые переменные t_i могут принимать одно из K значений
- Физически это соответствует делению изображения на K областей
- Такие задачи возникают при построении карт диспаритетов, коллажах, семантической сегментации и пр.

Итерационная схема

$$E(T|X) = \sum_{(i,j)\in E} E_{ij}(t_i, t_j) + \sum_i E_i(x_i, t_i) \to \min_T, \quad t_i \in \{0, 1, \dots, K-1\}$$

- Задача оптимизации энергии по K-значным скрытым переменным ($K > 2$) является NP-трудной
- Тем не менее, в ряде случаев можно построить итерационную процедуру, сходящуюся к близкому к глобальному оптимуму ответу
- Наибольшее распространение получил алгоритм т.н. α-расширения

Итерационная схема

- Начинаем с произвольного начального приближения
- В цикле для каждой метки $\alpha \in \{0, \ldots, K-1\}$ заменяем часть других меток на данную, так чтобы минимизировать энергию (выполняем α-расширение)
- Если хотя бы для одной метки энергию удалось уменьшить, то переходим в предыдущему шагу, иначе выход

Для сходимости такого алгоритма необходимо, чтобы каждое парное слагаемое было метрикой в пространстве $\{0, \ldots, K-1\}$, т.е. удовлетворяло следующим условиям $\forall \alpha, \beta, \gamma \in \{0, \ldots, K-1\}$

- $E_{ij}(\alpha, \beta) = E_{ij}(\beta, \alpha)$ (симметричность)
- $E_{ij}(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta$ (аксиома тождества)
- $E_{ij}(\alpha, \gamma) \leq E_{ij}(\alpha, \beta) + E_{ij}(\beta, \gamma)$ (неравенство треугольника)

α-расширение

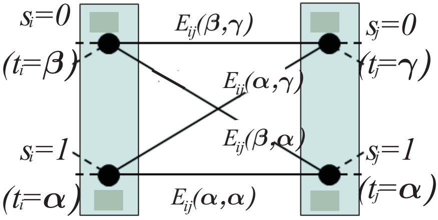
- В ходе альфа-расширения часть других меток принимает значение α, стремясь минимизировать энергию
- Введем вспомогательную марковскую решетку с бинарными переменными s_j, в которой значение 0 будет соответствовать тому, что исходные скрытые переменные t_j не изменились, а значение 1 будет означать, что исходные переменные приняли значение α

$$\forall j : t_j^{old} = \alpha \Rightarrow s_j \equiv 0$$

$$\forall j : t_j^{old} \neq \alpha \Rightarrow \begin{cases} s_j = 0 \Rightarrow t_j^{new} = t_j^{old} \\ s_j = 1 \Rightarrow t_j^{new} = \alpha \end{cases}$$

- Теперь относительно новых переменных можно построить минимальный разрез графа, минимизирующий энергию по всевозможным α-расширениям

α-расширение

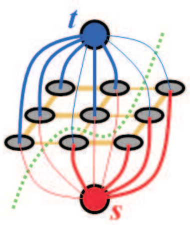


- Рассмотрим некоторую пару скрытых переменных (t_i, t_j), соединенную ребром
- Предположим, что старые значения переменных равнялись β и γ соответственно
- Для корректной репараметризации необходимо выполнение неравенства треугольника

$$E_{ij}(\beta,\gamma) \leq E_{ij}(\alpha,\gamma) + E_{ij}(\beta,\alpha)$$

- Теперь относительно новых переменных можно построить минимальный разрез графа, минимизирующий энергию по всевозможным α-расширениям

α-расширение



- Запретим изменять метки класса α

$$c(S, t_i) = E_i(x_i, \alpha), \quad c(T, t_i) = +\infty$$
$$c(t_i, t_j) = E_{ij}(\alpha, \alpha) = 0, \ \forall (i, j) \in E : \ t_i = t_j = \alpha$$

- Для остальных вершин ($t_i^{old} \neq \alpha$) определим пропускную способность следующим образом:

$$c(S, t_i) = E_i(x_i, \alpha), \quad c(T, t_i) = E_i(x_i, t_i^{old})$$
$$c(t_i, t_j) = E(\alpha, t_i^{old}), \quad c(t_j, t_i) = E(\alpha, t_j^{old}), \quad \forall (i, j) \in E$$

- Вершинам, попавшим в тот же подграф, что и исток S, будет присвоена метка α, энергия при этом уменьшится

Точность получающегося решения

- Алгоритм α-расширения является итерационным, полиномиальным, поэтому не гарантирует достижение глобального оптимума (NP-трудная задача)
- Можно показать, что значение энергии, получившейся в результате альфа-расширения, лежит в интервале

$$E(T^*) \le E(T) \le 2kE(T^*),$$

где T^* — оптимальное (наиболее вероятное) значение скрытых переменных, а

$$k = \frac{\max E_{ij}(\beta,\gamma)}{\min_{\beta\neq\gamma} E_{ij}(\beta,\gamma)}$$

степень контрастности парной энергии