Интерфейс передачи сообщений в марковских сетях

1 Поиск наиболее вероятной конфигурации ацикличной марковской сети.

Рассмотрим марковское случайное поле, задаваемое ацикличным графом $\mathcal{G} = (V, E)$. Совместное распределение на переменные марковской сети может быть записано в следующем виде

$$p(X) = \frac{1}{Z} \exp\left(-\sum_{i\in V} \phi_j(x_j) - \sum_{(i,j)\in E} \phi_{ij}(x_i, x_j)\right) = \frac{1}{Z} \exp\left(-E(X)\right).$$

Рассмотрим задачу поиска наиболее вероятной конфигурации переменных сети. Здесь и далее будем полагать, что переменные $x_i \in \{1, \dots, K\}$[1]. Очевидно, что

$$X^* = \arg\max p(X) = \arg\max \log p(X) = \arg\min E(X).$$

Задача поиска X^* является примером задачи дискретной оптимизации и, в общем случае, может быть решена только полным перебором по K^N возможным значениям X. Здесь $N = |V|$. Для ациклических марковских сетей существует более эффективный алгоритм известный как min-sum (belief propagation, динамическое программирвоание, алгоритм Витерби), основанный на т.н. интерфейсе передачи сообщений (message passing interface, MPI).

Ациклический граф является деревом. Выделим в нем произвольную вершину, которую будем считать корнем дерева. Пронумеруем все вершины графа от 1 до N так, чтобы путь от любого листа до корня состоял из вершин с убывающими номерами. Это можно сделать, например, последовательно нумеруя все вершины одного яруса и двигаясь по ярусам от корня к листьям. Обозначим T_i поддерево, порожденное вершиной i, т.е. вершины с номерами $j \geq i$, путь от которых до корня проходит через вершину i. Пусть $(i, j) \in E$, $i < j$. Не ограничивая общности, предположим, что количество ребер, связанных с корнем, равно двум. Рассмотрим следующую функцию

$$V_i(x_i) = \min_{\{x_j : j\in T_i\setminus\{i\}\}} \left[\sum_{j\in T_i} \phi_j(x_j) + \sum_{(j,k)\in E j,k\in T_i} \phi_{jk}(x_j, x_k)\right].$$

Очевидно, что $\min_X E(X) = \min_{x_1} V_1(x_1)$. С другой стороны,

$$\begin{aligned}
V_1(x_1) = \min_{x_2,\dots,x_N} \left(\phi_1(x_1) + \phi_{12}(x_1, x_2) + \phi_{13}(x_1, x_3) + \sum \dots\right) = \{\text{Distributivity}\} = \\
= \min_{x_2,x_3} \left(\phi_1(x_1) + \phi_{12}(x_1, x_2) + \phi_{13}(x_1, x_3) + \min_{x_4,\dots,x_N} \sum \dots\right) = \{\text{Acyclicity}\} = \\
= \min_{x_2,x_3} \left(\phi_1(x_1) + \phi_{12}(x_1, x_2) + \phi_{13}(x_1, x_3) + \min_{\{x_j : j\in T_2\setminus\{2\}\}} \sum \dots + \min_{\{x_j : j\in T_3\setminus\{3\}\}} \sum \dots\right) = \\
= \min_{x_2,x_3} \left(\phi_1(x_1) + \phi_{12}(x_1, x_2) + \phi_{13}(x_1, x_3) + V_2(x_2) + V_3(x_3)\right) = \\
= \phi_1(x_1) + \min_{x_2} \left(\phi_{12}(x_1, x_2) + V_2(x_2)\right) + \min_{x_3} \left(\phi_{13}(x_1, x_3) + V_3(x_3)\right) = \phi_1(x_1) + \mu_{2\to 1}(x_1) + \mu_{3\to 1}(x_1).
\end{aligned}$$

[1] Все последующие выкладки справедливы и для непрерывных переменных, но практическое применение описанных алгоритмов в последнем случае затруднительно за исключением нескольких специальных случаев, например, линейных динамических систем

Таким образом, нам удалось выразить $V_1(x_1)$ через $V_2(x_2)$ и $V_3(x_3)$. Функции

$$\mu_{j\to i}(x_i) = \min_{x_j}\left(\phi_{ij}(x_i, x_j) + V_j(x_j)\right)$$

называют сообщениями от вершины j к вершине i. При выводе формулы пересчета была использована дистрибутивность операции минимума и суммы, а также тот факт, что $T_2 \cap T_3 = \emptyset$, вытекающий из ацикличности графа $\mathcal{G}$. В общем случае, справедливо

$$V_i(x_i) = \min_{\{x_j : j\in Child(i),\ (i,j)\in E\}}\left(\phi_i(x_i) + \sum_{\{j\in Child(i),\ (i,j)\in E\}}\phi_{ij}(x_i, x_j) + \sum_{\{j\in Child(i),\ (i,j)\in E\}} V_j(x_j)\right) =$$
$$= \phi_i(x_i) + \sum_{\{j:(i,j)\in E, j\in T_i\}}\mu_{j\to i}(x_i)$$

и, следовательно,

$$\mu_{j\to i}(x_i) = \min_{x_j}\left(\phi_{ij}(x_i, x_j) + \phi_j(x_j) + \sum_{\{k:(j,k)\in E, k\in T_j\}}\mu_{k\to j}(x_j)\right).$$

В листьях ($T_j = \{j\}$) фукция $V_j(x_j)$, очевидно, инициализируется так $V_j(x_j) = \phi_j(x_j)$. Благодаря такому индуктивному соотношению между значениями функций $V_j(x_j)$ удается эффективно рассчитать функцию $V_1(x_1)$ и, следовательно, найти минимум энергии $E(X)$. Такой способ подсчета называется проходом вверх (от листьев к корню).

Дополнительно определим функцию

$$S_{j\to i}(x_i) = \arg\min_{x_j}\left(\phi_{ij}(x_i, x_j) + \phi_j(x_j) + \sum_{\{k:(j,k)\in E, k\in T_j\}}\mu_{k\to j}(x_j)\right).$$

Тогда для нахождения оптимальной конфигурации X^* достаточно выполнить проход сверху-вниз (backtracking). В качестве x_1^* возьмем $\arg\min V_1(x_1)$, а оптимальные значения последующих переменных пересчитываются по формуле

$$x_j^* = S_{j\to i}(x_i^*).$$

2 Передача сообщений. Подсчет мин-маргиналов.

Заметим, что $V_1(x_1) = \min_{x_2,\ldots,x_N} E(X|x_1)$, т.е. помимо минимума энергии мы дополнительно знаем т.н. мин-маргиналы (оптимальные значения энергии при условии, что значение первой переменной фиксировано). В принципе, для подсчета произвольных мин-маргиналов, можно было бы последовательно объявить каждую вершину корнем и сделать проход вверх, но есть способ экономичнее. Для этого пустим сообщения от корня к листьям

$$\mu_{1\to 2}(x_2) = \min_{x_1}\left(\phi_{12}(x_1, x_2) + \phi_1(x_1) + \sum_{\{j:(1,j)\in E, j\neq 2\}}\mu_{j\to 1}(x_1)\right) =$$
$$\min_{x_1}\left(\phi_{12}(x_1, x_2) + \phi_1(x_1) + \mu_{3\to 1}(x_1)\right).$$

В общем случае формула для подсчета нисходящего сообщения выглядит так

$$\mu_{i\to j}(x_j) = \min_{x_i}\left(\phi_{ij}(x_i, x_j) + \phi_i(x_i) + \sum_{\{k:(k,i)\in E, k\neq j\}}\mu_{k\to i}(x_i)\right).$$

Заметим, что на каждом шаге все величины, необходимые для подсчета очередного сообщения $\mu_{i\to j}(x_j)$, нам известны. Легко видеть, что

$$E(X|x_1) = V_1(x_1) = \phi_1(x_1) + \sum_{\{j:(1,j)\in E\}} \mu_{j\to 1}(x_1),$$

Аналогично, мин-маргиналы для других вершин рассчитываются как

$$E(X|x_j) = \phi_j(x_j) + \sum_{\{k:(j,k)\in E\}} \mu_{k\to j}(x_j).$$

Последняя формула становится очевидной, если мысленно сдвинуть корень дерева в вершину j. Ясно, что эта операция корректна в силу произвольности выбора корня в дереве.

3 Подсчет маргинальных распределений в дереве

Рассмотрим теперь задачу подсчета маргинальных распределений

$$p(x_j) = \sum_{X\setminus x_j} p(X) = \frac{\sum_{X\setminus x_j} \exp(-E(X))}{Z} = \frac{\sum_{X\setminus x_j} \exp(-E(X))}{\sum_X \exp(-E(X))}.$$

Перепишем совместное распределение в виде произведения

$$p(X) = \frac{1}{Z}\exp\left(-\sum_{i\in V}\phi_j(x_j) - \sum_{(i,j)\in E}\phi_{ij}(x_i,x_j)\right) = \frac{1}{Z}\prod_{i\in V}\psi_i(x_i)\prod_{(i,j)\in E}\psi_{ij}(x_i,x_j).$$

Будем действовать аналогично алгоритму поиска минимума. Определим функцию

$$V_i(x_i) = \sum_{\{x_j: j\in T_i\setminus\{i\}\}}\left[\prod_{j\in T_i}\psi_j(x_j)\prod_{(j,k)\in E j,k\in T_i}\psi_{jk}(x_j,x_k)\right].$$

Очевидно, что $p(x_1) = \sum_{X\setminus\{1\}} p(X)$. С другой стороны,

$$\begin{aligned}
V_1(x_1) &= \sum_{x_2,\dots,x_N}\left(\psi_1(x_1)\psi_{12}(x_1,x_2)\psi_{13}(x_1,x_3)\prod\dots\right) = \{\text{Distributivity}\} = \\
&= \sum_{x_2,x_3}\left(\psi_1(x_1)\psi_{12}(x_1,x_2)\psi_{13}(x_1,x_3)\sum_{x_4,\dots,x_N}\prod\dots\right) = \{\text{Acyclicity}\} = \\
&= \sum_{x_2,x_3}\left(\psi_1(x_1)\psi_{12}(x_1,x_2)\psi_{13}(x_1,x_3)\left(\sum_{\{x_j:j\in T_2\setminus\{2\}\}}\prod\dots\right)\left(\sum_{\{x_j:j\in T_3\setminus\{3\}\}}\prod\dots\right)\right) = \\
&= \sum_{x_2,x_3}\left(\psi_1(x_1)\psi_{12}(x_1,x_2)\psi_{13}(x_1,x_3)V_2(x_2)V_3(x_3)\right) = \\
&= \psi_1(x_1)\left(\sum_{x_2}\psi_{12}(x_1,x_2)V_2(x_2)\right)\left(\sum_{x_3}\psi_{13}(x_1,x_3)V_3(x_3)\right) = \psi_1(x_1)\mu_{2\to 1}(x_1)\mu_{3\to 1}(x_1).
\end{aligned}$$

Как и ранее, не ограничивая общности, мы предположили, что корень дерева связан с двумя вершинами. Данная формула позволяет индуктивно пересчитывать функцию $V_j(x_j)$ двигаясь по дереву от листьев к корню, т.е. уменьшая j. Функции-сообщения принимают вид

$$\mu_{j\to i}(x_i) = \sum_{x_j}\psi_{ij}(x_i,x_j)V_j(x_j) = \sum_{x_j}\left(\psi_{ij}(x_i,x_j)\psi_j(x_j)\prod_{\{k:(j,k)\in E, k\in T_j\}}\mu_{k\to j}(x_j)\right).$$

Сообщения от листьев имеют вид $\mu_{k\to i}(x_i) = \sum_{x_k} \psi_{ik}(x_i, x_k)\psi_k(x_k)$.

Для расчета маргинального распределения произвольной вершины, не являющейся корнем необходимо выполнить проход вниз, т.е. от корня к листьям. Нисходящие сообщения принимают вид

$$\mu_{i\to j}(x_j) = \sum_{x_i} \left(\psi_{ij}(x_i, x_j)\psi_i(x_i) \prod_{\{k:(k,i)\in E, k\neq j\}} \mu_{k\to i}(x_i) \right).$$

Маргинальное распределение с точностью до множителя имеет вид

$$p(x_i) = \frac{1}{Z}\psi_i(x_i) \prod_{\{j:(i,j)\in E\}} \mu_{j\to i}(x_i),$$

а нормировочная константа равна

$$Z = \sum_{x_i} \psi_i(x_i) \prod_{\{j:(i,j)\in E\}} \mu_{j\to i}(x_i), \quad \forall i.$$

В частности, можно взять $i = 1$ и получить $Z = \sum_{x_1} V_1(x_1)$. Таким образом, кроме подсчета маргинальных распределений, нам удалось оценить и значение нормировочной константы Z.

4 Произвольные полукольцевые операции на дереве

Оба описанных алгоритма (подсчет мин-маргиналов и подсчет маргинальных распределений) имеют много общего. В обоих случаях имеется некая функция от N переменных, представимая в виде композиции атомарных функций от одной и двух переменных с помощью некоторой алгебраической операции. Далее по всем переменным кроме одной выполняется маргинализация (т.е. исключение путем применения к их всевозможным значениям другой алгебраической операции). В обоих случаях используется схема передачи сообщений, позволяющая эффективно провести маргинализацию в случае, когда 1) при связывании ребром пеерменных, входящие в одну и ту же атомарную функцию, у нас получается ацикличный граф; 2) две алгебраические операции образуют полукольцо.

Напомним, что полукольцом называется следующее множество S с двумя определенными на нем операциями $\oplus$ и $\otimes$:

- $(a \oplus b) \oplus c = a \oplus (b \oplus c)$;
- $a \oplus b = b \oplus a$;
- $a \oplus 0 = a$;

} Коммутативный моноид с операцией $\oplus$ и единичным элементом 0

- $(a \otimes b) \otimes c = a \otimes (b \otimes c)$;
- $a \otimes 1 = 1 \otimes a = a$;

} Моноид с операцией $\otimes$ и единичным элементом 1

- $a \otimes (b \oplus c) = a \otimes b \oplus a \otimes c$;
- $(b \oplus c) \otimes a = b \otimes a \oplus c \otimes a$;

} Дистрибутивность

- $a \otimes 0 = 0 \otimes a = 0$.

} Обнуление умножения

Таким образом, если на ацикличном графе $G = (V, E)$ определена функция вида

$$F(X) = \bigotimes_{i \in V} f_i(x_i) \bigotimes_{(i,j) \in E} f_{ij}(x_i, x_j),$$

с помощью интерфейса передачи сообщений за два прохода по дереву для любого i может быть эффективно рассчитана функция вида

$$m(x_i) = \bigoplus_{X \setminus \{x_i\}} F(X).$$

Примеры полуколец на множестве действиетльных чисел (первая операция соответствует $\oplus$, а вторая $\otimes$): $(\max, \min)$, $(+, \cdot)$, $(\min, \max)$, $(\min, +)$, $(\max, \cdot)$.

5 Открытая проблема

В случае циклических графов можно иниицировать передачу сообщений из случано выбранных узлов и циклически передавать их до сходимости. Такой подход существет и называется loopy belief propagation, но он не гарантирует даже сходимости, не говоря уже о сходимоси к точному ответу. Более того, доказано, что в общем случае задача подсчета подобных маргиналов в циклических графах является NP-трудной. Тем не менее, для полукольца $(\min, +)$ существует важный частный случай, когда задача может быть решена полиномиально. Это так для $K = 2$ и т.н. субмоудлярных функций, т.е. функций, удовлетворяющих условию

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

Интересным вопросом является возможность выписывания условия, аналогичного субмодялурности в для случая произвольного полукольца (или хотя бы для полукольца $(+, \cdot)$, представляющего практическую ценность применительно к алгоритмам вывода в марковских случайных полях).