Аппроксимации свободной энергии для марковских случайных полей

Чистяков А.С.

18 февраля 2014 г.

Обзор

Связь MRF и распределения Гиббса

Понятие и свойства свободной энергии

Способы аппроксимации (Mean Field, Bethe, Kikuchi)

Алгоритм распространения доверия для аппроксимаций Бете и Кикучи

Эксперименты

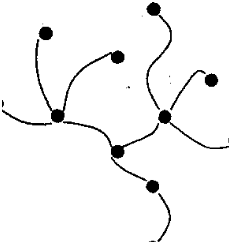
Обозначения

- G — неориентирванный граф без петель
- V — множество вершин графа ($|V| = S$)
- $\mathcal{C}(G)$ — множество клик графа G.
- $Y = (y_1, \dots, y_S)$ — семейство случайных величин с общим носителем
- $p(Y) = p(y_1, \dots, y_S)$ — совместное распределение во всех вершинах
- $p(Y_A)$ — плотность распределения на подмножестве вершин A
- $N'(v)$ — проколотая окрестность вершины v (соседние вершины)
- $N(v) = N'(v) \cup \{v\}$

Марковское случайное поле

Пара, состоящая из графа и семейства случайных величин на его вершинах, называется ***марковским случайным полем***, если для любой вершины v выполнено *локальное марковское свойство*:

$$p(y_v | Y_{V \setminus \{v\}}) = p(y_v | Y_{N'(v)})$$



Свойства марковского поля

Утверждение. Если все условные распределения $p(y_v|Y_{V\setminus\{v\}})$ положительны, то совместное распределение восстанавливается однозначно.

$$p(Y) = p(Y^*)\prod_{j=1}^{S}\frac{p(y_j|Y_{1,\dots,j-1},\,Y^*_{j+1,\dots,S})}{p(y^*_j|Y_{1,\dots,j-1},\,Y^*_{j+1,\dots,S})}$$

Пример. Если некоторые условные распределения принимают нулевые значения, то распределение $p(Y)$ может однозначно не восстанавливаться.

$$p(y_1|y_2) = p(y_2|y_1) = \delta_{y_1}^{y_2}$$

$$p(y_1, y_2) = \left(\begin{smallmatrix} x & 0 \\ 0 & 1-x \end{smallmatrix}\right)$$

Свойства марковского поля

Для марковского случайного поля выполняется **попарное марковское свойство**: *для любых двух не соседних вершин* v *и* v'

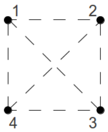
$$p(y_v, y_{v'} | Y_{V \setminus \{v, v'\}}) = p(y_v | Y_{V \setminus \{v, v'\}}) p(y_{v'} | Y_{V \setminus \{v, v'\}})$$

Свойства марковского поля

Утверждение. Если для всех пар вершин случайного поля со строго положительным распределением $p(Y)$ выполнено попарное марковское свойство, то оно является MRF.

Пример. Рассмотрим распределение на четырёх случайных величинах

$$p(0,0,0,0) = p(1,1,1,1) = \frac{1}{2}$$



Для любой пары вершин выполнено условие условной независимости, но все рёбра исключить нельзя. Также нельзя однозначно восстановить структуру графа.

Распределение Гиббса

Пусть $\mathbf{F} \subset \mathcal{C}(G)$. Распределение случайных величин на графе называется *распределением Гиббса*, если оно представимо в виде произведения

$$p(Y) = \prod_{C \in \mathbf{F}} \psi_C(Y_C)$$

Распределение Гиббса

Теорема

Любое распределение Гиббса на графе является марковским случайным полем.

Доказательство.

$$p(y_v|Y_{V\setminus\{v\}}) = \frac{p(y_1, Y_{V\setminus\{v\}})}{p(Y_{V\setminus\{v\}})} = \frac{p(Y)}{\sum_{y_v} p(Y)} = \frac{\prod_{C\in\mathbf{F}, C\ni v} \psi_C(Y_C)}{\sum_{y_v}\left(\prod_{C\in\mathbf{F}, C\ni v} \psi_C(Y_C)\right)}$$

$$p(y_v|Y_{N'(v)}) = \frac{p(y_v, Y_{N'(v)})}{p(Y_{N'(v)})} = \frac{\sum_{Y_{V\setminus N(v)}} p(Y)}{\sum_{y_v}\sum_{Y_{V\setminus N(v)}} p(Y)} = \frac{\prod_{C\in\mathbf{F}, C\ni v} \psi_C(Y_C)}{\sum_{y_v}\left(\prod_{C\in\mathbf{F}, C\ni v} \psi_C(Y_C)\right)}$$

$\square$

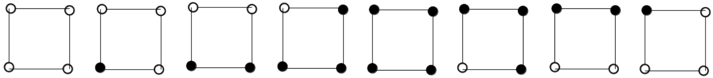
Распределение Гиббса

Теорема (Хаммерсли-Клиффорда)

Если распределение $p(Y)$ марковского случайного поля всюду положительно, то оно является распределением Гиббса.

Пример. Рассмотрим граф G — квадрат.

$$p(0,0,0,0) = p(0,0,0,1) = p(0,0,1,1) = p(0,1,1,1)$$
$$= p(1,1,1,1) = p(1,1,1,0) = p(1,1,0,0) = p(1,0,0,0) = \frac{1}{8}$$



Каждая клика принимает все свои значения с ненулевой вероятностью, а значит все $\psi_C(Y_C)$ должны быть положительны.

Энергия состояния марковской сети

Преобразуем вид распределения Гиббса для положительных распределений:

$$p(Y) = \prod_{C \in \mathbf{F}} \psi_C(Y_C) = \frac{1}{Z} e^{-\sum_{C \in \mathbf{F}} - \ln \psi_C(Y_C)}$$

$$= \frac{1}{Z} e^{-\sum_{C \in \mathbf{F}} E_C(Y_C)} = \frac{1}{Z} e^{-E(Y)}$$

Величину $E(Y)$ назовём *энергией состояния* марковской сети.

Напоминание о статфизе

- *Средней (Внутренней) энергией* $U(p)$ системы называется математическое ожидание энергии состояния

$$U(p) = \mathrm{E}_{p;Y} E(Y) = \sum_{Y} p(Y) E(Y)$$

- *Энтропией* $H(p)$ называется обычная энтропия распределения p

$$H(p) = \mathrm{E}_{p;Y}(-\ln p(Y)) = -\sum_{Y} p(Y) \ln p(Y)$$

- *Свободной энергией* $F(p)$ системы называется разность

$$F(p) = U(p) - H(p) = \sum_{Y} p(Y)(E(Y) + \ln p(Y))$$

Свойство распределения Гиббса

Теорема

Распределение Гиббса $p(Y) = \frac{1}{Z} e^{-E(Y)}$ *является точкой глобального минимума свободной энергии* $F(p)$ *по всем распределениям, а минимальное значение свободной энергии равно* $-\ln Z$.

Свойство распределения Гиббса
Доказательство

$$L(p,\lambda) = F(p) + \lambda\left(\sum_Y p(Y) - 1\right) = \sum_Y p(Y)(E(Y) + \ln p(Y) + \lambda) - \lambda$$

$$\frac{\partial L}{\partial p(Y_i)} = E(Y_i) + \ln p(Y_i) + \lambda + 1 = 0$$

$$\frac{\partial^2 L}{\partial p(Y_i)^2} = \frac{1}{p(Y_i)} > 0; \quad \frac{\partial^2 L}{\partial p(Y_i)\partial p(Y_j)} = 0$$

$$\ln p(Y_i) = -E(Y_i) - \lambda - 1 = \ln\left(\frac{1}{Z}e^{-E(Y_i)}\right)$$

Значит распределение Гиббса — действительно глобальный минимум.

$$F(p) = \sum_Y p(Y)(E(Y) - E(Y) - \ln Z) = -\ln Z$$

Способы аппроксимации свободной энергии

Mean field approximation

Предположим, что вершины независимы:

$$p(Y) = \prod_{j=1}^{S} p_j(y_j)$$

Свободная энергия имеет вид:

$$F_{\mathbf{MF}}(p_1, \ldots, p_S) = \sum_{Y} \prod_{j=1}^{S} p_j(y_j) \left(E(Y) + \sum_{j=1}^{S} \ln p_j(y_j) \right)$$

$$= \sum_{C \in \mathbf{F}} \sum_{Y_C} E_C(Y_C) \prod_{j: v_j \in C} p_j(y_j) + \sum_{j=1}^{S} \sum_{y_j} p_j(y_j) \ln p_j(y_j)$$

Mean field approximation

Оптимизировать будем методом покоординатного спуска.
Зафиксируем все распределения кроме j-го и сведём к задаче

$$\sum_{C \in \mathbf{F}: C \ni v_j} \sum_{Y_C} E_C(Y_C) \prod_{k: v_k \in C} p_k(y_k) + \sum_{y_j} p_j(y_j) \ln p_j(y_j) \to \min_{p_j}$$

Которую можно решить явно:

$$p_j(y_j) = \frac{1}{Z_j} e^{-\left(\sum_{C \in \mathbf{F}: C \ni v_j} \sum_{Y_{C \setminus v_j}} E_C(Y_C) \prod_{k: v_k \in C} p_k(y_k) \right)}$$

В силу ограниченности свободной энергии алгоритм сходится.

Mean field approximation

Достоинтсва:

- Задача сведена к разумной размерности: $(q-1)S$
- Задача минимизации имеет сходящееся решение (вариационный вывод)
- Даёт одностороннюю оценку $-\ln Z$

Недостатки:

- Может сходиться к локальному минимуму
- Результат может быть сколь угодно плохим
- Полученные распределения могут быть не похожи на реальные маргиналы

Bethe approximation

Рассмотрим пространство наборов распределений на кликах

$$\mathcal{P}_{\mathbf{F}} = \{p_C : C \in \mathbf{F}\}$$

Чтобы оно соответствовало распределению на всём графе, необходимо *условие согласованности*. Для любой тройки клик C, C', C'' такой что $\{C', C''\} \in \mathbf{F}$ и $C \subset C' \bigcap C''$ выполнено:

$$p_C(Y_C) = \sum_{Y_{C'\setminus C}} p_{C'}(Y_C, Y_{C'\setminus C}) = \sum_{Y_{C''\setminus C}} p_{C''}(Y_C, Y_{C''\setminus C})$$

Bethe approximation

Для графа с циклами условия согласованности не достаточно:

$$
\begin{tikzcd}
V_1\begin{pmatrix}0.5\\0.5\end{pmatrix} \arrow[rr, dash, "\left(\begin{smallmatrix}0.4 & 0.1\\0.1 & 0.4\end{smallmatrix}\right)"] \arrow[dr, dash, "\left(\begin{smallmatrix}0.1 & 0.4\\0.4 & 0.1\end{smallmatrix}\right)"'] & & V_2\begin{pmatrix}0.5\\0.5\end{pmatrix} \arrow[dl, dash, "\left(\begin{smallmatrix}0.4 & 0.1\\0.1 & 0.4\end{smallmatrix}\right)"] \\
& V_3\begin{pmatrix}0.5\\0.5\end{pmatrix} &
\end{tikzcd}
$$

Все $2^3 = 8$ состояний покрываются тремя 4-элементными множествами

$$
\begin{aligned}
p(Y \in \mathcal{Y}) \leqslant p(y_1 \neq y_2) + p(y_2 \neq y_3) + p(y_1 = y_3) = \\
= (0.1 + 0.1) + (0.1 + 0.1) + (0.1 + 0.1) = 0.6 < 1
\end{aligned}
$$

И сумма вероятностей событий, покрывающих достоверное событие, оказывается меньше 1.

Bethe approximation

Ограничимся только требованием локальной согласованности для одновершинных множеств C. Полученные семейства локальных вероятностных распределений $\mathcal{P}_{\mathbf{F}}$, не обязательно получающиеся из какого-то распределения на всём поле, называются *довериями (belief)*.
Заметим, что условие согласованности обеспечивает однозначное введение распределения на вершинах.

Bethe approximation

Пускай множество клик $\mathbf{F}$ также содержит все вершины. Если вершинам не приписана энергия, то её можно положить равной 0.

Введём обозначения:

$\mathbf{F}' = \mathbf{F} \setminus V$

$n_v = \#\{C \in \mathbf{F}' : C \ni v\}$

И свободную энергию Бете определим как

$$F_{\mathbf{F}}(\mathcal{P}_{\mathbf{F}}) = \sum_{C \in \mathbf{F}'} \left(U(p_C) + \sum_{v \in C} U(p_v) - H(p_C) \right) + \\ + \sum_{v \in V} (1 - n_v)(U(p_v) - H(p_v)) = \\ = \left(\sum_{C \in \mathbf{F}} U(p_C) \right) - \left(\sum_{C \in \mathbf{F}'} H(p_C) + \sum_{v \in V} (1 - n_v) H(p_v) \right) = \\ = U(\mathcal{P}_{\mathbf{F}}) + H_{\mathbf{F}}(\mathcal{P}_{\mathbf{F}})$$

Аппроксимация Bethe точна для деревьев

Утверждение. *Для распределения Гиббса на ациклическом графе*

$$p(Y) = \prod_{C \in \mathbf{F}'} \frac{p_C(Y_C)}{\prod_{v \in C} p_v(y_v)} \prod_{v \in V} p_v(y_v) = \prod_{C \in \mathbf{F}'} p_C(Y_C) \prod_{v \in V} (p_v(y_v))^{1-n_v}$$

Доказательство. Будем строить расширяющуюся последовательность поддеревьев $G_1 \subset G_2 \subset \ldots \subset G$, присоединяя по одному ребру. Правильность формулы доказывается по индукции:

$$\frac{p(Y_{G_i})}{p(Y_{G_{i-1}})} = p(Y_{C^{(i)}} | Y_{G_{i-1}}) = p(Y_{C^{(i)}} | Y_{v^{(i-1)}}) = \frac{p_{C^{(i)}}(Y_{C^{(i)}})}{p_{v^{(i-1)}}(y_{v^{(i-1)}})}$$

Следствие. Для ациклических графов энтропия Bethe совпадает со стандартной энтропией, а значит совпадают и свободные энергии.

Bethe approximation: итоги

Достоинства

- Задача сведена к разумной размерности: $(q^M - 1)KS$, где M — максимальный размер клики, K — количество типов клик;
- Точна для ациклических графов;
- Эффективно вычисляется при помощи алгоритма Belief Propagation;
- Успешно применяется на практике.

Недостатки

- Не даёт односторонних оценок;
- Алгоритмы могут сходиться к локальным минимумам;
- Алгоритмы могут сходиться не к распределениям;
- Алгоритмы могут вообще не сходится.

Kikuchi approximation

R — множество областей сети, не обязательно клики.
$\mathcal{P}_{\mathbf{R}}$ — семейство доверий, удовлетворяющих условию согласованности на **R**.
Каждой *области* припишем энергию

$$U(p_R) = \sum_{Y_R} p_R(Y) \sum_{C \in R} E_C(Y_C)$$

энтропию

$$H(p_R) = -\sum_{Y_R} p_R(Y) \ln p_R(Y)$$

и свободную энергию отдельной области

$$F(p_R) = U(p_R) - H(p_R)$$

Kikuchi approximation

Полную свободную энергию Кикучи можно представить в виде:

$$F_{\mathbf{R}}(\mathcal{P}_{\mathbf{R}}) = \sum_{R \in \mathbf{R}} c_R F(p_R) =$$

$$= \sum_{R \in \mathbf{R}} c_R \left(\sum_{Y_R} p_R(Y_R) E_R(Y_R) + \sum_{Y_R} p_R(Y_R) \ln p_R(Y_R) \right)$$

Для максимальных областей $c_R = 1$. Для остальных — обеспечивают точность вычисления средней энергии по формуле включений-исключений. ($c_R = 1 - \sum_{s \in super(R)} c_s$)

Пример:

Для тороидального графа-решётки

- $c_{square} = 1$
- $c_{edge} = 1 - 2c_{square} = -1$
- $c_{vertex} = 1 - 4c_{square} - 4c_{edge} = 1$

Kikuchi approximation: итоги

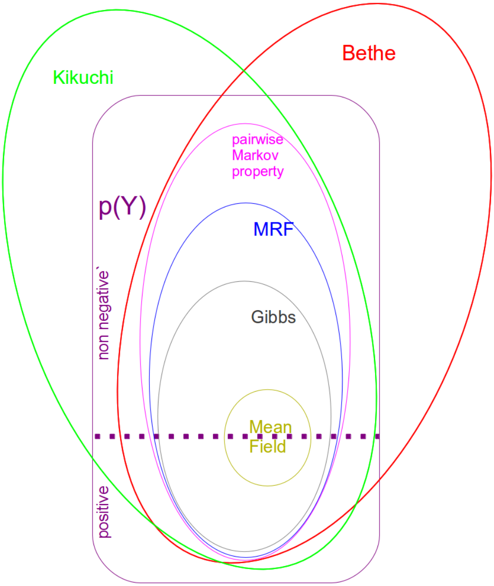
Достоинства

- Задача сведена к разумной размерности: $(q^M - 1)KS$, где M — максимальный размер области, K — количество типов областей;
- Эффективно вычисляется при помощи алгоритма GBP;
- Успешно применяется на практике.

Недостатки

- Не даёт односторонних оценок;
- Алгоритмы могут сходиться к локальным минимумам;
- Алгоритмы могут сходиться не к распределениям;
- Алгоритмы могут вообще не сходится.

Сравнение аппроксимаций



Loopy belief propagation (LBP)

Ограничимся рассмотрением распределений вида

$$p(Y) = \frac{1}{Z}\prod_{(vu)}\psi_{vu}(y_v, y_u)\prod_{v}\psi_v(y_v)$$

Для пересчёта *доверий* будем пользоваться формулами

$$m_{vu} \leftarrow \alpha\sum_{y_v}\psi_{vu}(y_v, y_u)\psi_v(y_v)\prod_{w\in N'(v)\setminus u} m_{wv}(y_w)$$

$$b_v(y_v) \leftarrow \alpha\psi_v(y_v)\prod_{u\in N'(v)} m_{uv}(x_v)$$

Дополнительно получим

$$b_{vu}(y_v, y_u) \leftarrow \alpha\psi_v(y_v)\psi_u(y_u)\psi_{vu}(y_v, y_u)\times$$
$$\times\prod_{w\in N'(v)\setminus u} m_{wv}(y_v)\prod_{t\in N'(u)\setminus v} m_{tu}(y_u)$$

Сходимость LBP

Утверждение. Пусть $\{m_{vu}\}$ — набор сообщений алгоритма BP и $\{b_v, b_{vu}\}$ — доверия, вычисленные по этим сообщениям. Данный набор доверий соответствует сходимости алгоритма BP тогда и только тогда, когда он соответсвует точке нулевого градиента свободной энергии Bethe:

$$F_{\mathbf{B}}(\{b_v, b_{vu}\}) = \sum_{vu} \sum_{y_v, y_u} b_{vu}(y_v, y_u)[\ln b_{vu}(y_v, y_u) - \ln \phi_{vu}(y_v, y_u)] + \\ + \sum_{v}(1 - n_v) \sum_{y_v} b_v(y_v)[\ln b_v(y_v) - \ln \psi_v(y_v)]$$

где $\phi_{vu}(y_v, y_u) = \psi_{vu}(y_v, y_u)\psi_v(y_v)\psi_u(y_u)$
с ограничениями на согласованность и нормированность:
$\sum_{y_v} b_v(y_v) = 1,$
$\sum_{y_v} b_{vu}(y_v, y_u) = b_u(y_u).$

Сходимость LBP

Доказательство. Выпишем лагранжиан задачи поиска нулевого градиента

$$\mathcal{L} = F_{\mathbf{B}} - \sum_{v \in V} \lambda_v \Big(\sum_{y_v} b_v(y_v) - 1\Big) - \sum_{(vu)} \lambda_{vu}^{(y_u)} \Big(\sum_{y_v} b_{vu}(y_v, y_u) - b_u(y_u)\Big)$$

$$\frac{\partial \mathcal{L}}{\partial b_{vu}(y_v, y_u)} = 0 \Rightarrow \quad \ln b_{vu}(y_v, y_u) = \ln \phi_{vu}(y_v, y_u) + \lambda_{uv}^{(y_v)} + \lambda_{vu}^{(y_u)} - 1$$

$$\frac{\partial \mathcal{L}}{\partial b_v(y_v)} = 0 \Rightarrow \quad \ln b_v(y_v) = \ln \psi_v(y_v) - 1 - \frac{\lambda_v}{n_v - 1} + \frac{\sum_{u \in N'(v)} \lambda_{uv}^{(y_v)}}{n_v - 1}$$

Сходимость LBP

Исходя из вида выражения для $\ln b_{vu}(y_v, y_u)$ положим

$$\lambda_{vu}^{(y_u)} = \ln \prod_{w \in N'(u) \setminus v} m_{wu}(y_u)$$

Если теперь набор доверий соответствует точке нулевого градиента, то полученный набор сообщений является стабилизировавшимся (следует из условия маргинализации доверий). И наоборот: если сообщения сошлись, то полученный из них набор доверий является точкой нулевого градиента свободной энергии.

Следствие. Если все $\psi_v(y_v)$ и $\psi_{vu}(y_v, y_u)$ положительны, то точка стабилизации сообщений существует.

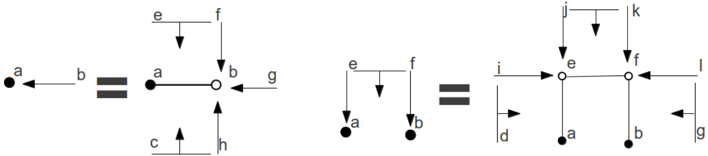
Generalized belief propagation (GBP)

Будем передавать сообщения не только между отдельными вершинами, но и между более крупными регионами графа. Добавим обозначений:

- $m_{rs}(y_s)$ — сообщение из региона $r \setminus s$ в регион s, где s — непосредственный подрегион r;
- $M(s)$ — множество сообщений $m_{r's'}$ таких что $(r' \setminus s') \cap s = \varnothing$ и $s' \subseteq s$;
- $M(r, s)$ — сообщения, начинающиеся в r и принадлежащие $M(s)$;

GBP: пересчёт сообщений

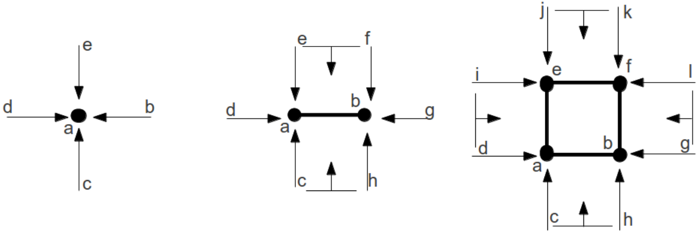
$$m_{rs} \leftarrow \alpha \left[\sum_{y_{r\setminus s}} \psi_{r\setminus s}(y_{r\setminus s}) \prod_{m_{r''s''} \in M(r)\setminus M(s)} m_{r''s''} \right] \Big/ \prod_{m_{r's'} \in M(r,s)} m_{r's'}$$



$$M_a^b(y_a) = \alpha \sum_{y_b} \psi_b(y_b)\psi_{ab}(y_a, y_b) M_{ab}^{ef} M_b^f M_b^g M_b^h M_{ab}^{ch}$$

GBP: Пересчёт доверий

$$b_r \leftarrow \alpha \psi_r(y_r) \prod_{m_{r's'} \in M(r)} m_{r's'}$$



$$b_{ab}(y_a, y_b) = \alpha \psi_{ab}(y_a, y_b) \psi_a(y_a) \psi_b(y_b) M_a^c M_a^d M_a^e M_{ab}^{ef} M_b^f M_b^g M_b^h M_{ab}^{ch}$$

Сходимость GBP

Утверждение. Пусть $\{m_{rs}(y_s)\}$ — набор сообщений алгоритма GBP и $\{b_r(y_r)\}$ — доверия, вычисленные по этим сообщениям. Данный набор доверий соответствует сходимости алгоритма GBP тогда и только тогда, когда он соответсвует точке нулевого градиента свободной энергии Кикучи:

$$F_{\mathbf{K}}(\mathcal{P}_{\mathbf{R}}) = \sum_{r \in \mathbf{R}} c_r \sum_{Y_R} b_r(Y_r) \left(\ln b_r(Y_r) - \ln \phi_r(Y_r)\right)$$

где $\phi_r(Y_r) = \prod_{s \subseteq r} \psi_s(Y_s)$

с ограничениями на согласованность и нормированность:

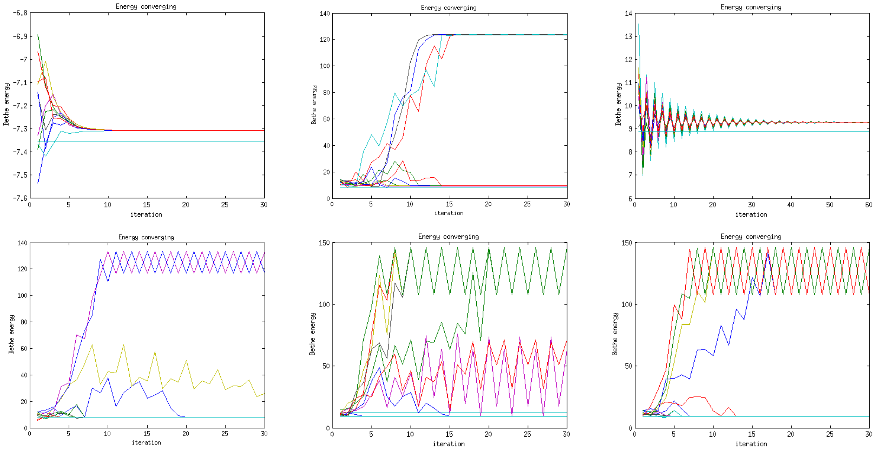
$\sum_{Y_r} b_r(Y_r) = 1$,

$\sum_{Y_{r \setminus s}} b_r(Y_r) = b_s(Y_s)$.

Эксперименты

Сходимость LBP

Зависимость энергии Bethe от итерации алгоритма BP.
Голубой чертой отмечено реальное значение $-lnZ$



Эмпирически энергия Bethe даёт одностороннюю оценку.

Сходимость LBP

Частоты попадания в локальный минимум, максимум и седловые точки.

	Лок. мин.	Лок. макс.	Седло	Сходимость
Рандом	15.7%	0.4%	83.9%	99%
Дом. диаг.	98.2%	0%	1.8%	70%

Проверка производилась путём рассчёта энергии Бете в 100 случайных точках из eps-окрестности и сравнения её со значением в точке стабилизации.

Сравнение LBP и GBP

Маргиналы, полученные при помощи LBP и GBP на бинарной прямоугольной решётке (10*10). (Результаты по одной строке)

Реальное значение	Маргиналы GBP	Маргиналы LBP
0.40131	0.40255	0.0043807
0.54038	0.54115	0.74502
0.48923	0.49184	0.32866
0.54506	0.54232	0.62190
0.44537	0.44812	0.37745
0.47856	0.48014	0.41243
0.51686	0.51501	0.57842
0.58108	0.57693	0.74555
0.57791	0.57710	0.85315
0.59881	0.59757	0.99632

Заключение

Проделанные шаги:

1. Распределение Гиббса описывает большой класс распределений на MRF;
2. Распределение Гиббса обладает минимумом свободной энергии по всем распределениям;
3. Для распределения Гиббса свободная энергия позволяет вычислить нормировочную константу;
4. Задача минимизации свободной энергии NP-полная, поэтому приходится её упрощать;
5. Мы разобрали упрощения двух видов:
   - Уменьшить пространство рассматриваемых распределений (Mean Field);
   - Расширить на семейство ”не совсем распределений” (Bethe, Kikuchi);

Ссылки

Y. Weiss J. S. Yeldia, W. T. Freeman.
Generalized belief propagation.
2000-06.

Andrea Montanari Marc Mezard.
Information, physics and computation.
2008-07.

А. Б. Мерков.
Введение в методы статистического обучения.
2014-01.