

基于异构双极信息的模糊查询研究

赵法信 金义富

(湛江师范学院智能信息处理研究所 湛江 524048)

摘要 在日常生活中,人们在表达其需求时,常常会对同一事物同时使用正面的描述和负面的描述。由于不要求正、负面的描述必须相互对称,此类具有异构语义的信息被称为异构双极信息。传统信息系统中的模糊查询方法,不能很好地处理用户查询需求中所含有的异构双极信息。基于经典数据库,以在模糊查询中能够处理异构双极信息为目标,将 Vague 集引入到异构双极信息的建模,给出了一个由独立的满意度和不满意度所组成的异构双极信息查询满意度模型框架,并在此基础上讨论了包含正面、负面信息的异构双极信息查询处理方法。

关键词 经典数据库,异构双极信息,Vague 集,双极查询满意度,模糊查询

中图分类号 TP311 文献标识码 A

Study on Fuzzy Query of Heterogeneous Bipolarity Information

ZHAO Fa-xin JIN Yi-fu

(Institute of Intelligent Information Processing, Zhanjiang Normal University, Zhanjiang 524048, China)

Abstract In daily life, people often give both positive and negative information to state what they desire and what they reject for the same things. Because positive and negative statements do not necessarily mirror each other, this results in so-called heterogeneous bipolar information. The fuzzy query in traditional information systems does not adequately support the handling of heterogeneous bipolar information. In this paper, based on the regular database, vague set was introduced into the modeling of heterogeneous bipolar information for dealing with heterogeneous bipolar information in fuzzy queries, and a bipolar query satisfaction modelling framework which is based on couples that consist of an independent degree of satisfaction and degree of dissatisfaction was given, then the processing of heterogeneous bipolar queries that contain both positive and negative criteria was also discussed.

Keywords Regular databases, Heterogeneous bipolar information, Vague set, Bipolar query satisfaction degrees, Fuzzy querying

1 引言

在现实世界中,人们常常会使用既包含正面信息又包含负面信息的语言来描述事物或表达自己的意愿。正面的信息主要是指包含可能性、满意度、期望值等自己认为可以接受的信息。负面的信息则主要是指包含不可能性、被拒绝或被禁止的信息。这里所说的双极(bipolar)就是指正、负面信息同时存在的情况^[1,2]。在某些情况下,正、负面信息具有明确的对称语义,因此彼此之间可以相互推理。例如,如果一个用户想买一部智能手机,显然非智能手机不会是该用户的考察目标。然而在现实中,也存在一些不能用类似的方法进行推理的情况,例如,一个用户声明他不喜欢白色的手机,但并不能据此推理出该用户同等地喜欢其他所有颜色的手机,即使已知该用户喜欢红色和蓝色的手机,但由于不可能列举出所有的颜色,所获得的信息仍然不够完整,因此很难获知该用户所喜欢手机的所有颜色,这一点恐怕用户自己也很难说全面了。此时,用户所表达的正、负面信息不相互对称,即该双极性信息具有异构的语义,这种具有异构语义的信息被称为异构双

极信息^[3],该类信息也普遍存在于我们的日常生活中。

多年来,面向经典数据库的柔性查询一直是数据库领域的热门研究方向之一。模糊查询^[4]作为柔性数据库查询的主要内容,其柔性查询技术是以模糊集^[5]为基础的。在模糊查询过程中,通常会根据查询结果满足查询条件的程度(称之为满意度)对包含在数据库中的查询结果进行排序。而满意度的取值范围为 $[0, 1]$,其中1表示完全满足查询条件,0表示完全不满足查询条件。如果一个记录满足查询条件的度为 d ,那么该记录不满足查询条件的度为 $1-d$ 。显然,模糊集所支持的单隶属度(用于表示查询满意度)决定了经典模糊数据库查询系统不能很好地处理异构双极信息。

目前,虽然在信息系统中使用双极性信息的研究还比较少,但对双极性信息的处理方法已经引起了越来越多学者的关注,并逐步成为一个热门的研究方向^[6,7]。文献^[8]是最早涉及双极性信息查询研究的成果,该文首次将查询条件分为强制性(mandatory)查询条件和偏好(preferences)查询条件两类。强制条件的反面指定了必须被拒绝的东西,因此,可将其视为关于查询条件的负面信息。而偏好条件所指定的东西

到稿日期:2012-09-16 返修日期:2013-01-03 本文受广东省自然科学基金项目(S2012010010438)资助。

赵法信(1974—),男,博士,高级工程师,主要研究方向为智能数据处理、数据挖掘,E-mail:zfx0405@163.com;金义富(1969—),男,博士,教授,主要研究方向为智能信息处理与数据挖掘。

可被视为正面信息。随后,这种思想被融合到模糊查询技术。在基于双向模糊集^[2]的模糊查询方法中,用户可以在一个基本的查询条件中指定哪些是偏好值,哪些是真正的期望值。文献[9]研究了元素查询条件之间存在双极性的情况,给出了需要和偏好条件之间的差别,并将模糊“winnow”操作符用于表达双极性查询。文献[8]研究了考虑查询偏好的双极性查询,其查询条件由一定数量的“正面”基本条件和一定数量的“负面”基本条件组成。正面的条件表达了所允许的,负面的条件则表达了不允许的。而那些既没有被明显允许也没有被明显禁止的值并没有被指定,主要是由于用户对被允许或不允许的东西不了解或无法列举出条件中所有(不)允许的值。

Vague集^[10]理论进一步扩展了模糊集理论,解决了模糊集理论中单值隶属度不能同时表示支持(正面信息)和反对(负面信息)的证据的问题,具有更强的表达数据模糊性和不精确性的能力^[11,12],也非常适合于异构双极信息的建模。为此,本文基于经典数据库,将Vague集引入到异构双极信息的建模,并根据实际情况对Vague集的限制条件进行了松弛化处理,将真隶属度和假隶属度分别用于表示满意度和非满意度,称之为双极满意度;并在此基础上,讨论了在选择操作(select)的查询条件中处理异构双极信息的方法以及对查询结果排序的方法。

2 基本知识

2.1 Vague集

定义1(Vague集^[10]) 设 U 为论域,其中的任意一个元素用 u 表示。 U 中的一个Vague集 V 用一个真隶属函数 t_V 和一个假隶属函数 f_V 表示, $t_V(u)$ 是从支持 u 的证据所导出的 u 的隶属度下界, $f_V(u)$ 则是从反对 u 的证据所导出的 u 的否定隶属度下界, $t_V(u)$ 和 $f_V(u)$ 都将区间 $[0,1]$ 中的一个实数与 U 中的每一个元素联系起来,即

$$t_V:U \rightarrow [0,1], f_V:U \rightarrow [0,1], \text{其中 } 0 \leq t_V(u) + f_V(u) \leq 1.$$

假设 $U = \{u_1, u_2, \dots, u_n\}$,那么论域 U 中的一个Vague集 V 可以表示如下:

$$V = \sum_{i=1}^n [t_V(u_i), 1 - f_V(u_i)] / u_i, \forall u_i \in U$$

其中, $0 \leq t_V(u) \leq 1 - f_V(u) \leq 1$ 且 $1 \leq i \leq n$ 。

Vague集的直观解释见图1。

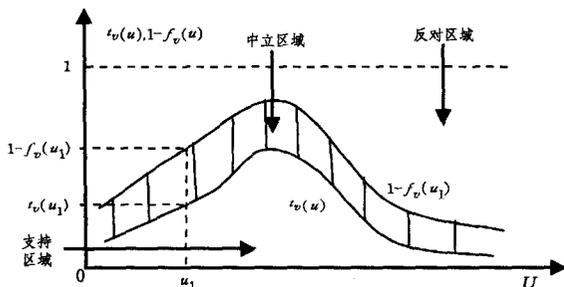


图1 Vague集的真(t_V)和假(f_V)隶属度函数

在Vague集 V 中,可用差值 $(1 - t_V(u) - f_V(u))$ 来描述对 u 了解的精度,且差值的大小与对 u 了解的精度成反比。当 $t_V(u)$ 等于 $1 - f_V(u)$ 时,对 u 的了解比较精确,此时Vague集还原至模糊集;当 $t_V(u)$ 和 $1 - f_V(u)$ 同时为0或1时,对 u 的了解最精确,此时Vague集还原至经典集合。

可以看出,Vague集可以同时表示支持(正面信息)和反对(负面信息)的证据,而差值 $1 - f_A(u) - t_A(u)$ 则可视作中立信息。显然,Vague集较模糊集具有更好的表达模糊信息的能力。

限于篇幅,下面仅给出部分Vague集的基本定义:

定义2 设Vague集 A 和Vague集 B 的并集是Vague集 C ,记作 $C = A \cup B$,其中 C 的真假隶属度函数分别为:

$$t_C = \max(t_A, t_B)$$

$$1 - f_C = \max(1 - f_A, 1 - f_B) = 1 - \min(f_A, f_B)$$

定义3 设Vague集 A 和Vague集 B 的交集是Vague集 C ,记作 $C = A \cap B$,其中 C 的真假隶属度函数分别为:

$$t_C = \min(t_A, t_B)$$

$$1 - f_C = \min(1 - f_A, 1 - f_B) = 1 - \max(f_A, f_B)$$

定义4 Vague集 V 的补集 V' 定义为: $t_{V'} = f_V, 1 - f_{V'} = 1 - t_V$ 。

2.2 面向经典数据库的模糊查询

在经典的模糊查询系统中,模糊查询主要是通过将柔性偏好和柔性表达式引入查询操作语句中来实现。通常情况下,将柔性引入查询操作语句分为两个层次:(1)将柔性引入到基本查询条件;(2)将柔性引入到复合查询条件。基本查询条件中的偏好表达允许灵活地设置搜索标准,而复合查询条件之间的偏好则可以通过为不同的查询条件设置不同的权重来实现。

面向经典数据库的模糊查询可通过在其基本查询条件中使用模糊集和模糊操作符来实现。数据库查询条件是由基本条件“ $X\theta A$ ”构成的,其中 X 表示属性, A 为模糊常量(通常用模糊集表示), θ 为模糊比较符,可视为 $=, >, <, \geq, \leq, \neq$ 等操作符的模糊扩展,如“approximately equal”和“(not) at least”等,具体可用一个带有隶属函数的模糊关系来表示 θ ,该模糊关系可视为两个取值区间为 $[0,1]$ 论域的笛卡尔积的模糊子集。

选择操作是查询操作的基本操作之一,也是研究查询操作的重要内容。本文将选择操作语句为研究对象,主要讨论将“ $X\theta A$ ”作为其基本查询条件时的柔性查询处理方法。在研究异构双极时,为了更好地表示异构双极信息,将约束条件松弛后的Vague集引入模糊常量 A 的表达。

在研究“ $X\theta A$ ”中包含异构双极信息的柔性查询处理方法之前,先看一下该条件中包含传统模糊信息(用模糊集表示)时的处理方法。

在传统的模糊查询方法中,对于基本查询条件“ $X\theta A$ ”,查询条件中带有隶属度函数 μ_A 的模糊集 A 可以用一个模糊词来表示。这个模糊集 A 表达了用户所需要的信息,即所有 X 的值“among”给定模糊集 A 的元组 t 。因此, A 中的所有值都表达了用户的需求, $\mu_A(x)$ 的值越大, x 就越满足用户的需求。

对于给定的元组 t ,基本查询条件“ $X\theta A$ ”的计算记为 $e(X\theta A)(t)$,计算所得的满意度 d 用于度量 $t[X]$ 的值与 A 所表示的值相匹配的程度,具体可使用以下公式进行计算:

$$e(X\theta A)(t) = \sup_{x \in dom} \min(\mu_A(x), \mu_\theta(t[X], x))$$

当“ θ ”取“ $=$ ”时,显然有

$$e(X\theta A)(t) = e(X=A)(t) = \mu_A(t[X])$$

设 C_1 和 C_2 为基本查询条件,根据模糊逻辑连接的语义,

组合查询条件的计算可使用以下公式:

$$e(C_1 \wedge C_2)(t) = \min(e(C_1)(t), e(C_2)(t))$$

$$e(C_1 \vee C_2)(t) = \max(e(C_1)(t), e(C_2)(t))$$

$$e(\neg C_1)(t) = 1 - e(C_1)(t)$$

3 基于扩展 Vague 集的双极满意度

在传统的模糊查询中仅使用单隶属度表示查询满意度 d , 其不满意度则可由 $1-d$ 计算而得, 但当负面信息和正面信息不再对合时, 就不仅仅需要查询满意度(正面信息), 还需要明确查询不满意度(负面信息)。仍以用户购买手机为例, 如果用户声明不想购买白色的手机, 那么此声明并不能保证所有非白色的手机都会被用户所接受。虽然用户没有明确指定, 但总会存在其它种颜色的手机也会被用户拒绝。反之, 如果用户声明将要买一部红色的手机, 同样, 此声明也不能确保用户会拒绝所有非红色的手机。虽然用户也没有明确指定, 但总会存在其它种颜色的手机也会被用户接受。为处理这种查询条件中存在异构双极的情况, 本文将引入 Vague 集来表示双极查询满意度。

双极满意度主要用于表示满意度和非满意度, 其值可用以下值对表示

$$(s, d), s, d \in [0, 1]$$

其中, s 为满意度, d 为非满意度。

s 和 d 的取值范围皆为 $[0, 1]$ 且相互独立, 分别表示满意度和非满意度。(1, 0) 表示“完全满意, 没有不满意的”; (0, 1) 表示“没有满意的, 完全不满意”, 设双极满意度为 ds , 那么有 $ds = \{(s, d) | s, d \in [0, 1]\}$

从语义上看, 双极满意度与 Vague 集是直接相关的。为了更好地反映人们的意愿, 保证满意度和非满意度的相互独立性, 我们将 Vague 集进行了扩展, 即允许满意度和非满意度的值突破 $0 \leq s + d \leq 1$ 的约束。实际上, 由于 s 和 d 在反映人们意愿时是完全独立的, 而且人们在表达自己的正、负面意愿时往往也会发生一定的冲突, 因此在用双极满意度表达人们的意愿时, 允许 $s + d > 1$ 的情况出现, 这种情况在现实世界中也是实际存在的。

例 1 图 2—图 4 所示的双极查询条件既包含了正面的信息, 也包含了负面的信息, 其相应的真隶属函数和假隶属函数分别用 t_{price} 和 f_{price} 表示, 这两个函数都定义在论域“手机的价格”上, 可视为用户查询手机时的一个查询条件。图示的 3 种情况分别表明了对“价格较高手机”的不满意度以及对“价格适中手机”的满意度。图 2 与图 3 中 t_{price} 和 f_{price} 满足 Vague 集的限制条件 $0 \leq t_{price} + f_{price} \leq 1$, 使用的是 Vague 集形式的示意图。图 4 中 t_{price} 和 f_{price} 则突破了 Vague 集的限制条件, 使用的是类 Vague 集形式的示意图。

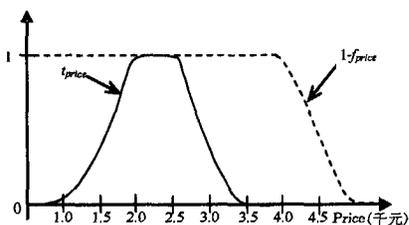


图 2 t_{price} 和 f_{price} 未发生交叉的情况 ($0 \leq t_{price} + f_{price} \leq 1$)

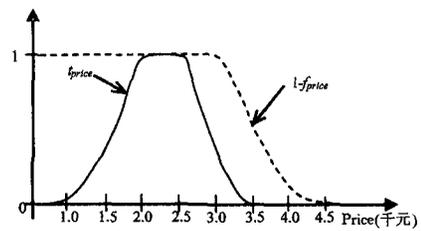


图 3 t_{price} 和 f_{price} 发生交叉的情况 ($0 \leq t_{price} + f_{price} \leq 1$)

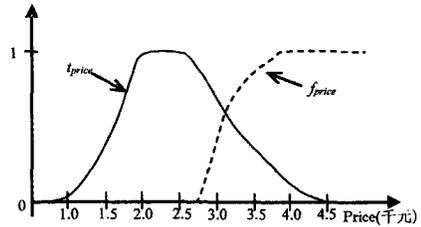


图 4 t_{price} 和 f_{price} 发生交叉的情况 (存在 $t_{price} + f_{price} > 1$)

在图 2 中, 真隶属函数和假隶属函数之间没有交叉, 且域中的所有值都满足 $0 \leq t_{price}(x) + f_{price}(x) \leq 1$ 。但对域中的一些值 ($x \in [3.5, 4.0]$) 有 $t_{price}(x) = 0$ 和 $f_{price}(x) = 0$ 同时成立, 这意味着用户完全忽视了 $(1 - t_{price}(x) - f_{price}(x) = 1)$ 这个价位的手机是否合适。

在图 3 中, 域中的所有值都满足 $0 \leq t_{price}(x) + f_{price}(x) \leq 1$, 因此两个函数之间没有冲突。而在真隶属函数和假隶属函数交叉区间 $(3.0, 3.5)$ 中的值皆满足 $t_{price}(x) + f_{price}(x) < 1$, 即用户对这些价位的手机存在一定的踌躇, 这意味着这些价位的手机并没有引起用户的足够重视。

在图 4 中, 隶属度函数和非隶属度函数在区间 $(2.5, 4.0]$ 之间有一定的交叉, 而且其中的一些值还发生了冲突, 使得 $t_{price}(x) + f_{price}(x) > 1$ 。例如, 对于 $x = 4.0$, 其不满意度是 1, 但同时其满意度为 0.1。当然, 这是相互矛盾的, 看起来有悖常理, 但也有可能发生。这就像用户说“我不想买 4 千元的手机, 但是在最坏情况下, 对于这个价位的手机, 我还是有 0.1 的满意度”。从用户的角度来看, 这种冲突情况并不能完全排除, 因为人们的意愿很难总是保持一致。

4 面向经典数据库的双极性信息处理方法

使用双极满意度的优点是用户能指定同时包含正面和负面信息的查询条件, 这在一定程度上解决了当用户对论域值不完全了解或用户无法对论域中的所有值都赋予一个相应的满意度和不满意度时所产生的问题。

4.1 基本双极查询条件

通常情况下, 双极性查询条件能为论域中的每个值都指定一个取值范围为 $[0, 1]$ 的满意度和不满意度。因此, 基本查询选择条件内用于表达用户偏好的异构双极条件能形式化地用模糊集的双极扩展进行建模。而 Vague 集正满足这样的要求。但在 Vague 集的定义(定义 1)中, 其一致性约束条件 $0 \leq t_v(u) \leq 1 - f_v(u) \leq 1$ 即 $0 \leq t_v(u) + f_v(u) \leq 1$ 的存在, 影响了异构双极信息中满意度和不满意度取值的独立性。为了充分反映现实世界中用户偏好突破了一致性约束条件的情况, 可以将 Vague 集的约束条件进行松弛化处理, 这就意味着 $0 \leq t_v(u) + f_v(u) \leq 1$ 不一定必须成立。

考虑一个针对属性 A 的基本选择条件 C , 属性 A 的论域

dom_A 表达了与属性 A 的元组值相关的用户的偏好。因此，条件 C 的最简单的通用形式可以用 Vague 集建模为：

$$C = \sum_{i=1}^n \langle t_C(x_i), f_C(x_i) \rangle / x_i, \forall x_i \in dom_A$$

其中，隶属度函数 t_C 定义了用户的正面偏好，也就是说，与域值 $x \in dom_A$ 相关的隶属度 $t_C(x)$ 表示 x 满足属性 A 的程度；非隶属度函数 f_C 定义了用户的负面偏好，也就是说，与域值 $x \in dom_A$ 相关的隶属度 $f_C(x)$ 表示 x 不满足属性 A 的程度。

双极查询条件大体有如下 3 类：

第一类是同时提供反映用户正面偏好的隶属度函数和反映用户负面偏好的非隶属度函数。

第二类是仅提供反映用户正面偏好的隶属度函数 t_C 。在这种情况下，用户未明确其负面的意愿，此时可假定非隶属度函数与隶属度函数相对称，那么条件 C 可表示为

$$C = \sum_{i=1}^n \langle t_C(x_i), 1 - t_C(x_i) \rangle / x_i, \forall x_i \in dom_A$$

第三类是仅提供反映用户负面偏好的隶属度函数 f_C 。在这种情况下，用户未明确其正面的意愿，此时同样可假定隶属度函数与非隶属度函数相对称，那么条件 C 可表示为

$$C = \sum_{i=1}^n \langle 1 - f_C(x_i), f_C(x_i) \rangle / x_i, \forall x_i \in dom_A$$

需要指出的是，仅具有满意度参数的传统模糊查询方法可视为双极性信息查询方法的一种特例（仅提供正面信息）。此时的双极性信息为对称双极，其约束的语义和经典模糊查询相同，即用于表达用户的偏好。

4.2 双极满意度的计算与聚合操作

通常情况下，双极性柔性查询可由几个独立的基本双极查询条件组合而成。一般来说，用基本双极性查询条件 C_A 对数据表 r 中的属性 A 进行查询，会产生一个双极满意度，计算方法如下：

$$(s_{C_A}, d_{C_A}) = (t_{C_A}(R[A]), f_{C_A}(R[A])) \quad (1)$$

其中， s_{C_A} 和 d_{C_A} 分别是 r 中元组的属性 A 对查询条件 C_A 的满意度和非满意度。

如果用户仅给定正面信息，式(1)可简化为：

$$(s_{C_A}, d_{C_A}) = (t_{C_A}(R[A]), 1 - t_{C_A}(R[A]))$$

类似地，如果用户仅给定负面信息，式(1)可简化为：

$$(s_{C_A}, d_{C_A}) = (1 - f_{C_A}(R[A]), f_{C_A}(R[A]))$$

而对于复合双极查询条件，当计算其查询结果时，须对组成该条件的基本双极性查询条件的查询结果进行聚合操作，因而，首先应分别计算出所有基本双极查询条件 C_i ($i=1, 2, \dots, n$) 对特定记录 R 的双极满意度 (s_i, d_i) ($i=1, 2, \dots, n$)，进而将这些满意度聚合为一个单一的双极满意度 (s_c, d_c) ，并将其作为整个查询的全局双极满意度。对双极满意度进行聚合的方法有很多，本文仅讨论其中最基本的双极交 (\wedge)、双极并 (\vee) 以及双极补 (\neg) 操作。

1. 双极交操作

两个基本查询条件 C_1 和 C_2 双极交操作的结果是同时满足查询条件 C_1 和 C_2 的记录的集合。在双极性查询中，双极查询条件使用类 Vague 集（没有一致性约束条件）的形式，假设 C_1 和 C_2 的双极隶属度函数对分别是 (t_{C_1}, f_{C_1}) 与 (t_{C_2}, f_{C_2}) ，根据 Vague 集中的交操作，我们有：

$$C_1 \wedge C_2 = (\min(t_{C_1}(x), t_{C_2}(x)), \max(f_{C_1}(x), f_{C_2}(x))) / x, x \in dom_A$$

相应地，我们可以得出同时满足条件 C_1 和 C_2 的双极满意度计算方法（见式(2)）：

$$(s_{C_1 \wedge C_2}, d_{C_1 \wedge C_2}) = (\min(t_{C_1}, t_{C_2}), \max(f_{C_1}, f_{C_2})) \quad (2)$$

需要说明的是，式(2)不仅在条件 C_1 和 C_2 定义在相同属性和相同论域时有效，而且定义在不同属性和论域时同样有效。因为要想满足合并后的复合条件，就必须先分别满足合并前的各个条件，因此取两个独立满意度的最小值。而如果查询结果不满足合并后的复合条件，则需要至少不满足其中的一个条件，因此取两个独立不满意度的最大值。因此即使 C_1 和 C_2 定义在不同属性和论域，式(2)同样有效。

2. 双极并操作

基本查询条件 C_1 和 C_2 双极并操作的处理方法与双极交操作的处理方法相同。一方面，要想满足 C_1 和 C_2 双极并操作后的结果，必须至少满足其中的一个条件，因此，其满意度应该取两个独立满意度的最大值。另一方面，要想不满足 C_1 和 C_2 双极并操作后的结果，必须同时不满足 C_1 和 C_2 ，因此，其不满意度应该取两个独立不满意度的最小值。根据 Vague 集中的并操作，可以得出 C_1 和 C_2 双极并操作后的双极满意度计算方法：

$$(s_{C_1 \vee C_2}, d_{C_1 \vee C_2}) = (\max(t_{C_1}, t_{C_2}), \min(f_{C_1}, f_{C_2}))$$

3. 双极补操作

双极性条件 C 双极补操作的双极满意度 $(s_{\neg C}, d_{\neg C})$ 可以通过交换条件 C 双极满意度中的满意度和非满意度值来获得，计算方法如下：

$$(s_{\neg C}, d_{\neg C}) = (f_C, t_C)$$

需要注意的是，双极补操作与初始条件 C 双极满意度中满意度和非满意度的取反操作不同，取反操作的结果是 $(t_{\neg C}, f_{\neg C}) = (1 - t_C, 1 - f_C)$ ，如果使用这种方法，完全漠视 $(d_s(0, 0))$ 和完全冲突 $(d_s(1, 1))$ 则互为补操作。显然这是不正确的，完全漠视的补还应该是完全漠视，完全冲突的补还应该是完全冲突。

4.3 双极性查询结果排序

双极模糊查询的每个查询结果元组都会有一个与之相关联的双极满意度，为了根据全局查询满意度对这些元组进行排序，就需要一个双极满意度排序函数。可以使用不同的排序函数。在此给出一种赋予了满意度和非满意度相同权重的基本排序方法，即取 $s-d$ 为全局查询满意度的值。

在进行双极性信息查询时，可按 $s-d$ 从大到小的顺序对查询结果进行排序。当 $s=1, d=0$ 时， $s-d$ 取最大值 1，此时查询结果完全满足查询要求，没有任何踌躇和冲突；当 $s=0, d=1$ 时， $s-d$ 取最小值 -1，此时查询结果完全不满足查询要求，也没有任何踌躇和冲突；当 $s=d$ ，即用户的满意度和非满意度相同时， $s-d$ 取中间值 0，此时用户对查询结果持中立态度。

需要注意的是，使用此排序方法，当查询结果的不满意度值为 1 或满意度值为 0 时，不应该优先将该查询结果作为完全不满足查询条件而从查询结果中剔除。实际上，例如，双极满意度 $(1, 1)$ 和 $(0, 0)$ 尽管分别有 $d=1, s=0$ 成立，但二者皆有中间排序值 $s-d=0$ ，因而不能视为完全不满足查询条件。

例 2 给定一手机数据表，其中包含数据表“phone”，数据表中包含“颜色”和“屏幕尺寸”两个字段，如表 1 所列。

（下转第 181 页）

[14] 郭志林. 随机 P-集合的数据筛选过滤[J]. 河南科技大学学报: 自然科学版, 2012, 33(2): 83-86

[15] 郭志林. P-集合的随机特征[J]. 模糊系统与数学, 2011, 25(2): 170-174

[16] 郭志林. F_p -规律与 F_p -规律积分[J]. 度量烟台大学学报: 自然科学与工 程版, 2010, 23(4): 260-264

[17] 史开泉. 函数 S-粗集[J]. 山东大学学报: 理学版, 2005, 40(1): 1-

[18] 史开泉, 姚炳学. 函数 S-粗集与规律辩证[J]. 中国科学 E: 信息科学, 2008, 38(4): 553-564

[19] 史开泉, 赵建立. 函数 S-粗集与规律安全-认证[J]. 中国科学 E: 信息科学, 2008, 38(8): 1234-1243

[20] 刘江臣, 陈保会. 基于 S-粗集的粗数据规律识别[J]. 山东大学学报: 理学版, 2008, 43(10): 60-66

(上接第 156 页)

表 1 关系 phone

编号	颜色	屏幕尺寸(英寸)
1	黑色	5
2	黑色	4
3	白色	4.5
4	红色	4

考虑以下双极性查询: 不想要白色的手机, 喜欢黑色且屏幕大小为 4 英寸左右的手机。

设对应于此要求的选择查询条件为

$$C_{颜色} \wedge C_{屏幕大小}$$

其中,

$C_{颜色} = \{ \langle \text{白色}, 0, 1 \rangle, \langle \text{黑色}, 1, 0 \rangle \}$ 为双极性偏好, 用类 Vague 集的形式表示。

$C_{屏幕大小} = \{ (3, 0, 1), (3.5, 0, 6), (3.6, 0, 8), (3.8, 1)(4, 1), (4.2, 1), (4.5, 0, 5), (5, 0, 1) \}$ 为正面偏好。

表 2 所列带有双极满意度和排序值 ($s-d$) 的查询结果。

表 2 查询结果集

编号	双极满意度 (颜色)	双极满意度 (屏幕尺寸)	双极满意度 (总评)	排序值
2	(1,0)	(1,0)	(1,0)	1
4	(0,0)	(1,0)	(0,0)	0
1	(1,0)	(0,1,0,9)	(0,1,0,9)	-0,8
3	(0,1)	(0,5,0,5)	(0,1)	-1

从表 2 可以看出, 2 号、4 号手机为前两个满足用户需要的手机, 4 号手机之所排在第二位, 主要是因为在使用 Vague 集对用户的双极性需求建模时忽略了黑色和白色以外颜色的手机。而在传统的不支持异构双极信息的查询系统中, 尽管用户有可能对红色的 4 号手机感兴趣, 但其也会被排除在检索结果之外。

另外, 在一些情况下, 可根据实际情况给满意度或非满意度一些较多的权重。一种方法是根据需要分别将满意度(非满意度)作为第一关键字或第二关键字进行排序。另一种方法是将双极满意度进行融合, 将其转换为单值满意度, 并在转换过程中根据需要给予满意度或非满意度以较高的权重, 具体方法暂不作为本文的研究内容。

结束语 在对信息系统进行模糊查询时, 用户以异构双极的形式表达查询条件具有一定的实用性和应用背景。本文基于经典数据库, 将 Vague 集引入异构双极信息的表达, 并根据异构双极信息中正、负面信息的相互独立性, 对 Vague 集进行了松弛化处理。在此基础上, 讨论了选择操作查询条件中异构双极信息的处理方法, 给出了一个基于双极满意度

的双极模糊查询处理逻辑框架。双极满意度为数据库用户提供了一个查询工具, 该工具不仅能够返回查询条件满意度信息, 还可以返回相关的不满意度信息。这种方法能够将数据库中既不满足正面条件也不满足负面条件的元组识别出来。这正好反映了人们在现实生活中选择问题的最佳解决方案时所采取的策略。

参 考 文 献

[1] Dubois D, Fargier H. Qualitative decision-making with bipolar information[C]//Proc. 10th Int. Conf. Principles Knowl. Representation Reason. Menlo Park, CA, 2006: 175-186

[2] Dubois D, Prade H. Handling bipolar queries in fuzzy information processing[M]//Galindo J, ed. Handbook of Research on Fuzzy Information Processing in Databases. New York: Inform., Science Ref., 2008: 97-114

[3] Dubois D, Prade H. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory[J]. Fuzzy Sets and Systems, 2009, 160(10): 1355-1366

[4] Zadrozny S, De Tré G, De Caluwe R, et al. An Overview of Fuzzy Approaches to Flexible Database Querying[C]//Galindo J, ed. Handbook of Research on Fuzzy Information Processing in Databases. IGI Global, New York, 2008: 34-54

[5] Zadeh L A. Fuzzy sets [J]. Information and Control, 1965, 8(3): 338-353

[6] Bosc P, De Tré G, Dujmović J, et al. On advances in soft computing applied to databases and information systems[J]. Fuzzy Sets and Systems, 2012, 196: 1-3

[7] Zadrozny S, Kacprzyk J. Bipolar queries: An aggregation operator focused perspective[J]. Fuzzy Sets and Systems, 2012, 196: 69-81

[8] Lacroix M, Lavency P. Preferences: Putting more knowledge into queries[C]//Proc. 13 Int. Conf. Very Large Databases. Brighton, U. K., 1987: 217-225

[9] Zadrozny S, Kacprzyk J. Bipolar queries and queries with preferences[C]//Proc. of the DEXA Conference (DEXA 2006). Krakow, Poland, 2006: 415-419

[10] Gau W L, Buehrer D J. Vague sets[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1993, 23(2): 610-614

[11] Lu A, Ng W. Vague sets or intuitionist fuzzy sets for handling vague data; which one is better [J]. Lecture Notes in Computer Science, 2005, 3716: 401-416

[12] 欧阳春娟, 李斌, 李霞, 等. 基于 Vague 集相似度量的图像隐写系统安全性测度[J]. 计算机学报, 2012, 35(7): 1510-1521