基于证据推理的显著误差检测方法 1)

王希若 荣冈
（浙江大学工业控制技术国家重点实验室，杭州 310027）

摘　要　提出一种基于证据推理的显著误差检测法，以约束残差作为基本信度分布，仿真实例表明，该方法切实可行.

关键词　证据推理，显著误差检测，数据校正

1 引言

计算机综合自动化系统中，直接反映生产状况的测量数据不可避免地带有各种误差，其中，显著误差给生产信息带来的“污染”特别严重. 作为数据校正的任务之一，必须对此进行检测并剔除. 已有的显著误差侦破和识别方法，如整体检验法、约束检验法等[3] [4]，虽然具有优越的误差侦破功能，但这些基于概率统计的方法往往需要对仪表精度等的先验概率有一个精确的了解. 证据理论[1]的首要出发点就是命题的置信度取遍 [0, 1] 之间的所有值，可以在精确的先验概率无法获得的情况下，根据专家或经验提供的证据对问题作出判断，目前已应用于数据融合和模式识别[2]. 本文提出一种基于证据推理的显著误差检测法，以约束残差作为基本信度分布，仿真实例表明，该方法是切实可行的.

2 方法简介

显著误差检测问题从证据决策的角度来看，就是针对问题“哪个流量含有显著误差” 收集证据并作出决策. 检测法的具体步骤如下，

(i) 建立识别框架

识别框架是某个问题所有可能答案的集合，设Θ是某个问题的可能答案的有限集，且设其中有且只有一个是正确的，此时称Θ为识别框架，或简称框架.

1) 国家 863 CIMS 主题基础研究课题(863-511-945-006)

例如，设在一个含有 6 条支路的流量系统中，有且仅有一个流量含有显著误差，则对“那一个流量含有显著误差”这个问题的识别框架可以选为Θ={流量 1，流量 2，流量 3，流量 4，流量 5，流量 6}．建立了识别框架后，感兴趣的命题都对应于Θ的一个子集．例如，“流量 1 含有显著误差”对应集合 A={流量 1}$\in 2^{\Theta}$，其中 2^{Θ}是Θ的幂集．

(ii) 收集证据

显著误差检测的任务就是要确定形如“流量 1 含有显著误差”，“流量 2 含有显著误差”等命题为真的程度．由日常经验可知，判断命题为真的程度应该有根据，这些根据可以是某条信息，也可以是某种知识或经验，在此统称为证据．

与“哪个流量含有显著误差”这一问题相关的证据很多，如测量值、结点的约束残差、流量计的维护和检修情况等．此外，还有一部分很重要的信息来自专家的经验．为叙述简明起见，取约束残差作为证据．

(iii) 求取各证据的基本信度分布

证据对各命题提供支持或反对的信息，每个证据对各命题的不同支持程度通过它的基本信度分布反映出来．

定义 1：设Θ为识别框架，如果集函数 $m: 2^{\Theta}\to[0,1]$ 满足[1]

① $m(\varnothing)=0$

② $\sum_{A\subseteq\Theta} m(A)=1$

则称 m 为框架Θ上的基本信度分配．$\forall A\subseteq\Theta$，$m(A)$ 称为 A 的基本信度值．

如果 $m(A)>0$，则称 A 为基本信度分配 m 的焦元（focal element），所有焦元的并称为它的核心（core），记为 C．

定义 2：设Θ是一个识别框架．如果集函数 $Bel: 2^{\Theta}\to[0,1]$ 满足

① $Bel(\varnothing)=0$ (1)

② $Bel(\Theta)=1$ (2)

③ $\forall A_1, A_2, \ldots, A_n\subseteq\Theta$（n 为任意自然数）

$$Bel(\bigcup_{i=1}^{n}A_i)\geq\sum_{i=1}^{n}Bel(A_i)-\sum_{i<j}Bel(A_i\cap A_j)+\cdots+(-1)^{n+1}Bel(\bigcap_{i=1}^{n}A_i)$$
$$=\sum_{I\subseteq\{1,2,\ldots,n\},I\neq\varnothing}(-1)^{|I|+1}Bel(\bigcap_{i\in I}A_i) \quad (3)$$

则称 Bel 为框架Θ上的信度函数．

(iv) 合成各证据

定义 3（Dempster 合成法则）：设 Bel_1 和 Bel_2 是同一识别框架Θ上的两个信度函数，m_1 和 m_2 分别是其对应的基本信度分配，焦元分别为 $A_1, A_2, \ldots, A_k$，

和 B_1, B_2, ..., B_l, 设

$$\sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j) < 1 \tag{4}$$

那么由下式定义的函数 m: $2^\Theta \rightarrow [0, 1]$ 是基本信度分布.

$$m(A) = \begin{cases} 0 & A = \varnothing \\ \dfrac{\sum\limits_{A_i \cap B_i = A} m_1(A_i) m_2(B_j)}{1 - \sum\limits_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j)} & A \neq \varnothing \end{cases} \tag{5}$$

基本信度分配 m 称为 m_1 和 m_2 的直和，记为 $m_1 \oplus m_2$. 所对应的信度函数也称为 Bel_1 和 Bel_2 的直和，同样记为 $Bel_1 \oplus Bel_2$.

(v) 作出决策

利用证据表示函数中所包含的信息作决策，即对问题“哪个流量含有显著误差”作出回答. 其中较常用的是求单点似真度的方法.

定义 4: 设 Bel: $2^\Theta \rightarrow [0, 1]$ 是识别框架Θ上的一个信度函数. 定义 Dou: $2^\Theta \rightarrow [0, 1]$ 和 Pl: $2^\Theta \rightarrow [0, 1]$ 如下

$$\forall A \subseteq \Theta \quad Dou(A) = Bel(\overline{A}) \tag{6}$$

$$Pl(A) = 1 - Bel(\overline{A}) \tag{7}$$

则称 Dou 为 Bel 的怀疑度函数，Pl 为 Bel 的似真度函数. $\forall A \subseteq \Theta$, Dou (A) 称为 A 的怀疑度，Bel (A) 称为 A 的似真度.

显然，Dou (∅) = 1, Dou (Θ) = 0; Pl (∅) = 0, Pl (Θ) = 1.

定义 5: 设Θ为识别框架，函数 m: $2^\Theta \rightarrow [0, 1]$ 是识别框架Θ上的一个基本信度分配. 函数 Q: $2^\Theta \rightarrow [0, 1]$ 由下式定义

$$\forall A \subseteq \Theta, \quad Q(A) = \sum_{A \subseteq B} m(B) \tag{8}$$

则称 Q 为框架Θ上的众信度函数. $\forall A \subseteq \Theta$, Q (A) 称为 A 的众信度值.

通过简单的推导可以得出，在单点集上似真度与众信度是相同的. 因此，我们在计算单点集时，可以通过求取它的众信度来得到它的似真度.

现以仿真系统为例来说明证据决策理论在显著误差检测问题上的应用.

3 仿真实例

设有如图 1 所示的流量系统，各流量的真值分别为 x1=30, x2=20, x3=10, x4=20, x5=10, x6=30，单位 kg/h. 且其中有且仅有一个流量含有显著误差.

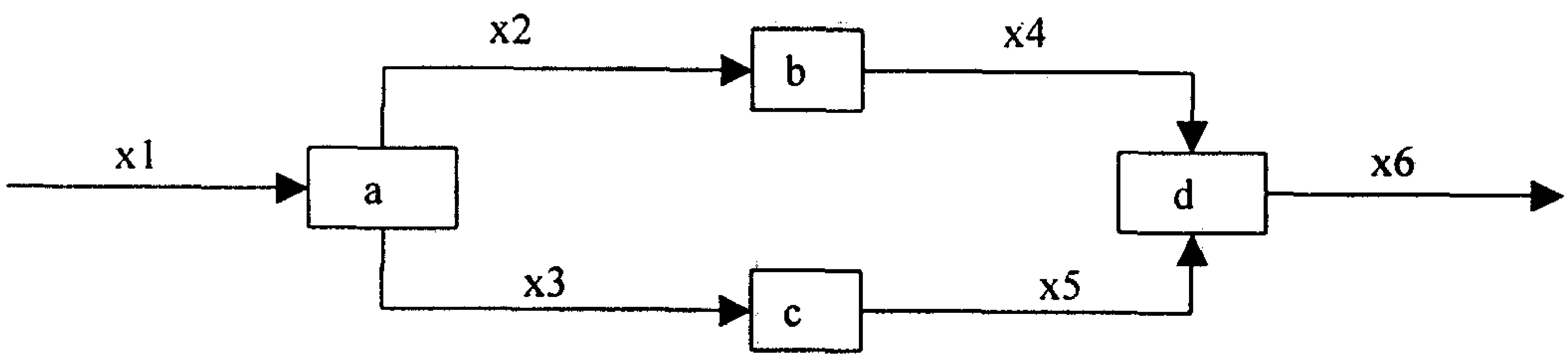

图 1 仿真系统

假设流量 x2 出现显著误差. 所得的测量值为：x1=29.5, x2=25, x3=10.2, x4=9.9, x5=10.1, x6=29.9. 针对该系统，我们可以依据证据决策理论来进行显著误差检测.

首先，建立识别框架. $\Theta = \{x1, x2, x3, x4, x5, x6\}$

在显著误差检测问题中，结点的约束残差是很重要的一个指标. 当与某结点相连的所有流量都不存在显著误差时，该结点的约束残差服从期望为零的正态分布. 因此，此处取各结点的约束残差的绝对值为证据.

例如，结点 a 的约束残差为 5.3，占该结点总入流量的 19%. 根据经验，我们可以以此为该证据的基本信度分布的度量标准，认为该证据以 0.19 的程度支持{x1, x2, x3}中有元素出现显著误差，但不支持Θ中的其他假设，那么证据将把信度 0.19 分配到集合{x1, x2, x3}上，而剩下的信度 1-0.19=0.81 将被分配到Θ上，因此，所得证据 1 的基本信度分配为：$m_1\{x1, x2, x3\}=0.19$； $m_1(\Theta)=0.81$；在Θ的其他子集上的信度分布均为零.

同理，可得其他证据的基本信度分配. 如表 1 所示.

表 1 各证据基本信度分配

证据	结点	约束残差与总流量之比	基本信度分配	在其他子集上信度分布
1	a	19%	$m_1\{x1, x2, x3\}=0.19$； $m_1(\Theta)=0.81$	0
2	b	25%	$m_2\{x2, x4\}=0.25$； $m_2(\Theta)=0.75$	0
3	c	1%	$m_3\{x1, x2, x3\}=0.01$； $m_3(\Theta)=0.99$	0
4	d	1%	$m_4\{x1, x2, x3\}=0.01$， $m_4(\Theta)=0.99$	0

用 Dempster 合成法则（3）（4）对以上证据进行合成，有

$m\{x2\}= 0.0489$； $m\{x1, x2, x3\}= 0.1411$； $m\{x2, x4\}=0.2086$； $m(\Theta)=0.6014$

求得各单点的似真度为

Pl{x1}= 0.7425; Pl{x2}= 1; Pl{x3}= 0.7425

Pl{x4}= 0.81; Pl{x5}= 0.6014; Pl{x6}= 0.6014

如图2所示.

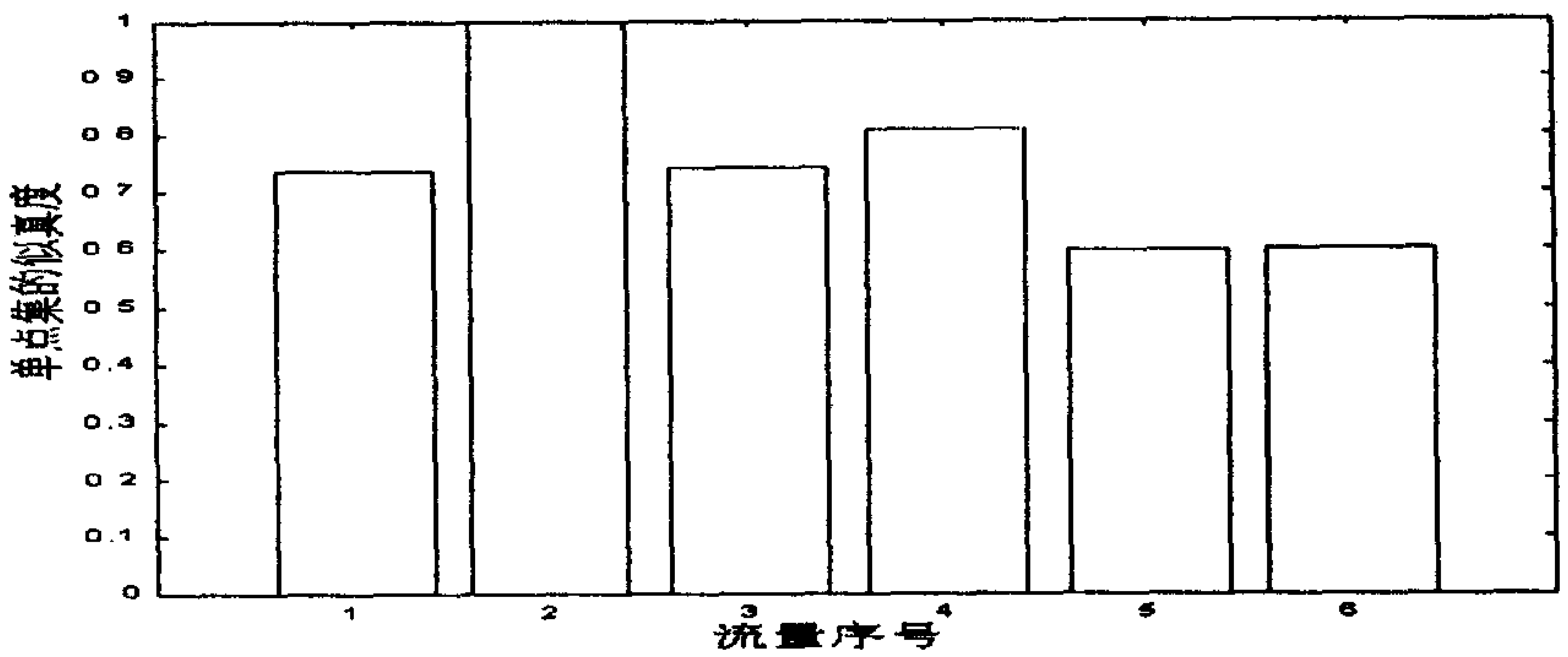

图2 各流量的单点似真度

可作出决策，流量x2含有显著误差．可见与实际情况一致．

4 结论

提出了基于证据决策理论的显著误差检测法．该方法在实际应用中还有以下两点需要进一步讨论．

1）实际系统所包含的物流、结点数目庞大，计算量也较大，必须将其分划为若干个小系统再进行计算，以减小计算量．

2）与其它方法相比，该方法能综合不同性质、来源的信息，包括约束残差、仪表精度等级等，也可将专家经验等信息量化后综合考虑．但此处也涉及到各证据基本信度分配的确定问题，需通过长期操作中经验的积累取得．

参考文献

1. Shafer,G., A Mathematical Theory of Evidence, Princetion University Press, Princeton, NJ,1976

2. 戴冠中等 证据推理的进展及存在的问题， 控制理论与应用，1999, 16(40)

3. 袁永根，李华生, 过程系统测量数据校正技术, 北京：中国石化出版社，1996.

4. 王希若, 荣冈，测量仪表过失误差的单结点识别方法，化工学报，2000，51（1）

王希若 1974年生，硕士． 研究领域为化工过程数据校正及计算机集成生产系统．