

基于模糊描述逻辑的 PNL 网络问答系统

高明霞, 刘椿年

(北京工业大学计算机学院多媒体与智能软件技术重点实验室, 北京 100022)

摘要: 将现有网络搜索引擎过渡到网络问答系统是 WI 的基本目标之一, 模糊问答系统是网络问答系统研究的重要问题。基于模糊描述逻辑, 该文特化了 PNL 过程, 提出了 PNL 式模糊网络问答系统。该系统将模糊描述逻辑规则作为 PNL 推理过程中用到的原型语言规则中的计算部分, 推理出原型语言规则中的知识符号, 并以一个具体的问答实例, 介绍了 PNL 式问答系统的工作流程。

关键词: 模糊描述逻辑; 精确的自然语言; 问答系统

PNL Question-answering System Based on Fuzzy Description Logic

GAO Mingxia, LIU Chunnian

(Key Laboratory of Multimedia and Intelligent Software Technology, College of Computer, Beijing University of Technology, Beijing 100022)

【Abstract】 Upgrading a search engine to a Web-based question-answering system is one of basic target concerning certain aspects of Web intelligence (WI). The question-answering system is more important for Web-based question-answering research. This paper specifies precise natural language process based on fuzzy description logic as fuzzy PNL question-answering system. The system uses fuzzy description logic as computational part in prototype language rules. It makes use of reasoning algorithms to study knowledge symbols, and gives an example of query film to show the system workflow.

【Key words】 Fuzzy description logic; Precise natural language(PNL); Question-answering system

随着网络信息的膨胀、服务的增多, 及时有效地从网络获得用户需要的知识, 或者借助网络解决复杂的实际问题是一个具有挑战性的工作, 因此WI的研究得到了广泛的关注, 并取得了一些成果^[1,2]。问答系统研究有很长的历史, 最早的问答系统大多是受限问答系统, 即答案是从有限的集合中获得的。最常见的受限问答系统有专家系统、常见问题回答系统等。随着网络知识的膨胀, 出现将网络作为信息源的网络问答系统。其中比较著名的有华盛顿大学的MULDER, 微软研究院的AskMSR等。网络问答系统和传统系统的最大不同是冗余网络数据驱动获得正确答案。

现有问答系统无论是受限的还是网络驱动的, 处理的对象大多是确定的知识。尽管有些问答系统可以处理模糊知识, 如基于数据库的模糊问答系统、模糊专家系统等。但是它们通常是将确定的知识模糊化, 利用已知的模糊规则分类, 再将结果精确, 这些系统更像是一个模糊分类器。处理能力受限于已知的模糊规则。

万维网上存在许多不确定、不精确的模糊知识。人的知识大多来自于感觉, 而感觉本质上就是不精确的, 它反映的是人的感觉器官在处理信息上的受限能力, 让网络问答系统处理这些模糊概念, 模糊关系是至关重要的。Zadeh从实际知识相关性、精确性的模糊本质出发, 提出了解决这些问题的新工具: 精确的自然语言, 概化约束语言, 原型形式语言等^[3-5]。但是这些知识是从宏观的角度对整个模糊问题的探讨, 具体领域的实现过程和应用涉及很少。

1 模糊描述逻辑

本文使用的模糊描述逻辑是文献[6,7]中介绍的模糊ALC。本节简要介绍模糊ALC, 关于该语言更多内容请参考

文献[6,7]。

定义 1 令 C 、 R 分别是模糊概念和模糊关系, 则 $C \cup D$, $C \cap D$, $\neg C$, $\forall R.C$, $\exists R.C$ 也是模糊概念。

定义 2 令 A 、 C 分别是简单模糊概念和模糊概念, 则 $A \approx C$ 是模糊概念定义, $A < C$ 是模糊概念特化。由模糊概念定义和模糊概念特化组成的知识库是模糊术语知识库。

定义 3 令 C 、 R 分别是模糊概念和模糊关系, a 、 b 是域中的个体对象, $n \in (0, 1]$, 则 $C(a) \geq n$, $C(a) \leq n$, $R(a, b) \geq n$ 是模糊声明。由模糊声明组成的知识库是模糊声明知识库。

定义 4 模糊 ALC 的语义。一个模糊解释 $FI = (\Delta, I)$, 其中 Δ 是域, I 是解释函数。 I 将域 U 中的不同个体映射成域中不相等的元素, 将域 U 中模糊概念 A 映射成隶属函数 $A^I: \Delta \rightarrow [0, 1]$, 将域 $U \times U$ 中模糊关系 R 映射成隶属函数 $R^I: \Delta \times \Delta \rightarrow [0, 1]$ 。对于所有 $d \in \Delta$, 满足条件如下:

$$\begin{aligned} U^I(d) &= 1 \\ (\neg U)^I(d) &= 0 \\ (C \cap D)^I(d) &= \min\{C^I(d), D^I(d)\} \\ (C \cup D)^I(d) &= \max\{C^I(d), D^I(d)\} \\ (\neg C)^I(d) &= 1 - C^I(d) \\ (\forall R.C)^I(d) &= \inf_{d' \in \Delta} \{\max\{1 - R^I(d, d'), C^I(d')\}\} \\ (\exists R.C)^I(d) &= \sup_{d' \in \Delta} \{\min\{R^I(d, d'), C^I(d')\}\} \end{aligned}$$

定义 5 模糊解释 $FI = (\Delta, I)$ 满足模糊概念特化 $A < C$, 当

基金项目: 国家自然科学基金资助项目(60173014); 北京自然科学基金资助项目(4022003); 重点实验室开放基金资助项目

作者简介: 高明霞(1973 -), 女, 博士生, 主研方向: 语义网技术, Ontology 信息管理, 问答系统; 刘椿年, 博士、教授、博导

收稿日期: 2006-01-17 **E-mail:** gaomx@emails.bjut.edu.cn

且仅当 $\forall d \in \Delta$ 时, 有 $A^l(d) \leq C^l(d)$ 。模糊解释 $FI = (\Delta, I)$ 满足模糊概念定义 $A \approx C$, 当且仅当 $\forall d \in \Delta$ 时, 有 $A^l(d) = C^l(d)$ 。模糊解释 $FI = (\Delta, I)$ 满足模糊术语知识库 K_{FT} , 当且仅当 FI 满足知识库中的每一条术语公理, 表示为 $FI \models K_{FT}$ 。

定义 6 模糊解释 $FI = (\Delta, I)$ 满足模糊声明: 当且仅当 $A^l(a') \geq n, A^l(a') \leq n, R^l(a', b') \geq n, A(a) \geq n, A(a) \leq n, R(a, b) \geq n$ 。当且仅当 FI 满足知识库中每一条声明, 模糊解释 $FI = (\Delta, I)$ 满足模糊声明知识库 K_{FA} , 表示为 $FI \models K_{FA}$ 。

定义 7 模糊知识库 $FK = \{K_{FT}, K_{FA}\}$ 如果有模型, 则该知识库可满足。模糊声明 α 和模糊术语公理 Φ 是知识库 FK 的模糊推论, 当且仅当知识库的所有模型也满足该模糊声明 α 或模糊公理 Φ , 表示为 $FK \models \alpha, FK \models \Phi$ 。

2 基于模糊描述逻辑的 PNL 式问答系统

2.1 PNL 的简介

简单地说, PNL 就是一个如图 1 所示的过程。其中, p 是自然语言中可精确化的命题, $p^*/GC(p)$ 是该命题精确化后形成的概化约束语言 (Generalized-constraint Language, GCL), $p^{**}/PF(p)$ 是精确化的命题通过抽象形成的原型形式语言 (Protoform Language, PFL)。

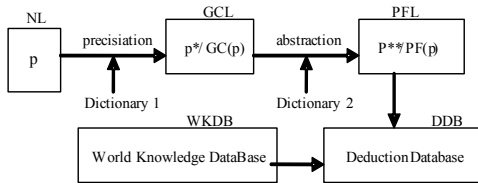


图 1 PNL 的基本结构

GCL 的表达式为 $X \text{ is } R$ 。其中, X 是约束变量, R 是约束关系, r 是特征标示符。不同的特征标示符表示不同的约束。PFL 是一个对象 (命令、命题、问题、答案等) 的抽象总结。在 PNL 过程中 PFL 主要抽象精确化的自然语言命题 p^* 。根据不同的特征标示符, GCL 有不同的约束形式, 从 p 到 $p^*/GC(p)$ 的精确规则和从 $p^*/GC(p)$ 到 $p^{**}/PF(p)$ 的抽象原则和特定的约束形式相关。针对不同的应用领域, 最简单的方法就是总结出对应的转换字典。一个简单的转换字典例子如表 1 所示。

表 1 转换字典

字典 1		字典 2	
自然语言命题 p	精确化 $GC(p)$	精确化 $GC(p)$	抽象 $PF(p)$
小李很高	Height(小李)is 很高	Height(小李)is 很高	A(X)is C

DDB 是管理概化约束传播的原型形式规则集合, 这些规则是推理的基础, 它们通常有符号部分和计算部分。符号部分是用原型形式表示的。计算部分因特征标示符 r 的不同而各异。WKDB 是实际知识的数据库。

2.2 基于模糊描述逻辑的 PNL 问答系统

PNL 的整个过程可以看作是模糊问答系统的准备过程。但是最重要的推理或者解决问题的算法依然需要根据 GCL 的约束领域确定。只有确定了特征标示符 r , DDB 中的原型形式规则需要的推理规则及计算部分才能确定。实现一个 PNL 式的问答系统需要下列步骤:

- (1) 确定 GCL 的约束领域, 即将特征标示符 r 具体化。并用 PFL 重写约束领域中的各种推理规则作为 DDB 的一部分;
- (2) 建立约束领域的精确化、抽象化字典;
- (3) 借助上述字典将该领域的问题和知识转换为对应的 PFL 形式;
- (4) 用 DDB 中约束领域的规则和算法解决问题;

(5) 将答案通过相应的字典转换为自然语言。

表 2 模糊描述逻辑对应的 PFL

模糊描述逻辑	PFL 形式
$\mu_{C \cap D}(u) = \min\{\mu_C(u), \mu_D(u)\}$	$X \text{ is } C$ $X \text{ is } D$ $X \text{ is } C \cap D$
$\mu_{C \cup D}(u) = \max\{\mu_C(u), \mu_D(u)\}$	$X \text{ is } C$ $X \text{ is } D$ $X \text{ is } C \cup D$
$\mu_{\neg C}(u) = 1 - \mu_C(u)$	$X \text{ is } C$ $X \text{ is } \neg C$
$\mu_{\exists R.C}(u) = \sup_{u' \in U} \{\min\{\mu_R(u, u'), \mu_C(u')\}\}$	$X \text{ is } C$ $(X, Y) \text{ is } R$ $Y \text{ is } \exists R.C$
$\mu_{\forall R.C}(u) = \inf_{u' \in U} \{\max\{1 - \mu_R(u, u'), \mu_C(u')\}\}$	$X \text{ is } C$ $(X, Y) \text{ is } R$ $Y \text{ is } \forall R.C$

表 3 声明规则的 PFL 形式

模糊描述逻辑计算形式	PFL 表示形式
$(\neg \geq) \langle \neg A(w) \geq n \rangle \rightarrow \langle A(w) \leq 1-n \rangle$	$(X \text{ is } \neg A) \geq n$ $(X \text{ is } A) \leq 1-n$
$(\neg \leq) \langle \neg A(w) \leq n \rangle \rightarrow \langle A(w) \geq 1-n \rangle$	$(X \text{ is } \neg A) \leq n$ $(X \text{ is } A) \geq 1-n$
$(\cap \geq) \langle (A \cap B)(w) \geq n \rangle \rightarrow \langle A(w) \geq n \rangle, \langle B(w) \geq n \rangle$	$(X \text{ is } A \cap B) \geq n$ $(X \text{ is } A) \geq n$ $(X \text{ is } B) \geq n$
$(\cap \leq) \langle (A \cap B)(w) \leq n \rangle \rightarrow \langle A(w) \leq n \rangle, \langle B(w) \leq n \rangle$	$(X \text{ is } A \cap B) \leq n$ $(X \text{ is } A) \leq n \mid (X \text{ is } B) \leq n$
$(\cup \geq) \langle (A \cup B)(w) \geq n \rangle \rightarrow \langle A(w) \geq n \rangle \mid \langle B(w) \geq n \rangle$	$(X \text{ is } A \cup B) \geq n$ $(X \text{ is } A) \geq n \mid (X \text{ is } B) \geq n$
$(\cup \leq) \langle (A \cup B)(w) \leq n \rangle \rightarrow \langle A(w) \leq n \rangle, \langle B(w) \leq n \rangle$	$(X \text{ is } A \cup B) \leq n$ $(X \text{ is } A) \leq n$ $(X \text{ is } B) \leq n$
如果 $\Psi = \langle R(w, w') \leq 1-n \rangle$ $(\forall \geq) \langle (\forall R.A)(w) \geq n \rangle, \Psi' \rightarrow \langle A(w') \geq n \rangle$	条件: $\Psi = ((X, Y) \text{ is } R) \leq 1-n$ $(X \text{ is } \forall R.A) \geq n$ 且 Ψ^c $(Y \text{ is } A) \geq n$
如果 $\Psi = \langle R(w, w') \leq n \rangle$ $(\exists \leq) \langle (\exists R.A)(w) \leq n \rangle, \Psi' \rightarrow \langle A(w') \leq n \rangle$	条件: $\Psi = ((X, Y) \text{ is } R) \leq n$ $(X \text{ is } \exists R.A) \leq n$ 且 Ψ^c $(Y \text{ is } A) \leq n$
如果 x 是新变量, 模糊声明集中没有 $\langle R(w, w') \geq n \rangle, \langle A(w') \geq n \rangle$ 这 2 种形式 $(\exists \geq) \langle (\exists R.A)(w) \geq n \rangle \rightarrow \langle R(w, x) \geq n \rangle, \langle A(x) \geq n \rangle$	条件: Y 是新变量, 模糊声明集中没有 $\langle R(w, w') \geq n \rangle, \langle A(w') \geq n \rangle$ 这 2 种形式 $(X \text{ is } \exists R.A) \geq n$ $((X, Y) \text{ is } R) \geq n$ $(Y \text{ is } A) \geq n$
如果 x 是新变量, 模糊约束集中没有 $\langle R(w, w') \geq 1-n \rangle, \langle A(w') \leq n \rangle$ 这 2 种形式 $(\forall \leq) \langle (\forall R.A)(w) \leq n \rangle \rightarrow \langle R(w, x) \geq 1-n \rangle, \langle A(x) \leq n \rangle$	条件: Y 是新变量, 模糊声明集中没有 $\langle R(w, w') \geq 1-n \rangle, \langle A(w') \leq n \rangle$ 这 2 种形式 $(X \text{ is } \forall R.A) \leq n$ $((X, Y) \text{ is } R) \geq 1-n$ $(Y \text{ is } A) \leq n$

上述过程中最重要的步骤是确定约束领域, 重写约束领域中的推理规则。表 2 就是将特征标示符 r 具体化为模糊描述逻辑后, 用 PFL 格式重写的语义规则。则 PFL 格式是 DDB 中推理规则的符号部分, 模糊描述逻辑的计算方式就是对应推理规则的计算部分。从模糊描述逻辑的计算公式可以看出, 要得出最终的结果, 需要 $p^{**}/PF(p)$ 的具体真值。相关规则 PFL 形式如表 3 所示。在实际应用领域, 这个值基本在一个认知估计的范围。在模糊描述逻辑中就是以声明 $C(a) \geq n, C(a) \leq n, R(a, b) \geq n$ 的形式出现的, 问题的解决可以归结到声明知识库的满足。这样就需要将模糊描述逻辑和声明相关的规则以 PFL 格式重写。这样不但需要将特征标示符 r 具体化为模糊描述逻辑, 也要结合上特征标示符 r 具体化为

≥、≤、>、< 这样的形式。

关系符 >、< 的传播规则和 ≥、≤ 相似，所以没有列出。规则中的 Ψ, Ψ^c 是一对共轭声明。当满足表 4 中条件的行列的声明时，可认为其就是一对共轭声明。

表 4 共轭声明

	$a < m$	$a = m$
$a < n$	$n = m$	$n > m$
$a > n$	$n = m$	$n = m$

2.3 模糊问答系统框架

有了这些基于模糊描述逻辑的 PFL 格式的规则，就可以借用 PNL 过程，以自然语言为具体的问题和答案语言，完成某一具体领域的模糊问答系统，系统框架如图 2 所示。

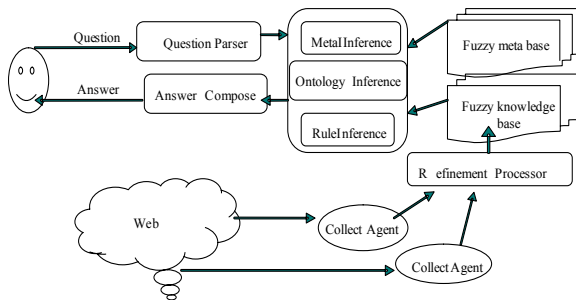


图 2 模糊问答系统

问题分析模块确定 GCL 中的特征标示符 r 并转换问题的，即根据实际问题确定问题领域并利用该领域的转换字典将问题转换为 PFL 格式。问题集成模块将推理结果转换为自然语言式的答案。整个系统的核心就是各个层次的推理机，它们利用领域推理规则解决 PFL 的问题和知识并获得对应的结果。其中的精化处理模块就是将 Agent 收集到的知识转换为 PFL 格式并存储到对应的知识库中。

3 实例

模糊描述逻辑的 PNL 式问答系统的适用范围受 2 个因素影响：(1) 该领域的问题解决可以归结到模糊描述逻辑；(2) 领域问题比较集中，可以使用转换字典将自然语言的问题和知识转换为 PFL 格式。本文用一个电影推荐网站的常用问答系统作为一个实例说明 PNL 式模糊问答系统的工作过程。

网站知识库中存储着从 Web 上收集到的关于电影的知识。这些知识除了发行时规定的标准说明外，也包含着观众的个人认知，而这些认知大多可以用模糊描述逻辑中的声明来表示。有了 PFL 格式的知识后，将 PFL 格式的问题和知识提交到推理机，推理机在 DDB 的规则中查找，找到下式的模糊规则。

$$\frac{(X \text{ is } A \wedge B) \geq n}{(X \text{ is } A) \quad n}$$

$$(X \text{ is } B) \quad n$$

同时查找常识知识库，有这样的规则：当 $n_1 \leq n$ 、 $X \text{ is } A \geq n_1$ 时，则认为 $(X \text{ is } A) \geq n$ 。这个问题完全可以转换为模糊描述逻辑的推理问题来解决，根据解决结果生成自然语言问题，即“file(《国家宝藏》)is 好电影 喜剧 0.5”是成立的，所以答案就是“《国家宝藏》是好喜剧电影”。表 5 表示了自然语言格式的知识到 PFL 格式的转换。

表 5 知识转换

	自然语言	P*格式	PFL
标准说明	中文片名：《国家宝藏》 英文片名：《National Treasure》 类型：喜剧	file(《国家宝藏》)is 喜剧 1	$X \text{ is } A \geq n_1$
观众认知	《国家宝藏》是好电影 我喜欢电影《国家宝藏》 《国家宝藏》不错 《国家宝藏》真好	(file(《国家宝藏》)is 好电影) 0.5	$X \text{ is } B \geq n$
问题	《国家宝藏》是好喜剧电影吗	file(《国家宝藏》)is 好电影 喜剧 0.5 ?	$(X \text{ is } A \wedge B) \geq n$

4 总结

本系统以模糊描述逻辑为基础，将 PNL 结构中的 DDB(Deduction Database)包含的原型语言规则，转换成了模糊描述逻辑的表示形式。并将收集到的知识以相同的形式保存在 WKDB(World Knowledge Database)。这样就可以利用模糊描述逻辑的推理算法来完成实际问答。

进一步的工作就是实现模糊推理机，并结合其他领域的推理机完成更复杂的问答任务，另一项具有挑战性的研究是结合特定领域建立自然语言和 PFL 格式的转换字典。

参考文献

- Zhong N, Wu J L, Liu C. Building a Data Mining Grid for Multiple Human Brain Data Analysis[C]. Proc. of International Workshop on Knowledge Grid and Grid Intelligence, Halifax, Canada, 2003: 29-35.
- Zhong N. Representation and Construction of Ontologies for Web Intelligence[J]. International Journal of Foundations of Computer Science, 2002, 13(4): 555-570.
- Zadeh L A. A Note on Web Intelligence[J]. World Knowledge and Fuzzy Logic, Data & Knowledge Engineering, 2004, 50(3): 291-304.
- Zadeh L A. Precisiated Natural Language[J]. AI Magazine, 2004, 25(3): 74-91.
- Zadeh L A. A New Direction in AI—Toward a Computational Theory of Perceptions[J]. AI Magazine, 2001, 22(1): 73-84.
- Straccia U. Reasoning Within Fuzzy Description Logics[J]. Journal of Artificial Intelligence Research, 2001, 14(1): 137-166.
- Straccia U. A Fuzzy Description Logic[C]. Proceedings of the 15th National Conference on Artificial Intelligence, Madison, USA, 1998: 594-599.

(上接第 14 页)

参考文献

- Hand D, Mannila H, Smyth P. 数据挖掘原理[M]. 北京: 机械工业出版社, 2003.
- 何友全, 肖建, 黄碧霞等. 一种用于数据挖掘的二进制挖掘算法[J]. 计算机应用研究, 2004, 21(5): 15-16.
- 颜跃进, 李舟军, 陈火旺. 一种挖掘最大频繁项集的深度优先算

法[J]. 计算机研究与发展, 2005, 42(3): 462-467.

- Vincent To-Yee Ng, Jacky Man-Lee Wong, Bao P. Incremental Mining of Association Patterns on Compressed Data[C]. Proc. of IFSA World Congress and the 20th NAFIPS International Conference, Vancouver, 2001: 441-446.
- 毛国君, 刘椿年. 基于项目序列集操作的关联规则挖掘算法[J]. 计算机学报, 2002, 25(4): 417-422.