

基于相似关系的社会集合论

李金屏 黄艺美

(济南大学信息科学与工程学院模式识别与智能系统研究所 济南 250022)

(山东省网络环境智能计算技术重点实验室 济南 250022)

摘要 在利用经典相容关系研究人际关系的基础上,引入了人际关系中的模糊属性。利用截集的概念,研究了满足模糊相容关系即相似关系下的人际关系社会中社会交际圈的划分、成员活跃度和成员完成各项任务的代价等,并给出了相关算法和讨论。计算结果表明,该模型能够有效描述满足相似关系的社会关系网络。

关键词 社会集合论,相似关系,社会交际圈,活跃度,代价

中图分类号 TP182 **文献标识码** A

Social Set Theory Based on Similarity Relations

LI Jin-ping HUANG Yi-mei

(Institute of Pattern Recognition and Intelligent System, School of Information Science and Engineering, University of Jinan, Jinan 250022, China)

(Shandong Provincial Key Lab of Network Based Intelligent Computing, Jinan 250022, China)

Abstract On the basis of the study of social relations by using classical similarity relation, we began to consider the fuzziness in the social relations. By using the concept of cut set, we studied many properties of social relations satisfying similarity relation (i. e., fuzzy tolerance relation), such as the segmentation of social relation circles, the activity degree of individuals in the social relations and the cost when an individual accomplishes various tasks. The corresponding algorithms were also presented. Simulations show the proposed model can describe the social relation network satisfying similarity relation effectively.

Keywords Social set theory, Similarity relation, Social relation circle, Activity degree, Cost

1 引言

现实生活中,物以类聚,人以群分。每个人、单位或者国家总是生活在一定的“圈子”里面,小至家庭,大至国家,有各种各样的“圈子”,例如家庭、棋友、球友、驴友、校友、相同信仰的人、北大西洋公约组织、阿拉伯国家联盟、社会主义国家、资本主义国家、世界贸易组织、各个学术团体(例如计算机学会、自动化学会、中国图象图形学学会、中国人工智能学会、电子学会等)。可以这样说,具有相同或者相似兴趣、共同利益等的人、机构和组织就构成了所谓的“圈子”。

随着互联网的发展,还出现了各种各样的虚拟社区。因此,社会计算和社会网络的研究迅速引起人们的关注^[1-4],其中一个重要的研究内容就是网络中每个成员影响力的评价。在文献[5]中,我们研究了满足经典相容关系的社会网络中成员社会交际圈(简称社交圈)的划分、成员活跃度和成员完成各项任务的平均代价的计算,取得了良好效果。

然而,实际的人际关系十分复杂,很多情况下具有模糊的属性,例如人与人之间的关系具有不同的程度,因人而异;在多个国家所组成的不同联盟和组织内部,不同成员之间的关系也是各不相同的。因此,有必要引入成员关系中的模糊属

性。

描述模糊属性的工具是模糊数学^[6]。本文将在经典关系的基础上引入模糊关系,进而利用截集的概念,通过经典的相容关系,研究模糊相容关系下的人际关系,包括社交圈的划分、活跃度和代价的计算。迄今为止,尚未见到与此相关的文献^[7]。

为了论文内容的完整性,本文首先给出关系和模糊集合的相关概念,尤其是给出成员活跃度和代价的定义;然后介绍基于截集的模糊相容关系下社交圈的划分、成员活跃度、平均代价和最大代价的计算;最后是结论。

2 基本概念和相关定义

设 R 是集合 U 上的一个经典相容关系,则 R 具有自反性和对称性。

定义 1 设 $F(U)$ 是 U 上模糊集的全体,则

$$F(U) = \{S | \mu_S : U \rightarrow [0, 1]\} \quad (1)$$

式中, $\mu_S(u)$ 是模糊集 S 的隶属度函数^[8]。

定义 2 对于模糊集合 $S \in F(U)$, $\lambda \in [0, 1]$, 称 $S_\lambda = \{u | \mu_S(u) \geq \lambda\}$ 为模糊集 S 的 λ 水平截集^[8], 其中 λ 称为 S 的一个截集水平。当 $\lambda > 0$ 时, $S_\lambda = \text{supp}[\mu_S(u)]$ 称为模糊集 S 的支

到稿日期:2012-07-11 返修日期:2012-11-09 本文受国家自然科学基金(61203341),山东省高等学校科技计划(J12LN19),济南大学科技发展计划(XKY1035)资助。

李金屏(1968—),男,博士,教授,硕士生导师,主要研究方向为图像处理、机器视觉、人工智能、模式识别等, E-mail: ise_lijp@ujn.edu.cn。

集;当 $\lambda=1$ 时, S_λ 称为模糊集 S 的核。

需要说明,截集是利用经典集合研究模糊集合的重要工具,是本文的主要工具。

定义3 在 $U \times U \times \dots \times U$ 上的一个 n 元模糊关系 $\tilde{R}$ 是指以 $U \times U \times \dots \times U$ 为论域的一个模糊集合,记为

$$\tilde{R} = \int_{U \times U \times \dots \times U} \mu_{\tilde{R}}(u_1, u_2, \dots, u_n) / (u_1, u_2, \dots, u_n) \quad (2)$$

式中, $\mu_{\tilde{R}}(u_1, u_2, \dots, u_n)$ 是模糊关系 $\tilde{R}$ 的隶属度函数,将 $U \times U \times \dots \times U$ 上的每一个成员 $(u_1, u_2, \dots, u_n)$ 映射为 $[0, 1]$ 上的一个实数,该实数反映出 n 元组 $(u_1, u_2, \dots, u_n)$ 具有关系 $\tilde{R}$ 的程度^[8]。本文只讨论 $n=2$ 的模糊关系。

定义4 设 $F(U \times U)$ 是 U 上二元模糊关系的全体^[8],则 $F(U \times U) = \{\tilde{R} | \mu_{\tilde{R}}: U \times U \rightarrow [0, 1]\}$ (3)

定义5 具有自反性和对称性的二元模糊关系 $\tilde{R} \in F(U \times U)$ 称为相似关系^[8]。其中,自反性表明 $(\forall u)(u \in U \rightarrow \mu_{\tilde{R}}(u, u) = 1)$,对称性表明 $(\forall u)(\forall v)(u \in U \wedge v \in U \rightarrow \mu_{\tilde{R}}(u, v) = \mu_{\tilde{R}}(v, u))$ 。

定义6 如果 $(\forall u_1)(\forall u_2)(u_1, u_2 \in T \rightarrow u_1 R u_2)$,则 $T \subseteq U$ 是一个相容类^[9]。

定义7 设 $T \subseteq U$ 是一个相容类,并且 T 无法包含在其他相容类中, T 称为一个最大相容类^[9]。

定义8 设 A 是一个映射,即 $A: U \rightarrow [0, 1]$ 。若满足下面条件中任何一个^[5]:

1) $A_1 = N/M$,其中 M 表示 U 中所有的最大相容类的数目, N 表示成员 $u \in U$ 所在的最大相容类的数目;

2) $A_2 = n/m$,其中 m 表示 U 中所有成员的数目, n 表示 U 中与 $u \in U$ 有直接关系的成员数目;

则 A_1 和 A_2 均称为成员 u 的活跃度。对于某成员 $u \in U$,为了体现两种活跃度的含义, u 的活跃度可以表示为一个有序对 $A(u) = (A_1(u), A_2(u))$ 。

定义9 对于 $u \in U$,其通过各种社会关系完成各项任务的代价的平均值称为 u 的平均代价 $C(u)$,其中的最大值称为最大代价 $C_{\max}(u)$ ^[5]。

3 基于模糊相容关系的社会集合论

文献^[5]研究了经典相容关系下社交圈的划分、成员活跃度和成员完成各项任务的平均代价等。事实上,人与人之间的社会关系非常复杂,具有很强的模糊特点,例如,在一个机构内部,不同成员之间的亲密关系就具有模糊的特点。

下面利用截集的概念和经典社会集合论的方法计算模糊关系下的人际关系。本文只讨论满足自反性和对称性的相似关系。

3.1 社交圈的划分

首先计算在不同截集水平下的社交圈的划分。

一般情况下,同一个社交圈内的人、单位或者国家在合作时会比较方便,而社交圈之外的人或机构在合作时相对来说难度就会大一些。在需要求助时,人们会首先想到社交圈内部的人和机构,当然有时候还需要考虑成本和便捷性。朋友多了好办事,显然,一个人或者机构所在的社交圈越多,该人、机构就具有越大的影响。不同的人其重要性是不同的。因此,研究社会网络中社交圈的划分至关重要。

在一定的截集水平下,模糊关系可以变成经典关系,于是

就可以利用经典的极大相容类算法来划分各个社交圈,进而计算该截集水平下每个成员的活跃度、平均代价和最大代价。

设 $\tilde{R} \in F(U \times U)$ 是 U 上一个满足自反性和对称性的相似关系。在截集水平 λ 下, $\tilde{R}$ 相应的经典关系 R_λ 定义为

$$R_\lambda = \{(u, v) | \mu_{\tilde{R}}(u, v) \geq \lambda\} \quad (4)$$

容易证明, R_λ 是一个经典相容关系。

于是就可以利用经典的极大相容类算法(参见文献^[5]中的算法1)计算每个成员的社交圈。

显然,随着截集水平的变化,每个成员所在的社交圈也会发生不同程度的变化。一般情况下,当截集水平很小时,多数成员处于同一个社交圈内。当截集水平不断提高时,社交圈会逐渐增多,而各个社交圈一般会变小。其中,某些成员所在社交圈的数量急剧减少,圈子规模也急剧减小;而另一些成员所在社交圈的数量不会明显减少,所在的社交圈规模也没有明显减小。显然,前者表明其成员社交能力偏弱,社会影响力较小,而后者所对应的成员社交能力较强,社会影响较大;换言之,前者活跃度小,后者活跃度大。

3.2 仿真计算和讨论

图1(a)是文献^[5]中的一个简单的例子,是一个小规模的社会网络,一共有9个成员和19对关系(由于相容关系具有自反性和对称性,因此每对连线只作一条线段,节点处的自回路也省略)。每条连线上的数字表示权重,对应两个成员之间某种关系的程度。

3.2.1 不同截集水平下的计算结果

当取 $\lambda=0.1, 0.2, 0.5$ 和 0.8 时,根据定义2,可以获得这4个截集水平下的经典社会关系(见图1)。

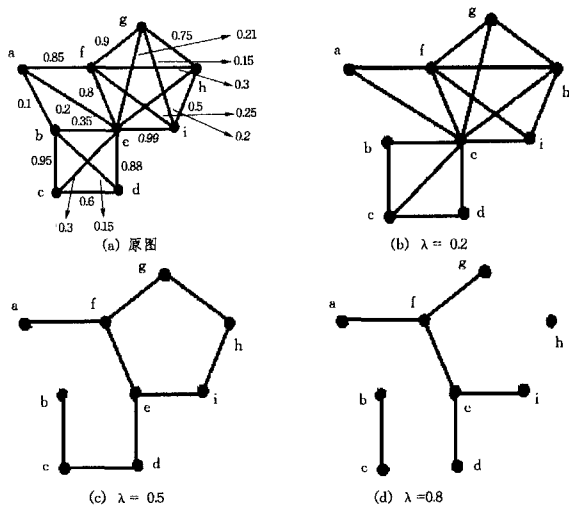


图1 一个小的社会关系网络及其不同截集水平下的社会关系

1) $\lambda=0.1$

社会网络结构没有任何变化,只是由模糊关系变成了经典相容关系,即将图1(a)中所有不等于0的权重置为1。此时,有最大相容类4个:

$$\{a, b, e\}, \{a, e, f\}, \{b, c, d, e\}, \{e, f, g, h, i\}$$

此时,根据活跃度的定义8,可以得到9个成员的活跃度:

$$A(a) = (2/4, 4/9); A(b) = (2/4, 5/9);$$

$$A(c) = (1/4, 4/9); A(d) = (1/4, 4/9);$$

$$A(e) = (4/4, 9/9); A(f) = (2/4, 6/9);$$

$$A(g) = (1/4, 5/9); A(h) = (1/4, 5/9);$$

$$A(i)=(1/4, 5/9)。$$

根据代价的定义 9, 各个成员通过社会关系履行某项任务时, 通常会首选路径最短的路径。例如在图 1(a) 中, 成员 a 求助成员 f , 会直接通过 (a, f) , 一般不会经过 $(a, e)(e, f)$ ¹⁾。因此, 在最短路意义下 (见图 2), a 和 b 的平均代价和最大代价分别为

$$C(a)=\frac{1}{N-1}\sum_{i=1}^n w_i L(w_i)=1.625, C_{\max}(a)=2;$$

$$C(b)=1.50, C_{\max}(b)=2。$$

式中, N 表示所有成员的个数, n 表示在最短树中与树根相关的成员个数, w_i 表示权重, 由于该相容关系是取交集之后的经典关系, 因此 $w_i=1$ 。

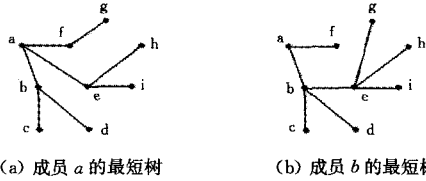


图 2 图 1 所示例子中与成员 a 和 b 对应的最短树

这里需要说明, 最短路算法利用了 Dantzig 算法^[5,9]。

其他成员平均代价和最大代价分别为:

$$C(c)=1.625, C_{\max}(c)=2;$$

$$C(d)=1.625, C_{\max}(d)=2;$$

$$C(e)=1.0, C_{\max}(e)=1;$$

$$C(f)=1.375, C_{\max}(f)=2;$$

$$C(g)=1.50, C_{\max}(g)=2;$$

$$C(h)=1.50, C_{\max}(h)=2;$$

$$C(i)=1.50, C_{\max}(i)=2。$$

具体计算过程请见文献[5], 不再赘述。

$$2)\lambda=0.2$$

截集之后的社会网络结构发生一些变化。此时的最大相容类有 5 个:

$$\{a, e, f\}, \{b, c, e\}, \{c, d, e\}, \{e, f, g, h\}, \{e, f, h, i\}$$

此时, 根据活跃度的定义, 可以继续计算 9 个成员的活跃度:

$$A(a)=(1/5, 3/9); A(b)=(1/5, 4/9);$$

$$A(c)=(2/5, 4/9); A(d)=(1/5, 3/9);$$

$$A(e)=(5/5, 9/9); A(f)=(3/5, 6/9);$$

$$A(g)=(1/5, 5/9); A(h)=(2/5, 5/9);$$

$$A(i)=(1/5, 5/9)。$$

各个成员的代价分别为:

$$C(a)=1.75, C_{\max}(a)=2;$$

$$C(b)=1.75, C_{\max}(b)=2;$$

$$C(c)=1.625, C_{\max}(c)=2;$$

$$C(d)=1.75, C_{\max}(d)=2;$$

$$C(e)=1.0, C_{\max}(e)=1;$$

$$C(f)=1.375, C_{\max}(f)=2;$$

$$C(g)=1.625, C_{\max}(g)=2;$$

$$C(h)=1.50, C_{\max}(h)=2;$$

$$C(i)=1.625, C_{\max}(i)=2。$$

$$3)\lambda=0.5$$

社会网络结构进一步发生变化。此时的最大相容类有 9 个:

$$\{a, f\}, \{f, e\}, \{f, g\}, \{g, h\}, \{h, i\}, \{e, i\}, \{d, e\}, \{c, d\}, \{b, c\}$$

继续计算 9 个成员的活跃度:

$$A(a)=(1/9, 2/9); A(b)=(1/9, 2/9);$$

$$A(c)=(2/9, 3/9); A(d)=(2/9, 3/9);$$

$$A(e)=(3/9, 4/9); A(f)=(3/9, 4/9);$$

$$A(g)=(2/9, 3/9); A(h)=(2/9, 3/9);$$

$$A(i)=(2/9, 3/9)。$$

各个成员的代价分别为

$$C(a)=2.875, C_{\max}(a)=5;$$

$$C(b)=3.625, C_{\max}(b)=5;$$

$$C(c)=2.75, C_{\max}(c)=4;$$

$$C(d)=2.125, C_{\max}(d)=3;$$

$$C(e)=1.75, C_{\max}(e)=3;$$

$$C(f)=2.0, C_{\max}(f)=4;$$

$$C(g)=2.5, C_{\max}(g)=5;$$

$$C(h)=2.625, C_{\max}(h)=5;$$

$$C(i)=2.25, C_{\max}(i)=4。$$

$$4)\lambda=0.8$$

社会网络结构发生了更大变化, 出现了不连通现象。此时的最大相容类有 7 个:

$$\{a, f\}, \{f, e\}, \{f, g\}, \{e, i\}, \{d, e\}, \{b, c\}, \{h\}$$

继续计算 9 个成员的活跃度:

$$A(a)=(1/7, 2/9); A(b)=(1/7, 2/9);$$

$$A(c)=(2/7, 2/9); A(d)=(1/7, 2/9);$$

$$A(e)=(3/7, 4/9); A(f)=(3/7, 4/9);$$

$$A(g)=(1/7, 2/9); A(h)=(1/7, 1/9);$$

$$A(i)=(1/7, 2/9)。$$

由于出现了不连通, 因此每一个成员的平均代价和最大代价都是无穷大(∞)。

3.2.2 讨论

下面分析一些成员的活跃度和代价随着截集水平的变化规律。

在截集水平 $\lambda=0.1, 0.2, 0.5$ 和 0.8 时, 对于成员 a 而言, 其活跃度分别是 $(2/4, 4/9)$ 、 $(1/5, 3/9)$ 、 $(1/9, 2/9)$ 和 $(1/7, 2/9)$, 其代价则是 $1.625, 1.75, 2.85$ 和 ∞ ; 对于成员 f 而言, 其活跃度分别是 $(2/4, 6/9)$ 、 $(3/5, 6/9)$ 、 $(3/9, 4/9)$ 和 $(3/7, 4/9)$, 其代价则是 $1.375, 1.375, 2.0$ 和 ∞ ; 对于成员 e 而言, 其活跃度分别是 $(4/4, 9/9)$ 、 $(5/5, 9/9)$ 、 $(3/9, 4/9)$ 和 $(3/7, 4/9)$, 其代价则是 $1.0, 1.0, 1.75$ 和 ∞ 。可以看出, 一般情况下, 成员活跃度呈明显减小的趋势, 而代价则呈单调递增的趋势。这是自然的, 因为随着截集水平的提高, 成员之间的联系会逐步减少, 成员所在的圈子数量通常会逐步减少, 所以其活跃度会呈下降的趋势 (根据定义 8, D_2 必定单调下降), 其完成各项任务的能力也必将随之下降, 而代价则随之增加。

一般情况下, 不同截集水平下的社会网络结构适合于完

¹⁾ 这是在考虑截集之后的网络结构。在考虑实际权重时, 需要另外考虑。例如, 虽然 a 与 e 直接相连, 但是权重太小, 反而通过 $\langle a, f \rangle$ 和 $\langle f, e \rangle$ 有更高效率和更低代价。

成不同层次的任务。

1) 活跃度

为了更加准确地描述成员的活跃度,可以利用成员活跃度随着不同截集水平的变化函数,即 $A(u)$ 随着 λ 的变化曲线,将其记作 $A_u(\lambda)$ 。由于 $A(u)$ 有两种定义方式,因此活跃度变化函数也有两个: $A_{u1}(\lambda)$ 和 $A_{u2}(\lambda)$ 。

显然,只利用成员活跃度变化函数尚无法对一个成员的活跃度进行总体描述。因此我们利用对归一化的活跃度变化函数的积分进行描述,即

$$A(u) = \int_0^1 A_u(\lambda) d\lambda \quad (5)$$

在具体计算时,可以利用求和取代积分。在要求不高的情况下,还可以只对部分截集水平下的活跃度进行求和。例如,利用前述 9 个成员的活跃度计算结果,可以获得如下 9 个成员的归一化活跃度求和结果(只给出第一种活跃度定义下的计算结果):

$$A(a) = 0.238; A(b) = 0.238;$$

$$A(c) = 0.289; A(d) = 0.204;$$

$$A(e) = 0.69; A(f) = 0.465;$$

$$A(g) = 0.204; A(h) = 0.254;$$

$$A(i) = 0.204。$$

显然,9 个成员活跃度可以排序如下:

$$1. A(e) = 0.69; 2. A(f) = 0.465;$$

$$3. A(c) = 0.289; 4. A(h) = 0.254;$$

$$5. A(a) = 0.238; 6. A(b) = 0.238;$$

$$7. A(d) = 0.204; 8. A(g) = 0.204;$$

$$9. A(i) = 0.204。$$

这与我们观察图 1 的结果基本上是一致的。如果考虑到活跃度的第二种定义 D_2 ,就可以将上述的相同次序区分得更加仔细。

2) 代价

同样,也可以利用代价随着截集水平的变化即 $C(u)$ 随着 λ 的变化来更加准确地描述成员代价,将其记作 $C_u(\lambda)$ 。当然这种描述在网络结构出现不连通之前才有意义,我们暂时不讨论不连通情况下的代价计算。

显然,只利用 $C_u(\lambda)$ 尚无法对一个成员的活跃度进行总体描述。同样可以利用对归一化代价函数的积分来进行描述,即

$$C(u) = \int_0^{\lambda_0} C_u(\lambda) d\lambda \quad (6)$$

式中, λ_0 表示出现不连通网络结构之前的最大截集水平。在具体计算时,也可以利用求和取代积分。在要求不高的情况下,还可以只对部分截集水平下的代价进行求和。例如 9 个成员的归一化代价求和结果如下(只有 3 个截集水平 $\lambda = 0.1, 0.2$ 和 0.5):

$$C(a) = 2.08; C(b) = 2.29; C(c) = 2.0;$$

$$C(d) = 1.83; C(e) = 1.25; C(f) = 1.58;$$

$$C(g) = 1.875; C(h) = 1.875; C(i) = 1.79。$$

显然,办事代价从小到大的排序结果是:

$$1. C(e) = 1.25; 2. C(f) = 1.58;$$

$$3. C(i) = 1.79; 4. C(d) = 1.83;$$

$$5. C(g) = 1.875; 6. C(h) = 1.875;$$

$$7. C(c) = 2.0; 8. C(a) = 2.08;$$

$$9. C(b) = 2.29。$$

可以看出,上述结果与活跃度的排序总体上一致。例如,成员 e 的活跃度最大,其代价也最小;成员 f 的活跃度也比较大,其代价也比较小;成员 b 的活跃度比较小,其代价也很大。当然,活跃度与代价的排序结果并非严格对应,这是需要注意的地方。

当然,对一个成员的影响力和能力进行排序有多种方法。

结束语 本文引入了人际关系的模糊属性,在截集方法的基础上,利用经典相容关系初步研究了满足相似关系的社会网络,包括网络中社交圈的划分、成员活跃度、完成各项任务的平均代价和最大代价等。通过仿真计算,取得了符合人们常识的研究结果。

需要说明,这种研究尚处于起步阶段,有许多问题需要解决,例如成员活跃度与平均代价之间究竟有没有一些量化的关系,有没有更好的、科学合理的描述活跃度和代价的方法和计算公式等。对于代价公式而言,就存在一些问题:在连通情况下,该公式问题不大;但是在不连通的情况下,尚未探讨合适的代价计算公式。

人们关心的一个重要问题是如何获得具有模糊相容关系的人际关系。一般情况下,形式可以多样,问卷调查、网络投票等都是可行的方法。

此外,人际关系是极为复杂的,许多关系并不对称,甚至有些关系的权值可以是负值。如何研究这种更加复杂的人际关系中的社交圈的划分、成员活跃度、代价以及成员综合影响力等都是值得研究的问题,也是我们下一步的研究课题。

参 考 文 献

- [1] 刘挺. 社会计算[J]. 中国计算机学会通讯, 2011, 7(12): 6-7
- [2] 毛文吉, 曾大军, 王飞跃. 社会计算的研究现状与未来[J]. 中国计算机学会通讯, 2011, 7(12): 8-11
- [3] 程学旗, 沈华伟. 社会信息网络中的社区分析[J]. 中国计算机学会通讯, 2011, 7(12): 12-19
- [4] Yager R R, Reformat M Z. Determining Affinity of Users in Social Networks Using Fuzzy Sets[J]. Communications in Computer and Information Science, Advances in Computational Intelligence, 2012, 298(2): 149-160
- [5] Li Jin-ping. Study on Social Relations between Individuals Based on Graph Theory[C]// The 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'07). 2007: 423-427
- [6] Zadeh L A. Fuzzy sets [J]. Information and Control, 1965, 8: 338-353
- [7] Chen M-S. Information Processing in Social Networks[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12). 2012: 1403-1403
- [8] 李洪兴, 汪培庄. 模糊数学[M]. 北京: 国防工业出版社, 1994
- [9] 左孝凌, 李为鑑, 刘永才. 离散数学[M]. 上海: 上海科学技术出版社, 2002