自然语言中不确定性的语法逻辑

聂文龙 [1]

(1.中山大学逻辑与认知研究所，广东 广州 510275)

摘要： 本文意图将 Lambek 演算 LC 中的范畴概念推广到模糊模糊模糊范畴，从而使 LC 推广为模糊 Lambek 演算 FLC，并给出易于应用的简化演算 FLC*。通过示例的成功应用，说明 FLC 是 LC 的一种合适的推广。。

关键词：范畴；演算；Lambek

中图分类号： B81 **文献标识码：**A

人们在使用自然语言的过程中，不可避免地遇到不确定性问题，其原因是多方面的。不完全的知识、不同源的知识是最主要的原因。反映在语言的语义上有岐义现象；反映在语言的语形上有多种词汇或短语的组合构造。自从 J.Lambek 给我们带来了第一个语法逻辑，之后结合 Montague 的内涵逻辑形成了范畴（类型）逻辑语法以来，不确定性对逻辑语法作用的研究现在仍然十分少见。这是由于在实际的语言计算中，处理不确定性的方法不够系统化，所考虑的计算问题只是局部性的，缺乏对整体的考虑（请参考相关的资料，不难产生出这一印象）。在近来对不确定知识的表达与推理的理论研究已经成果极为丰富的条件下，给逻辑语法研究领域引入对不确定性问题的研究，是实际的需要，也是理论进一步发展的需要。本文试图讨论不确定性对语法逻辑的作用，并且暂时只限于对不确定现象中的模糊现象，目的是为自然语言的分析计算提供一个系统的、整体的逻辑理论基础。

Lambek 演算 LC 是第一个关于句法的逻辑证明系统，它使用范畴作为逻辑公式并且采用 Gentzen 风格的后承演算形式，构造了一个亚结构逻辑系统（参见文献[3]）。Moortgat 研究了 LC 中的范畴取并运算的内部不确定结构形式（文献[4].），从而对 LC 作出了在不确定信息应用场合的一定程度的推广。从另一个角度看，范畴整体也有不确定的表现形式，具体表现在对语言的词汇作范畴指派时，可能有多个或者模糊的指派情况出现，这样就对范畴提出一个问题：应该允许多个或者模糊的范畴参与运算。但就作者所知，迄今为止，还没有研究者考虑 LC 在范畴是模糊的情况下的推广研究，本文意图将 LC 中的范畴概念推广到模糊范畴，从而使 LC 演算推广为模糊 Lambek 演算。

经典的 LC 演算有各种等价的版本，下面是文献[2]中给出的 LC 推理规则形式：

$$\frac{\Delta \to B \quad \Gamma_1, A, \Gamma_2 \to C}{\Gamma_1, A/B, \Delta, \Gamma_2 \to C} \ (/L) \qquad \frac{\Delta \to B \quad \Gamma_1, A, \Gamma_2 \to C}{\Gamma_1, \Delta, B\backslash A, \Gamma_2 \to C} \ (\backslash L)$$

收稿日期：2005-01-10

基金项目：教育部哲学社会科学研究重大课题攻关项目资助。

作者简介：聂文龙(1963-)，男，中山大学逻辑与认知研究所，副教授，博士

$$\frac{\Gamma, A \to B}{\Gamma \to B/A} \quad (/R) \qquad \frac{A, \Gamma \to B}{\Gamma \to A \backslash B} \quad (/R)$$

其中大写希腊字母如Γ，Γ_1，Δ和Γ_2等表示范畴序列；英文字母 A、B 等表示范畴。在这些范畴及其序列推广为模糊范畴和模糊序列的情况下，推理规则等应取何种形式就是本文研究的目的。

1 模糊 L ambek 演算 FLC

设论域 T={p_1,p_2,……}是范畴语言，<T, ο, /, \>是范畴代数(参见文献 1.)，L_B=<{0,1},∨, ∧, ¬>是经典布尔代数。

函数μ:T→ L_B是 T 上模糊子集，称为模糊类型。记所有模糊类型的集合为 FT，其中有空类型 0，0 表示映射μ: μ(a)=0，对所有 a∈T。

FT 上运算ο、/、\ 的定义如下(在不引起混淆的情况下，使用与 T 上运算一样的符号)。

定义 设 f1,f2∈FT，f1οf2，f1/f2，f1\f2 满足对任意的 a∈T

$$f1 \circ f2(a) = \vee_{a=a1 \circ a2} (f1(a1) \wedge f2(a2))\ ;$$

$$f1\ /\ f2(a) = \vee_{a=a1/a2} (f1(a1) \wedge f2(a2))\ ;$$

$$f1 \backslash f2(a) = \vee_{a=a1\backslash a2} (f1(a1) \wedge f2(a2))\ ;$$

还有经典的模糊子集的包含关系⊆，显然，一般地 f1οf2 ≠ f2οf1，即运算ο不可交换。

定义 三元关系 R(f,f1,f2)成立，如果 f=f1οf2,且 f≠0。

由此，我们得到模糊 Lambek 演算 FLC = < FT, ⊆ , R, ο, /, \ , Infer >，其中推理规则的集合 Infer 在下面给出。对于 FT，设Γ是 FT 的元素的序列，f ∈ FT，那么我们有

定义 称 Γ |- f 为后承，如果 f ≠ 0。

现在，我们给出 FLC 的后承演算推理规则 Infer。一般地，我们有下述对 LAMBEK 演算 LC 直接推广的 Gentzen 风格的推理规则：

$$f \mid\!- f \qquad [id]$$

$$\frac{f \subseteq g}{f \mid\!- g} \qquad [\subseteq]$$

$$\frac{\Gamma \mid\!- f1 \quad R(f2, f, f1) \quad \Delta, f2, \Pi \mid\!- C}{\Delta, f, \Gamma, \Pi \mid\!- C} \qquad [\circ\text{-L}]$$

$$\frac{\Gamma \mid\!- f1 \quad R(f2, f1, f) \quad \Delta, f2, \Pi \mid\!- C}{\Delta, \Gamma, f, \Pi \mid\!- C} \qquad [\text{-}\circ\text{L}]$$

$$\frac{\Gamma, f1 \mid\!- f2 \quad R(f2,f,f1)}{\Gamma \mid\!- f} \qquad [\circ\text{-R}]$$

$$\frac{f1, \Gamma \mid\!- f2 \quad R(f2,f1,f)}{\Gamma \mid\!- f} \qquad [\text{-}\circ\text{R}]$$

Cut 规则

$$\frac{\Gamma \vdash f \quad \Delta, f, \Pi \vdash C}{\Delta, \Gamma, \Pi \vdash C} \quad [Cut]$$

上述推理规则中，R的出现使推理难度增大，下一节提出一种在推理规则中不包含的R的简化方法，于是形成演算 FLC*。

2 演算 FLC*

对于 LC 中的任一范畴 $a \in T$，我们给出左论元范畴、右论元范畴和值范畴的定义：
设 a 是任何范畴，$a \in T$，作如下规定

larg(a) = b，如果 a = b\ c；

rarg(a) = b，如果 a = c/b；

val(a) = c，如果 a = b\ c 或 a = c/b。

通过对 T 中范畴的左（右）论元范畴和值范畴定义，可以定义 FT 中的模糊范畴的模糊左（右）论元范畴和模糊值范畴。设 larg*(f)、rarg*(f)与 val*(f)分别表示 f 的模糊左论元范畴、模糊右论元范畴和模糊值范畴，它们都可看作是 f 的模态算子，定义如下：

larg*(f) = $g \in FT$: 对任何 $a \in T$，g(a)=1，如果存在 $b \in T$，f(b) = 1，a = larg(b)；否则 g(a) = 0；

rarg*(f) = $g \in FT$: 对任何 $a \in T$，g(a)=1，如果存在 $b \in T$，f(b) = 1，a = rarg(b)；否则 g(a) = 0；

val*(f) = $g \in FT$: 对任何 $a \in T$，g(a)=1，如果存在 $b \in T$，f(b) = 1，a = val(b)；否则 g(a) = 0。

将推理规则

$$\frac{\Gamma \vdash f1 \quad R(f2, f, f1) \quad \Delta, f2, \Pi \vdash C}{\Delta, f, \Gamma, \Pi \vdash C} \quad [\circ\text{-}L]$$

简化为

$$\frac{\Gamma \vdash rarg^*(f) \quad \Delta, val^*(f), \Pi \vdash C}{\Delta, f, \Gamma, \Pi \vdash C} \quad [\circ\text{-}L]*$$

类似地，推理规则

$$\frac{\Gamma \vdash f1 \quad R(f2,f1,f) \quad \Delta, f2, \Pi \vdash C}{\Delta, \Gamma, f, \Pi \vdash C} \quad [\text{-}\circ L]$$

简化为

$$\frac{\Gamma \vdash larg^*(f) \quad \Delta, val^*(f), \Pi \vdash C}{\Delta, \Gamma, f, \Pi \vdash C} \quad [\circ\text{-}L]*$$

规则

$$\frac{\Gamma, f1 \vdash f2 \quad R(f2,f,f1)}{\Gamma \vdash f} \quad [\circ\text{-}R]$$

简化为

$$\frac{\Gamma, f1 \vdash f2}{\Gamma \vdash f2 / f1} \quad [\circ\text{-}R]*$$

规则

$$\frac{f1, \Gamma \vdash f2 \quad R(f2,f1,f)}{\Gamma \vdash f} \quad [\text{-}\circ R]$$

简化为

$\underline{\Gamma,\ f1 \mid\text{-}\ f2}$ [o-R]*
$\Gamma \mid\text{-}\ f1 \backslash f2$

通过以上简化，我们得到易于应用的简化 FLC，称之为 FLC*。

例子说明

考虑一个语言学的应用例子。选择范畴的集合 T 由原子范畴{n，np，s}在运算o，/，\ 下递归生成（文献[2]），n:名词性范畴，np:名词短语范畴，s:句子范畴。
现在应用演算 FLC*分析句子“张三跑步”，按下述步骤：
(1) 指派(用符号?表示)模糊范畴和序列
“张三”? f，“跑步”? g，其中
f(a) = 1，当 a = np;
0，其他情况。
g(a) = 1，当 a = np\s;
0，其他情况。
“张三跑步” ? f，g
(2) 后承的证明
设句子的模糊范畴为 sent，满足
sent(a) = 1，当 a = s;
0，其他情况。
构造后承 f,g |- sent，给予证明。使用逆向证明搜索策略，我们有如下证明树：
$\underline{f,g \mid\text{-}\ sent}$ [-oL]
f |- f h |- sent
其中 h1 = val*(g)，按 val*的定义可计算出 h1(s)=1；当 a ≠ s 时，h1(a) = 0。故有 h1 = sent，结果得
$\underline{f,g \mid\text{-}\ sent}$ [-oL]
f |- f $\underline{h1 \mid\text{-}\ sent}$
sent |- sent
由此可见，证明成功。
(3) 结论
因 f,g |- sent 可证，所以句子是合法的。
以上所指派的范畴 f、g 实际上是等价于 LC 中所用的范畴，并不是真正模糊的。如果对“张三”、“跑步”的范畴指派是模糊的，那么可能有下面的分析情况：
(1) 指派(用符号?表示)模糊范畴和序列
“张三”? f，“跑步”? g，其中
f(a) = 1，当 a = np,n;
0，其他情况。
g(a) = 1，当 a = np\s,n\n;
0，其他情况。
“张三跑步” ? f，g
(2) 后承的证明
同样地，设句子的模糊范畴为 sent，满足
sent(a) = 1，当 a = s;
0，其他情况。
构造后承 f,g |- sent，给予证明。使用逆向证明搜索策略，我们有如下证明树：
$\underline{f,g \mid\text{-}\ sent}$ [-oL]
f |- f h1 |- sent
其中 h1 = val*(g)，按 val*的定义可计算出 h1(s)=1，h1(n)=1；当 a 是其他情况时，h1(a) = 0。故有 sent ⊆ h1，结果得

f,g |- sent [-oL]
f |- f h |- sent
sent |- h1 h1 |- sent [⊆]
sent |- sent

由此可见，证明成功。

(3) 结论

因 f,g |- sent 可证，所以句子是合法的。

在例子所示的真模糊情况下，我们仍然能够获得句子的合法性结论，这说明在不确定的信息下能够发现那些合法的句子，从而可以在大量的语言资源中，例如互联网上挖掘出有用的语言资源。

3 FLC*的语义和完全性定义

考虑在第 3 节给出的演算 FLC*的完全性问题，现更为集中地如下给出 FLC*的各个组成部分。

符号表

英文字母串 V ::= { A,B,C,f,g,…}；
算子 O ::= {rarg*, larg*, val*}；
其他符号 S ::= { |-, (,) }。

合适公式

φ ::= V | O(V) | V/V | V\V

序列

Seq ::= φ,φ,…, φ

后承

Con ::= Seq |- φ

公理

φ |- φ

推理规则

以下假定大写希腊字母代表 Seq 的任何成员；f, f1, f2, A, B, C 表示φ的任何成员；

$$\frac{\Gamma \;|\!-\; rarg^*(f) \qquad \Delta,\ val^*(f),\ \Pi \;|\!-\; C}{\Delta,\ f,\ \Gamma,\ \Pi \;|\!-\; C}\ [o\text{-}L]^*$$

$$\frac{\Gamma \;|\!-\; larg^*(f) \qquad \Delta,\ val^*(f),\ \Pi \;|\!-\; C}{\Delta,\ \Gamma,\ f,\ \Pi \;|\!-\; C}\ [o\text{-}L]^*$$

$$\frac{\Gamma,\ f1 \;|\!-\; f2}{\Gamma \;|\!-\; f2\ /\ f1}\ [o\text{-}R]^*$$

$$\frac{\Gamma,\ f1 \;|\!-\; f2}{\Gamma \;|\!-\; f1 \setminus f2}\ [o\text{-}R]^*$$

证明

后承的证明或称后承是可证的完全符合常规的证明定义。

设论域 W = {μ | μ: T → L_B}，代数框架 W = <W, ⊆, o, /, \, larg*, rarg*, val*>, 其中关系⊆和各个二元运算与一元运算的定义参见第 2、3 节。

定义　模型 M = < W, v >，其中解释函数 v：$\varphi \to W$ 满足下列条件：

（1）　v(f1\f2) = v(f1) \ v(f2)，v(f1/f2) = v(f1) / v(f2)。

（2）　v(val*(f)) = val*(v(f))，v(larg*(f)) = larg*(v(f))，v(rarg*(f)) = rarg*(v(f))。

设Γ = A，B,…，记$\overline{v}(\Gamma)$ = v(A) $\circ$v(B) $\circ$…，我们给出真与完全性的定义。

定义　真。如果$\overline{v}(\Gamma) \subseteq$ v(C)，称后承Γ |- C 在模型 M 中为真。

定义　完全性。如果若Γ |- C 在任何模型中为真，则Γ |- C 是可证明的，那么称演算 FLC* 是完全的。

上述演算的完全性研究，将是进一步的研究工作。

4　结论

本文提出 FLC 演算系统，是具有 Gentzen 风格的后承演算系统，其简化的演算 FLC*更易于应用。通过例子的演算，我们说明了 FLC 可以处理经典 LC 能够处理的范畴演算，也能够处理模糊范畴的演算，因此 FLC 是 LC 的一种合适的推广。进一步的研究工作可以考虑多模态的 MLC 演算系统（文献[5]）的模糊化问题，最终我们将在这一系列的语法逻辑向模糊语法逻辑的推广中，获得越来越强大的语言处理能力。

参考文献：

[1] Bar-Hillel Y. On Categorical and Phrase Structure Grammars. In: Language and Information, Reading: Addison-Wesley, 1964.

[2] Carpenter, B. Type-Logical Semantics. Cambridge/London, MIT Press. 1997.

[3] Lambek,J. The mathematics of sentence structure [J]. Amer. Math. Monthly. 1958, 65: 154-170.

[4] Moortgat M. Categorical grammar and formal semantics. In: the Encyclopedia of Cognitive Science, Nature Publishing Group, Macmillan Publishers Ltd. 2002.

[5] Van Benthem J. etc. Handbook of Logic and Language[A]. Moortgat M. Categorial Structure [C]. Amsterdam: Elsevier Science. 1997. 50-140.

The Grammar Logic of Uncertainty in Natural Language

Wenlong Nie[1]

(1. The Institute of Logic and Cognition, Zhongshan University, Guangzhou, 510275, China)

Abstract：The paper intended to give a generalization of category in LC to fuzzy category so that Lambek calculus LC is extended to a calculus FLC that takes fuzzy category into account and we provide a simplified calculus FLC*. We shown that FLC indeed is a rational generalization of LC by the success of processing some examples.

Key words：category; calculus; Lambek