УДК 519.21, 621.39

Г.Ф. КОЛОМИЕЦ, Е.И. ЛАВРИКОВА, М.Н. СИНЯКОВ

БАЙЕСОВА ОЦЕНКА ГИПОТЕЗ ДЛЯ ПРИЧИННО-СЛЕДСТВЕННОГО АНАЛИЗА МНОГОМЕРНЫХ ПРОЦЕССОВ

Решена задача причинно-следственного анализа детерминированных бесконечно возвратных процессов на примере построения байесовой оценки гипотезы о синхронной ОМВ, полученной в результате анализа конечного отрезка детерминированного бесконечно возвратного процесса.

Очевидно, что обобщенные модели взаимодействия (ОМВ), построенные по конечным отрезкам процессов [1], можно рассматривать лишь как гипотезы об истинных взаимодействиях между переменными исследуемых процессов. Поэтому следующий этап решения задачи причинно-следственного анализа (ПСА) детерминированных бесконечно возвратных (д.б.в.) процессов состоит в оценке таких гипотез.

Ниже приведен пример построения байесовой оценки гипотезы о синхронной ОМВ, полученной в результате анализа конечного отрезка детерминированного бесконечно возвратного процесса.

Формулировка задачи. Заданы: $M = \{1, ..., m\}$ – множество меток переменных $\mathsf{x}_1, ..., \mathsf{x}_m$ [2], характеризующих процесс x; $X_1, ..., X_m$ – конечные множества допустимых значений этих переменных; x – детерминированный бесконечно возвратный процесс порядка N ($N \geq 1$) из множества $\mathcal{X}_N$ процессов вида

$$\mathcal{X}_N \underset{Df}{=} \{\mathsf{x} \in (\prod_{i \in M} X_i)^{\mathsf{N}^*} : (\forall t \in \mathsf{N}^*)(\exists t' \in \mathsf{N}^*)$$
$$(\forall \tau \in [1, N])(t < t' \wedge \mathsf{x}(t+\tau) = \mathsf{x}(t'+\tau))\}.$$

В течение конечного отрезка времени $T \in \mathsf{N}^*$ наблюдается отрезок процесса $\tilde{\mathsf{x}} \subset \mathsf{x}$.

Требуется построить такую гипотезу о синхронных взаимодействиях переменных $\mathsf{x}_1, ..., \mathsf{x}_m$ процесса x, для которой отношение ее апостериорной вероятности к априорной достаточно велико (например, больше некоторой заданной величины).

Для решения задачи в такой формулировке

необходимо: 1) определить множество E гипотез о синхронных взаимодействиях между переменными процесса X и построить на нем априорное распределение вероятностей; 2) найти способ оценивания апостериорных вероятностей гипотез из E по наблюдениям состояний процесса $\tilde{\mathsf{X}} \subset \mathsf{X}$; 3) построить собственно алгоритм выделения и оценки гипотез о синхронных взаимодействиях переменных $\mathsf{x}_1, ..., \mathsf{x}_m$ процесса X при наблюдении конечного отрезка процесса $\tilde{\mathsf{X}} \subset \mathsf{X}$.

Решение задачи. Пусть $\mathsf{R}^1 = \wp(\prod_{i \in M} X_i) \setminus \{\varnothing\}$ – множество всех отношений порядка 1, ассоциированных со всеми процессами порядка 1 из X_N. Каждому $R \in \mathsf{R}^1$ можно поставить в соответствие единственную ОМВ, которую мы обозначим $\mathsf{G}(R)$. Тогда множество гипотез E представляет собой множество ОМВ, таких, что $E = \mathsf{G}(\mathsf{R}^1)$. Отображение $\mathsf{G} : \mathsf{R}^1 \to E$ не взаимно однозначно, так как легко показать, что одна и та же ОМВ может соответствовать различным отношениям из R^1, т.е. могут существовать $R \in \mathsf{R}^1$ и $R' \in \mathsf{R}^1$ такие, что $\mathsf{G}(R) = \mathsf{G}(R')$. Отсюда следует, что $|E| \le |\mathsf{R}^1|$. Очевидно, что если $P(R)$ есть вероятность (априорная либо апостериорная) отношения $R \in \mathsf{R}^1$, а $B(\mathsf{G}(R))$ – вероятность ОМВ $\mathsf{G}(R) \in E$, то $P(R) \le B(\mathsf{G}(R))$. Это соотношение позволяет оценивать снизу вероятности на множестве E с помощью соответствующих вероятностей на множестве R^1. Таким образом, задачу построения априорного и апостериорного распределения вероятностей на множестве E можно заменить задачей их оценивания с помощью построения соответствующих распределений на множестве R^1.

Априорное распределение на множестве R^1 строится следующим образом. Среди всех наборов из множества $S = \prod_{i \in M} X_i$ проводится жеребьевка, в результате которой каждый набор либо принимается с вероятностью p, либо отвергается с вероятностью $1 - p$. Результатом каждой такой жеребьевки является некоторое отношение R из R^1, представляющее собой совокупность принятых наборов. Вероятность $Q(R)$ того, что результатом жеребьевки окажется конкретное отношение $R \in \mathsf{R}^1$, и принимается за его априорную вероятность. Эта вероятность

$$Q(R) = \frac{p^{|R|} \cdot (1-p)^{|S|-|R|}}{1-(1-p)^{|S|}},$$

где $|S|$ и $|R|$ – мощности множеств S и R. (В этом выражении знаменатель осуществляет нормировку так, чтобы $\sum_{Q \in \mathsf{R}^1} Q(R)$ была равна 1. Необходимость нормировки вызвана отсутствием в R^1 множества $\{\varnothing\}$).

В ситуации, когда отсутствует дополнительная информация о вероятности появления произвольного набора из S в конкретном отношении $R \in \mathsf{R}^1$ (такая ситуация достаточно типичная) естественно предположить, что $p = 0{,}5$. В этом случае априорная вероятность отношения R

$$Q(R) = \frac{1}{2^{|S|} - 1} \tag{1}$$

и одинакова для всех $R \in \mathcal{R}^1$.

Примем следующее допущение. Будем считать, что если наблюдается процесс x, у которого ассоциированное отношение есть R_{x}, то для любого момента времени $t \in N^*$ вероятность того, что любой набор из R_{x} станет состоянием процесса x в этот момент времени, одна и та же и не зависит ни от момента времени t, ни от данного набора, ни от предыстории процесса. (Такое допущение приемлемое для бесконечно возвратных детерминированных процессов).

Апостериорной вероятностью $P_T(R \mid ¥)$ любого отношения $R \in \mathcal{R}^1$ будем называть вероятность того, что $R = R_{x}$ при условии, когда наблюдался отрезок процесса ¥ в течение времени T. Достаточной статистикой для вычисления апостериорной вероятности является частичное отношение $R_{¥}$:

$$R_{¥} = \{(x_i(t) \in X_i \mid i \in M) : t \in T\}.$$

(Очевидно, что $R_{¥} \subset R_{x}$). Поэтому вместо $P_T(R \mid ¥)$ будем писать $P_T(R \mid R_{¥})$.

Рассмотрим ситуацию, при которой $R = R_{¥}$. Тогда

$$P_T(R_{¥} \mid R_{¥}) = P(R_{¥} = R_{x} \mid R_{¥}, T) = \frac{P(R_{¥}, T \mid R_{¥} = R_{x}) Q(R_{x})}{P(R_{¥}, T)}, \tag{2}$$

где $Q(R_{x})$ – априорная вероятность отношения R_{x}.

Обозначим X^T множество начальных отрезков длины T д.б.в. процессов порядка 1 из множества X_N. Для всякого $x^T \in X^T$ R_{x^T} обозначим частичное отношение порядка 1, построенное по отрезку длины T процесса x. Определим $P(R_{¥}, T \mid R_{¥} = R_{x})$ – вероятность того, что за время T наблюдалось частичное отношение $R_{¥}$ при условии, что это отношение является ассоциированным, – следующим образом.

В каждый момент времени $t \in T$ с вероятностью $\frac{1}{|R_{¥}|}$ в отрезке ¥ имеется некоторый набор из отношения $R_{¥}$. Тогда $\left(\frac{1}{|R_{¥}|}\right)^T$ – вероятность такого отрезка процесса ¥, которому соответствует частичное отношение $R_{¥}$. Величина $|\{x^T \in X^T : R_{x^T} = R_{¥}\}|$ определяет количество д.б.в. процессов в множестве X_N, начальным отрезкам которых длины T соответствует частичное отношение $R_{¥}$.

Тогда

$$P(R_{\text{¢}}, T \mid R_{\text{¢}} = R_{\mathsf{x}}) = \left(\frac{1}{|R_{\text{¢}}|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}|. \tag{3}$$

По формуле полной вероятности

$$P(R_{\text{¢}}, T) = \sum_{\substack{R \in \mathcal{R}^1 \\ R_{\text{¢}} \subset R}} Q(R) \cdot P(R_{\text{¢}}, T \mid R = R_{\text{¢}}),$$

где, в свою очередь,

$$P(R_{\text{¢}}, T \mid R = R_{\text{¢}}) = \begin{cases} \left(\dfrac{1}{|R_{\text{¢}}|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}| & \text{при } R_{\text{¢}} \subset R \\ 0 & \text{иначе.} \end{cases}$$

Тогда

$$P(R_{\text{¢}}, T) = \sum_{\substack{R \in \mathcal{R}^1 \\ R_{\text{¢}} \subset R}} Q(R) \cdot \left(\frac{1}{|R|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}|.$$

Таким образом, выражение (2) примет следующий вид:

$$P(R_{\text{¢}} = R_{\mathsf{x}} \mid R_{\text{¢}}, T) = \frac{\left(\dfrac{1}{|R_{\text{¢}}|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}| \cdot\ Q(R_{\text{¢}})}{\sum\limits_{\substack{R \in \mathcal{R}^1 \\ R_{\text{¢}} \subset R}} Q(R) \cdot \left(\dfrac{1}{|R|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}|}. \tag{4}$$

В соответствии с выражением (1) для априорных вероятностей упростим (4):

$$\begin{aligned} P(R_{\text{¢}} = R_{\mathsf{x}} \mid R_{\text{¢}}, T) &= \frac{\left(\dfrac{1}{|R_{\text{¢}}|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}| \cdot \dfrac{1}{2^{|S|} - 1}}{\sum\limits_{\substack{R \in \mathcal{R}^1 \\ R_{\text{¢}} \subset R}} \dfrac{1}{2^{|S|} - 1} \cdot \left(\dfrac{1}{|R|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}|} = \\ &= \frac{\left(\dfrac{1}{|R_{\text{¢}}|} \right)^{T} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}| \cdot \dfrac{1}{2^{|S|} - 1}}{\dfrac{1}{2^{|S|} - 1} \cdot |\{\mathsf{x}^{T} \in X^{T} : R_{\mathsf{x}^{T}} = R_{\text{¢}}\}| \cdot \sum\limits_{\substack{R \in \mathcal{R}^1 \\ R_{\text{¢}} \subset R}} \left(\dfrac{1}{|R|} \right)^{T}} = \frac{1}{(|R_{\text{¢}}|)^{T} \cdot \sum\limits_{\substack{R \in \mathcal{R}^1 \\ R_{\text{¢}} \subset R}} \left(\dfrac{1}{|R|} \right)^{T}}. \end{aligned} \tag{5}$$

При байесовом подходе оценкой гипотезы является отношение апостериорной вероятности к априорной.

Учитывая, что при $p = 0{,}5$ (вероятность появления произвольного набора из $\prod_{i \in M} X_i$ в частичном отношении) априорные вероятности всех частичных отношений равны, выражение (5) позволяет оценивать гипотезы из множества E: для любого частичного отношения $R \in \mathcal{R}^1$ апостериорная вероятность (5) является оценкой снизу для гипотезы $G(R)$ в силу соотношения $P(R) \le B(G(R))$ для всякого $R \in R^1$.

Упростим выражение (5). В этом выражении величина $|R|$ изменяется от $|R_{Ӿ}|$ до $|S|$, а количество отношений R разной мощности изменяется от $C^0_{|S|-|R_{Ӿ}|}$ до $C^{|S|-|R_{Ӿ}|}_{|S|-|R_{Ӿ}|}$ соответственно. Тогда

$$P(R_{Ӿ} = R_{x} \mid R_{Ӿ}, T) = \frac{1}{(R_{Ӿ})^T \cdot \sum\limits_{\substack{k=0 \\ n=|R_{Ӿ}|}}^{\substack{|S| \\ |S|-|R_{Ӿ}|}} C^k_{|S|-|R_{Ӿ}|} \cdot \frac{1}{n^T}}. \qquad (6)$$

В выражении (6) величины T и $|S|$ известны из условий задачи. По конечному отрезку процесса Ӿ легко построить $R_{Ӿ}$ и, соответственно, вычислить апостериорную вероятность отношения $R_{Ӿ}$.

Теперь можно сформулировать алгоритм выделения и оценки гипотез о синхронном взаимодействии между переменными $x_1, ..., x_m$ процесса x по конечному отрезку процесса $Ӿ \subset x$, наблюдаемому в течение времени T.

1. Определить значение $|S|$ и по формуле (1) вычислить величину априорной вероятности $Q(R)$ произвольного отношения из R^1.

2. По наблюдаемому отрезку процесса Ӿ построить частичное отношение $R_{Ӿ}$.

3. Определить $|R_{Ӿ}|$ и по формуле (6) вычислить значение апостериорной вероятности отношения $R_{Ӿ}$.

4. Вычислить значение $q = P(R_{Ӿ} = R_{x} \mid R_{Ӿ}, T) \mid Q(R)$.

5. Если значение q достаточно велико (с точки зрения пользователя), то построить ОМВ $MV(R_{Ӿ})$ в соответствии с аппаратом, описанным в [2]. Работу алгоритма прекратить.

6. Если значение q мало (с точки зрения пользователя), то продолжить наблюдение процесса x (при наличии такой возможности) и выполнить алгоритм с п.2. Если возможность дальнейшего наблюдения процесса отсутствует, работу алгоритма прекратить.

Таким образом, сформулирована задача байесовой оценки гипотезы о синхронной обобщенной модели взаимодействия между переменными процесса и предложен алгоритм формирования байесовой оценки гипотезы.

1. *Коломиец Г.Ф., Лаврикова Е.И., Синяков М.Н.* Вопросы постановки задач причинно-следственного анализа // Засоби комп’ютерної техніки з віртуальними функціями і нові інформаційні технології. – К.: Ін-т кібернетики ім. В.М. Глушкова НАН України, 2002. – 1. – С.79 – 86.
2. *Математический* аппарат качественного анализа многомерных процессов / Н.Н. Дидук, В.Н. Коваль, Г.Ф. Коломиец и др. // Нові комп’ютерні засоби, обчислювальні машини та мережі. – К.: Ін-т кібернетики ім. В.М. Глушкова НАН України, 2001. – 1. – С. 81 – 92.

Получено 01. 07. 2002