doi:10.11887/j.cn.201405029 http://journal.nudt.edu.cn

不确定数据流多维建模方法*

李 明,张维明,刘青宝
(国防科技大学 信息系统工程重点实验室,湖南 长沙 410073)

摘 要:由于不确定数据流应用的出现,给传统的精确、静态数据环境下的多维建模带来了巨大挑战。针对不确定数据流动态、无限和不确定等特征,提出了一种不确定数据流多维模型。该模型中引入了不确定对象来描述不确定事实元组,并且通过定义时间维度的层次时间窗口,很好地反映了数据流的动态性和无限性,最后还对此多维不确定数据流模型的基本代数操作和分析代数操作进行了形式化定义,为不确定数据流多维查询与分析提供了理论依据。

关键词:多维模型;不确定数据流;时间维度;连续查询
中图分类号:TP311 **文献标志码**:A **文章编号**:1001-2486(2014)05-174-06

Multi-dimensional modeling method of uncertain data stream

LI Ming, *ZHANG Weiming*, *LIU Qingbao*
(Key Laboratory of Information System Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Because of the emergence of uncertain data stream applications, the traditional multi-dimensional model based on precise and static data is confronted with some new challenges. A multi-dimensional uncertain data stream model based on characters of uncertain data stream was presented, such as dynamic, infinite, uncertainty and so on. This model introduced uncertain object to depict uncertain facts, and defined the time dimension with multi-level time windows to reflect the dynamic and infinity of data stream. The basic algebraic operations and analytic algebraic operations of the multi-dimensional uncertain data stream was defined formally, which gives theoretical support to multidimensional query and analysis of uncertain data stream were provided.

Key words: multi-dimensional model; uncertain data stream; time dimension; continuous query

数据仓库作为一种重要的智能决策支持工具,被广泛地使用在军事、银行、电力等领域并发挥了重要作用。数据仓库可以简单地视为一个面向主题的、一致的、随时间变化的并且非易失的数据集合,它其中存储了各个异构分布数据源的海量历史数据,用户可以按照分析需要将其组织为多维立方体结构,这样可以辅助用户高效地进行决策分析。多维数据模型是数据仓库的核心,通过这种模型可以使得数据仓库能够提供高效的在线分析查询。但是现在的多维模型均是建立在一个统一的假设之上,即要求事实数据必须是精确的、静态的,这种假设在现实应用中往往无法满足,例如通过传感器采集到的数据则是动态实时产生的非精确数据[1]。正是由于这些不确定数据流应用的出现,使得不确定数据流研究受到了越来越多学者的关注,其基本思想是对传统的数据管理技术进行扩展以满足不确定数据流的需要,目前针对不确定数据流在模型、索引、存储以及查询等方面已经提出了一些方法[2-5],但是对不确定数据流的多维查询和分析的研究还很匮乏,重点从数据仓库多维模型角度来探索不确定数据流在数据仓库中的表示方式,提出了一种新型的多维模型来支持不确定数据流,为不确定数据流的多维建模提供了一定的理论支持。

1 相关工作比较

多维数据模型最早由 Codd 在 1993 年提出[6],在此之后,大量学者对此问题进行了研究,提出了维度、度量、维度层次等多维模型的基本概念,这为多维分析技术奠定了理论基础[7-10]。然而这些传统多维数据模型均要求事实数据是静态的、精确的,随着新型应用的出现,尤其是传感器数据大量的收集,多维建模也开始扩展到更多的数据类型,韩家炜在文献[11]中对数据流立方体

* **收稿日期**:2014-04-08
基金项目:国家自然科学基金资助项目(70771110)
作者简介:李明(1985—),男,河南洛阳人,博士研究生,E-mail:mingli1985@nudt.edu.cn;
张维明(通信作者),男,教授,博士,博士生导师,E-mail:zhangweiming@nudt.edu.cn

进行了完整的研究,包括其计算及查询技术,但并未给出严密的数据流多维模型描述。侯东风在文献[12]中针对数据流的多维建模给出了形式化的数学表述,其模型可以很好地描述数据流的动态性和无限性。刘青宝在文献[13]中提出了一种模糊动态多维建模方法,通过模糊维度来描述维层次之间的不确定划分,同时模型也可以支持数据的动态性,然而他们均未考虑度量数据的不确定问题。借鉴现有关于不确定数据流的研究成果,将不确定数据流引入多维建模描述,并通过在时间维度上定义多层次时间窗口体现了数据流的动态性和无限性,最后给出了不确定数据流多维模型的操作代数。

2　不确定数据流多维分析基础及应用

不确定数据流通常表现为实时到达的不确定对象序列,在现有的研究中多对其定义如下[14-16]:

定义 1　不确定数据流是由不确定对象构成无限序列集,表示为 $S=\{O_{t_1},O_{t_2},\cdots\}$,其中 O_{t_i} 为不确定对象,它表示在时刻 t_i,不确定数据流 S 的取值为 $S[t_i]=O_{t_i}=\{o_1^{t_i},o_2^{t_i},\cdots,o_n^{t_i}\}$,$o_j^{t_i}$ 为 O_{t_i} 的实例,每个 $o_j^{t_i}\in O_{t_i}$ 均有一个概率 $\Pr(o_j^{t_i})>0$ 与之对应,并且满足 $\sum_{j=1}^{n}\Pr(o_j^{t_i})=1$。

不确定世界语义模型是研究不确定数据的基础模型,由于时间域上的无限性,不确定数据流的语义表示则需要时间窗口约束。

定义 2　基于滑动窗口的不确定数据流可能世界模型。令 $S=\{O_{t_1},O_{t_2},\cdots\}$ 为一个不确定数据流,其在滑动时间窗口 $W_\omega^{t_i}$ 的可能世界 $w=\{v_{i-\omega},\cdots,v_{i-1},v_i\}$ 为一组实例的集合,其中 $v_j\in O_{t_j}, i-\omega\leqslant j\leqslant i$,对于可能世界 w 而言其包含的元组数为 $|w|=\omega$,出现概率为 $\Pr(w)=\prod_{j=i-\omega}^{i}\Pr(v_j)$。不确定数据流 S 在滑动时间窗口 $W_\omega^{t_i}$ 上的所有可能世界集合表示为 $\Omega(W_\omega^{t_i}(S))$。

例 1　对于火灾预警的多点温度监测应用,为了提升监测精度,往往会在同一地点布置多个传感器,由于物理误差和多源监测,数据中常常含有不确定性。每个温度监测器在特定时间源源不断地返回监测温度值,而在同一位置可以部署多个温度传感器,每个传感器带有一个可信度属性来标示其可靠程度,其关系模式可表示为 S(时间, 位置, 传感器, 温度, 可信度),其对应的数据流如表 1 所示。

表 1　不确定数据流示例

Tab. 1　Examples of uncertain data stream

时间	位置	传感器	温度(℃)	可信度
11:00:05	W_1	S_1	50	0.5
11:00:05	W_1	S_2	40	0.5
11:00:05	W_1	S_3	44	0.8
11:00:05	W_1	S_4	37	0.2
11:10:10	W_3	S_5	23	0.3
11:10:10	W_3	S_6	32	0.7
11:10:10	W_4	S_7	53	0.4
11:10:10	W_4	S_8	42	0.6
⋮	⋮	⋮	⋮	⋮

上述例子属于一个典型的不确定数据流应用,用户往往希望从更高层次对这些数据进行异常检测处理,例如一段时间内的最高温度等信息,这就有必要对其进行多维数据分析。对于例 1 的多维数据模式可以表示为 S-cube(时间, 位置, 温度),维度包含时间维和位置维,度量包含温度值。通过分析可以发现不确定数据流的多维分析对多维模型提出了新的需求,具体如下:

1)时间维度为默认维度不可缺失,并且通过时间维度可以反映数据的动态特性;

2)度量值中包含不确定数据,需要对其重新定义多维数据操作的语义;

3)支持持续的数据查询,其查询的结果常表现为流式信息。

3　不确定数据流的多维数据模型

在本节中重点介绍不确定数据流的多维数据模型,首先根据不确定数据流在时间上具有无限性和动态性特征来定义带有滑动时间窗口的时间维度模型,然后提出了不确定数据流度量的模型表示,基于以上两个模型进一步给出了不确定数据流的多维数据模型。

3.1　带有滑动时间窗口的时间维模型

对于不确定数据流的多维数据模型而言,时间维必须给出明确定义,这也是体现其动态性和无限性的重要基础,所以需要对其时间维度单独进行研究。

定义 3　时间维模式为 $\psi_t=\{L_t,\leq_t,\perp_t,All_t\}$,其中:

1)L_t 表示时间级别集合,$L_t=\{l_i\mid i=1,2,\cdots,m\}$,其中 l_i 表示为第 i 个层次级别;

2)$\leq_t$ 表示在时间维度级别之间的偏序关系,对于 $\forall l_i,l_j\in L_t, l_i\neq l_j$,若存在 $l_i\leq_t l_j$,则称 l_j 的层次级别粗于 l_i 的层次级别,反之则称 l_j 的层次

级别细于 l_i 的层次级别。若满足$\nexists l_k \in L_t$ 使得 $l_i \leq_t l_k \leq_t l_j$,则称 l_i, l_j 是相邻层次级别;

3) $\perp_t$ 表示时间维度上的基本层次级别,$\perp_t \in L_t$,且满足 $\forall l_i \in L_t$,均有 $\perp_t \leq_t l_i$;

4) All_t 表示时间维度上的最高层次级别,$All_t \in L_t$,且满足 $\forall l_i \in L_t$,均有 $l_i \leq_t All_t$。

对于不确定数据流而言,时间维模式中的最高层次级别 All_t 并不表示所有时间全域,因为其全域具有无限性,所以此处的 All_t 仅表示在一个时间窗口内的时间域。

对于例1而言,由于数据是快速变化的且返回的周期相对较短,所以粗时间粒度的定义并无实际意义,于是将其定义为 minute→quarter→hour,最细粒度为 minute,最粗粒度为 hour。

定义4 时间维度是时间维模式的一个实例,它可以表示为 $D_t = \{L_t, \psi_t, W_\omega^{t_i}\}$,其中:

1) L_t 表示时间维度上的层次集合,$L_t = \{l_i \mid i = 1,2,\cdots,m\}$,其中 $l_i = \{C_{i1}, C_{i2}, \cdots, C_{in}\}$,$C_{ik}$表示时间维第 i 个层次级别中的第 k 个成员,$l_\perp$ 中的所有成员称之为时间维的基本域。

2) ψ_t 表示维层次级别的映射关系,即 $\psi_t = \{\varphi_{ij}\}$,若 $l_i, l_j \in L_t$ 且满足 $l_i \leq_t l_j$,则二者之间存在映射关系 $\varphi_{ij}: 2^{l_i} \to l_j$,对于 $\forall C_{jk} \in l_j$,总是存在一个子集 $Z \subseteq l_i$,使得 $C_{jk} = \varphi_{ij}(Z)$。若 l_i 和 l_j 是相邻的,那么 φ_{ij}为直接映射关系。

3) $W_\omega^{t_i}$ 表示时间维度上的滑动时间窗口,其中 t_i 为时间域的起始时刻,ω 为时间窗口的大小,通过 $W_\omega^{t_i}$ 约束了整个时间域的取值范围$dom(D_t)$。

对于不确定数据流的时间维度建模最重要的就是引入滑动时间窗口,这样就可以将时间维度的取值限定在一定的范围之内,很好地规避了不确定数据流在时间域上潜在的无限性,伴随着时间往前推移,滑动时间窗口也不断地向前滚动,所以时间域上的取值范围是随时间不断变化的。

普通维度 D_i 由于其取值域固定,所以可以表示为清晰的层次结构,$l_i^\perp$ 表示维度 D_i 的基本域集合,l_i^{All} 表示顶层概念层次成员集合。由于普通维度的定义和一般多维数据模型一致,所以此处不再赘述。

3.2 度量模型

在不确定数据流中,由第2节所述可知,它在任意时刻 t_i 可以视为一个不确定对象,这也就是说在不确定数据流的多维模型中度量取值不再是一个确定的数值,而是一个随机变量。

定义5 度量可以表示为集合 $M = \{M_j \mid j = 1,2,..,q\}$。其中 M_j 表示度量属性,其实例对应一个不确定数据流,即 M_j 的基本取值域表示为 $dom(M_j) = \{O_1^j, O_2^j, \cdots\}$,$O_i^j$ 表示不确定对象。

例1中对应的度量为温度值,它的取值域为每个时刻每个位置对应的可能温度取值。

定义6 事实模式定义了维度和度量之间的映射关系,表示为 $F = \{\psi, M\}$,其中 $\psi = \{\psi_1, \cdots, \psi_n, \psi_t\}$ 为维度模式,M 为度量属性集合。

令不确定数据流的事实集合为 I,可以将其划分为基本事实 I_B 和聚集事实 I_A。

定义7 不确定数据流基本事实表示为 $f_B = \{a_1, \cdots, a_n, a_t, m_1, \cdots, m_q\}$,其中 a_i 表示维度 D_i 基本层次级别的成员,$a_i \in l_\perp^i$ $(1 \leq i \leq n)$;a_t 表示时间维度基本层次级别的成员,$a_t \in l_\perp^t$;m_j 表示 M_j 的一个取值,$m_j \in dom(M_j)$。

定义8 等价事实。对于 $\forall f_i, f_j \in I, D_k(f)$ 表示 f 在维度 D_k 上的取值,若满足 $D_k(f_i) = D_k(f_j)$,其中 $k \in [1,n] \cup t$,那么则称 f_i 与 f_j 为等价事实。

定义9 不确定数据流聚集事实。表示为 $f_A = \{a_1', \cdots, a_n', a_t', m_1', \cdots, m_q'\}$,其中 $\exists a_i \in l_j^i, l_\perp^i \leq_t l_j^i (j \in [1,n] \cup t), m_j' = Agg_{m_j}(I_B(f_A))$,$I_B(f_A)$表示一组基本事实,它们在聚集事实 f_A 的维度取值上等价,Agg_{m_j}表示对基本事实集合在度量属性 m_j 上的聚合函数。

例1中当前时间窗口的事实集合可表示为表2。

表2 当前窗口的事实集合

Tab.2 Facts in current time window

ID	时间	位置	温度(℃)
f_1	11:00:05	W_1	{50, 0.5}{40, 0.5}
f_2	11:00:05	W_2	{44, 0.8}{37, 0.2}
f_3	11:10:10	W_3	{23, 0.3}{32, 0.7}
f_4	11:10:10	W_4	{53, 0.4}{42, 0.6}
⋮	⋮	⋮	⋮

表2中均为基本事实,其中温度度量用不确定对象来表示,对于时间维而言,事实 f_1 和 f_2 以及 f_3 和 f_4 分别为等价事实,沿着时间维度或位置维度可以得到对应的聚合事实。

3.3 多维模型

根据前面两个小节提出的维度模型和度量模型,可以进一步给出不确定数据流的多维数据模型。

定义10 不确定数据流多维模型表示为一

个四元组 $UMDS=\{\psi,M,F,\alpha\}$,其中:

1)$\psi=\{\psi_1,\cdots,\psi_n,\psi_t\}$ 为一个维度模式集合,$\psi_1,\cdots,\psi_n$ 表示普通维度模式,ψ_t 表示时间维度模式;

2)$M=\{M_j\mid 1\leqslant j\leqslant q\}$ 为度量属性集合;

3)$F=\{\psi,M\}$ 为不确定数据流事实模式;

4)$\alpha=\{A_{i,j}\mid i=1,\cdots n,t;\ j=1,\cdots,q\}$,其中 $A_{i,j}$ 表示在维度 D_i 上依据维成员之间的依赖关系在度量 M_j 上进行的聚集计算。

定义 11　不确定数据流多维实例表示为三元组 $UMDI=\{D,M,I\}$,其中 D 表示维度实例集合,M 表示度量属性集合,I 表示不确定数据流事实集合。

在例 1 中所有事实集合就构成了不确定数据流多维实例。

4　不确定数据流多维代数操作

本节重点讨论不确定数据流多维模型的代数操作,以实现对多维不确定数据流的分析。对于不确定数据流而言,其具有两个显著特征,一是时间域上无限性,二是度量取值是不确定对象,所以需要根据这两种特征来扩展确定数据的基本操作和分析操作。

4.1　基本代数操作

不确定数据流多维数据的基本代数操作类似于关系数据的代数操作,包括选择、投影和聚集操作。

定义 12　选择操作是对不确定数据流事实按照一定约束规则进行的筛选操作,最终返回一组满足条件的事实集合。选择操作可以定义在维度上亦可在度量上,同时选择操作可以在维度的各个级别进行约束。选择操作可以表示为 $\sigma_p(UMDI)=UMDI'=(D',M',I')$,其中:

1)$D'=D$,其中时间维度 D_t 的取值域为当前时间窗口 $W_\omega^{t_i}$;

2)$M'=M$;

3)$\forall f\in I,f=\{a_1,\cdots,a_n,a_t,m_1,\cdots,m_q\}$,若 p 是关于维度上的选择操作,$p_D(a_1,\cdots,a_n,a_t)=\text{true}$,则 $f\in I'$;若 p 是关于度量上的选择操作,不失一般性,我们假设 p_M 仅对 m_1 进行约束,则 $p_M(m_1)=\{o\mid p_M(o)=\text{true}\wedge o\in m_1\}=m_1'$,令 $f'=(a_1,\cdots,a_n,a_t,m_1',\cdots,m_q)$,只要 m_1' 中实例数不为 0,则 $f'\in I'$。

若在例 1 查询温度值高于 40℃ 的记录,即 p_M(温度)≥40℃,那么按照定义查询的结果为 f_1,f_2,f_4,其中 f_2 出现的概率为 0.8。

对于不确定数据流而言,选择操作输入是随时间变化的实例流,输出结果同样是一组满足约束条件的元组流,我们将其称之为持续选择操作,$\sigma_p(UMDI)=\bigcup_t[UMDI'(t_i)-UMDI'(t_{i-1})]\cup UMDI'(t_0)$,其中 $UMDI'(t_i)$ 表示在时刻 t_i 的选择结果,这样就可以随着时间的推移不断更新结果集。

定义 13　投影操作是对多维不确定数据流在维度属性和度量属性上的筛选操作,其输入为当前时间窗口的多维不确定数据流实例,输出为投影操作后的多维数据流实例。投影操作可以表示为 $\pi_q(UMDI)=UMDI'=(D',M',I')$,其中:

1)$D'=D_q$,D_q 是 D 的一个维度子集,表示对原始多维不确定数据流的投影;若 $D_t\in D_q$,则 $dom(D_t)$ 变为限定范围的有限时间维度;

2)$M'=M_q$;

3)$I'\in\{f'\}$,$\forall f\in I$ 在维度投影 D_q 上可以得到若干等价事实集合 I_E,f' 则与等价事实集合 I_E 相对应,即 $f'=Agg(f)$,$f\in I_E$。

若将表 2 中的多维数据实例投影到时间维度上,即 $D_t=D_q$,此时 f_1 与 f_2 以及 f_2 与 f_3 分别形成一个等价事实集合,它们分别对应新的事实记录 f_1' 和 f_2',其聚合计算则在定义 14 中给出了描述。

对不确定多维数据流的持续投影操作可表示为 $\pi_q(UMDI)=\bigcup_t[UMDI'(t_i)-UMDI'(t_{i-1})]\cup UMDI'(t_0)$,其中 $UMDI'(t_i)$ 表示在时刻 t_i 的投影结果,显然,在持续投影操作中必须包含时间维度。

定义 14　聚合操作是将聚合函数应用到多维实例,得到相应的聚合结果。其操作是输入为当前时间窗口的多维不确定数据流实例,输出为聚合计算后得到的多维数据实例。设 φ_A 为聚合函数,l_A 为聚合的维度层次级别,M_A 为聚合的度量属性,聚合操作可以表示为 $Agg[\varphi_A,l_A,M_A](UMDI)=UMDI'=(D',M',I')$,其中:

1)$D'=D$,其中 $dom(D_t')$ 变为限定范围的有限时间维度;

2)$M'=M\cup\{M_A\}$;

3)根据等价事实的定义,可以得到 f_B 在 l_A 上的等价事实集合,换句话说,f' 对应一个由 f_B 构成的等价事实集合 I_E,且 $f'.M_A'=\varphi_A(I_E.M_A)$,由于 $\forall f_B\in I_E$,$f_B.M_A$ 对应一个不确定对象,所以

$I_E.M_A$ 对应一个不确定对象集合，可以表示为 $I_E.M_A=\cup w_i$，w_i 表示 $I_E.M_A$ 的可能世界实例，由此 $f'.M_A'=\varphi_A(w_i)$。

在例1中如果沿着位置维向上进行最大值聚集，则可以得到一组新的聚集事实，如表3所示。

表3 当前时间窗口内沿位置维的聚合事实集合

Tab. 3 Facts aggregated by location dimension in current time window

ID	时间	温度(℃)
f_1'	11:00:05	{50, 0.5}{44, 0.4}{40, 0.1}
f_2'	11:10:10	{53, 0.4}{42, 0.6}
⋮	⋮	⋮

对于不确定多维数据流持续聚合操作可以表示为：$Agg[\varphi_A,l_A,M_A](UMDI)=\bigcup_t[UMDI'(t_i)-UMDI'(t_{i-1})]\cup UMDI'(t_0)$，其中 $\bigcup_t[UMDI'(t_i)-UMDI'(t_{i-1})]$ 表示随时间变化的增量部分。

4.2 分析代数操作

不确定数据流的多维模型分析操作与传统的多维模型类似，包含了切片、切块、上钻及下钻操作。

定义15 切片操作是对某个维度的取值进行的等值约束，可以用选择和投影操作组合完成。假设切片维度为 D_k，切片的约束条件表示为 $p_k:d_k=a$，令 $D_r=D-D_k$，$q=(D_r,M)$，多维不确定数据流切片可表示为 $Slice(UMDI,D_k,p_k)=\pi_q[\sigma_{p_k}(UMDI)]$。

定义16 切块操作是对某个维度取值进行的范围约束，其约束条件可以表示为 $p_k:a\leqslant d_k\leqslant b$，令 $D_r=D$，$q=(D_r,M)$，其中 $dom(D_k)$ 变为约束条件 p_k 给定的范围，多维不确定数据源流切块操作可表示为 $Dice(UMDI,D_k,p_k)=\pi_q[\sigma_{p_k}(UMDI)]$。

在例1中若要求时间维度限定在 11:00:00 – 11:05:00 区间内，即 p_k: 11:00:00 ≤ d_t ≤ 11:05:00，那么对应的事实记录则变为 f_1 和 f_2。

持续切片(块)操作的结果可以表示为 $UMDI'=\bigcup_{t_i}[UMDI'(t_i)-UMDI'(t_{i-1})]\cup UMDI'(0)$，其中 $UMDI'(t_i)$ 为在时刻 t_i 的切片(块)。

定义17 流上钻操作完成多维不确定数据流的聚集运算，以便用户可在不同粒度上查看数据中隐含的信息和趋势。令上钻操作的维度为 D_k，聚集目标层次 $l_k^i=\{C_k^{i1},\cdots,C_k^{im}\}$，对应上钻操作可以表示为 $RollUp(UMDI,D_k,l_k^i)=Agg[\varphi_A,l_k^i,M_A](UMDI)$。

持续的上钻操作可以表示为 $UMDI'=\bigcup_{t_i}[UMDI'(t_i)-UMDI'(t_{i-1})]\cup UMDI'(0)$，其中 $UMDI'(t_i)$ 为时刻 t_i 的上钻操作结果。

若 $D_k=D_t$，则称为沿着时间维的上钻。

下钻操作等同于上钻操作的逆过程，对应的是将基本事实聚集到相对较低的维层次上，这里不再重复定义。

定理1 以上各种分析操作对于多维不确定数据流实例封闭。

证明：通过对分析操作的定义可以看出，其操作结果仍然为多维不确定数据流实例，所以其封闭性是显然的。

5 结论

提出了一种多维不确定数据流的形式化描述模型，在此模型中引入了不确定对象来进行描述数据中出现的不确定性，同时提出了基于时间窗口的时间维度模型来对时间域上的无限性进行描述，并在此基础上给出了多维不确定数据流模型中的流事实、流多维模型，流多维数据实例等基本概念的定义。最后根据多维不确定数据流的基本特征，定义了支持不确定数据流的多维分析基本操作代数和分析操作代数。在未来的工作中还需进一步针对多维不确定数据流的基于约束的清洗、聚集计算优化以及多维查询等问题展开更为深入的研究。

参考文献(References)

[1] Cuzzocrea A. Data warehousing and knowledge discovery from sensors and streams[J]. Knowledge and Information Systems, 2011, 28(3):491 – 493.

[2] 周傲英,金澈清,王国仁,等. 不确定性数据管理技术研究综述[J]. 计算机学报,2009,32(1):1 – 16.
ZHOU Aoying, JIN Cheqing, WANG Guoren, et al. A survey on the management of uncertain data[J]. Journal of Computer, 2009, 32(1): 1 – 16. (in Chinese)

[3] 蒋涛,高云君,张彬,等. 不确定数据查询处理[J]. 电子学报. 2013,41(5): 966 – 976.
JIANG Tao, GAO Yunjun, ZHANG Bin, et al. Query processing on uncertian data[J]. ACTA Electronica Sinica. 2013, 41(5): 966 – 976. (in Chinese)

[4] 王意洁,李小勇,祁亚斐,等. 不确定数据查询技术研究[J]. 计算机研究与发展,2013,49(7):1460 – 1466.
WANG Yijie, LI Xiaoyong, QI Yafei, et al. Uncertain data query technologies[J]. Journal of Computer Research and Development, 2013, 49(7): 1460 – 1466. (in Chinese)

[5] Suciu D, Olteanu D, Ré C, et al. Probabilistic databases[M]. US: Morgan & Claypool Publishers, 2011.

[6] Codd E F. Providing OLAP to user-analysts: an IT mandate[R]. Technical Report, TR-9300011, 1993.

[7] Arfaoui N, Akaichi J. Data warehouse: conceptual and logical schema survey [J]. International Journal of Enterprise Computing and Business Systems, 2012, 2(1), 1-31.

[8] 李建中, 高宏. 一种数据仓库的多维数据模型[J]. 软件学报,2000,11(7):908-917.
LI Jianzhong, GAO Hong. Multidimensional data modeling for data warehouses[J]. Journal of Software, 2000, 11(7): 908-917. (in Chinese)

[9] Pedersen T B, Jensen C S, Dyreson C E. A foundation for capturing and querying complex multidimensional data[J]. Information Systems,2001, 26(5): 383-423.

[10] 陆昌辉. 复杂多维数据模型的描述、构建与查询处理方法研究[D]. 长沙:国防科学技术大学,2006.
LU Changhui. Complex multidimensional data model description[D]. Changsha: National University of Defense Technology, 2006. (in Chinese)

[11] Han J W, Chen Y X, Dong G Z, et al. Stream cube: an architecture for multi-dimensional analysis of data streams[J]. Distributed and Parallel Databases, 2005, 18(2): 173-197.

[12] 侯东风. 流式数据多维建模与查询关键技术研究[D]. 长沙:国防科学技术大学,2010.
HOU Dongfeng. Research on key issues of data stream multi-dimensional modeling and querying[D]. Changsha: National University of Defence Technology, 2010. (in Chinese)

[13] 刘青宝. 模糊、动态多维数据建模理论与方法研究[D]. 长沙:国防科学技术大学,2006.
LIU Qingbao. Theory and method of fuzzy, dynamic multi-dimensional data modeling [D]. Changsha: National University of Defence Technology, 2006. (in Chinese)

[14] Pei J, Jiang B, Lin X M, et al. Probablistic skylines on uncertain data[C]//Proceedings of ACM VLDB, 2007: 15-26.

[15] Hua M, Pei J. Continuously monitoring top-k uncertain data stream: a probabilistic threshold method[J]. Distribute and Parallel Databases,2009, 26(1): 29-65.

[16] Jin C Q, Chen L, Yu J X, et al. Sliding-window top-k queries on uncertain streams[J]. The VLDB Journal,2010, 19(3): 411-435.