变精度粗糙集推广模型及其性质研究*

范年柏[a]，张 宁[a]，孙 涛[b]
（湖南大学 a. 信息科学与工程学院；b. 数学与计量经济学院，长沙 410082）

摘　要：数据挖掘的主要目标之一是进行有效分类，粗糙集的上下近似空间正是为了对信息系统进行分类。变精度粗糙集作为经典粗糙集的推广模型，目前研究仅局限于有限集。针对变精度粗糙集模型无法处理无限集合的问题，在变精度粗糙集和测度的理论基础上，提出了基于 Lebesgue 测度的变精度粗糙集模型。引入 Lebesgue 测度的概念，构造了一种基于 Lebesgue 测度的变精度粗糙集模型，将变精度粗糙集理论推广到无限集，定义了该模型的上、下近似空间，并证明了其相关性质。通过理论研究表明，该模型能有效处理无限集合问题，对变精度粗糙集的理论研究形成突破，也将极大地扩充其应用范围。

关键词：粗糙集；Lebesgue 测度；变精度粗集；数据挖掘
中图分类号：TP391　　**文献标志码：**A　　**文章编号：**1001-3695(2018)05-1399-04
doi:10.3969/j.issn.1001-3695.2018.05.025

Study on extensions of variable precision rough set and its proposition

Fan Nianbai[a], Zhang Ning[a], Sun Tao[b]
(*a. College of Computer Science & Eletronic Engineering, b. College of Mathematics & Econometric, Hunan University, Changsha* 410082, *China*)

Abstract: One of the main goals of data mining is to make an effective classification, and the upper and lower approximation space of rough set is to classify the information system. The variable precision rough set is an extended model of rough set, however the research of the variable precision rough set has been limited to the finite set so far. To address the problem the variable precision rough set could not deal with the infinite set, this paper proposed the variable precision rough set model based on Lebesgue measure on the basis of the variable precision rough set and measure theory. Firstly, this paper introduced the Lebesgue measure, and it constructed the variable precision rough set model based on Lebesgue measure which extended the variable precision rough set to the infinite set. Secondly, it defined the upper and lower approximation space of the new model and provided the properties. The theoretical study show that the new model can deal with the infinite set effectively. It will make a breakthrough in the theory of the variable precision rough sets, and greatly expand its scope of application.

Key words: rough set; Lebesgue measure; variable precision rough set; data mining

随着信息的增加，数据挖掘需要处理的数据越来越多且具有强烈的不确定性[1,2]。数据挖掘的主要目标是对数据进行有效分类，粗糙集是进行数据分类的主要工具之一，也能处理不确定、具有噪声的数据。自 Pawlak[3] 于 20 世纪 80 年代研究粗糙集以来，很多学者对其进行了大量的研究。为了提高分类效率，Ziarko[4] 引入了变精度粗糙集，变精度粗糙集是对经典粗糙集的推广，它允许一定的错误分类率存在[5,6]。粗糙集理论的研究及应用主要集中在有限集[7~9]，然而实际应用中会涉及到无限集，例如天气预报中强台风覆盖的范围是不确定的，且是无限集；体育竞技中射击项目，测量命中环数不是一个点而是一个区域等，这些用传统粗糙集理论已经无法解决。

Liu[10,11] 将经典粗糙集从有限集推广到无限集，并对粗糙集进行了公理化刻画；刘耀峰[12] 提出了基于 Sugeno 测度的粗糙集模型和基于 Sugeno 测度的变精度粗糙集模型；薛占熬等人[13] 将等价关系推广到覆盖，提出了基于覆盖的 Sugeno 测度的粗糙集模型。目前对变精度粗糙集的研究主要局限于有限集[14,15]，这极大地限制了变精度粗糙集的应用，因此有必要对变精度粗糙集模型进行推广。本文首先介绍一个基于有界无限集合的变精度粗糙集模型——基于 Lebesgue 测度[16] 的变精度粗糙集模型，进而对相关性质进行了充分的研究。

1　基础知识介绍

粗糙集（rough set）是波兰学者 Pawlak 于 1982 年提出的一种数学理论，目前被认为是数据挖掘技术中一种有效和常见的方法。它的基础是模糊集和概率论，采用等价关系将所研究的论域粒化为若干个互斥的等价类，是一种用于处理不精确、不一致、不完整数据的一种有效的数据分析工具。后来许多学者将等价关系推广到一般二元关系，相应的应用范围也得到大大的扩展。但是 Pawlak 教授提出的传统粗糙集理论只能处理名义型或符号型的数据，对现实生活中大量存在的数值型或连续型数据就无法直接处理。本文讨论的问题可以是连续性数据，并仍然以等价关系作为讨论对象。

1.1　粗糙集理论

定义 1　在信息系统 $S=(U,A,V,f)$ 中，对象集合 $X\subseteq U$，属性集合 $B\subseteq A$，则 X 对于 B 的上近似集和下近似集为

收稿日期：2016-12-25；**修回日期：**2017-03-02　**基金项目：**湖南省科技计划应用基础研究重点项目(2016JC2014)

作者简介：范年柏(1962-)，男，湖南澧县人，副教授，博士，主要研究方向为形式化方法、数据挖掘(nbfan6203@163.com)；张宁(1990-)，女，山东枣庄人，硕士，主要研究方向为数据挖掘；孙涛(1987-)，男，湖南岳阳人，博士研究生，主要研究方向为拓扑学.

$$\overline{B}(X)=\{x\in U|[x]\cap X\neq\varnothing\}$$
$$\underline{B}(X)=\{x\in U|[x]\subseteq X\}$$

此外，X 的 B 边界域定义为 $\mathrm{bnr}_B(X)=\overline{B}(X)-\underline{B}(X)$，它展示的是上近似和下近似的差值。边界又称为模糊区域，反映了集合的粗糙程度，也就是说，边界越大，集合就越粗糙，越不能精确地表示；相反，边界越少，集合就越精确。X 的 B 正域定义为 $\mathrm{POS}_B(X)=\underline{B}(X)$，即 X 关于属性集合 B 的下近似；X 的 B 负域定义为 $\mathrm{NEG}_B(X)=U-\overline{B}(X)$。属性 B 对应一个等价关系 R，元素 x 相应的等价类记为$[x]$，为了简洁起见省略下标 B。

推论1　对于全集 U 及其等价关系 R，X 是其子集。

$$\overline{R}(X)=\{x\in U|[x]\cap X\neq\varnothing\}=\cup\{[x]_R\in U/R|[x]\cap X\neq\varnothing\}$$
$$\underline{R}(X)=\{x\in U|[x]\subseteq X\}=\cup\{[x]_R\in U/R|[x]\subseteq X\}$$

1.2　测度理论

本节的结论来自文献[16]。

定义2　设 E 为 R^n 中任一点集，对于每一列覆盖 E 的开球 I_i，给出它的体积总和 $\mu=\sum_i|I_i|$，所有这一切的 μ 组成一个下方有界的数集，它的下确界（由 E 完全确定）称为 E 的勒贝格外测度，简称 L 外测度或外测度，即

$$m^*(E)=\inf_{E\subset\cup I_i}\sum_i|I_i|$$

E 的内测度 $m_*(E)=|I|-m^*(I-E)$。

如果 $m^*E=m_*E$，则称 E 可测，并记做 mE。

性质1　Lebesgue 外测度性质：

a）$m\varnothing=0$；

b）$m^*E\geqslant mE\geqslant 0$；

c）$A\subseteq B,m^*B\geqslant m^*A,mB\geqslant mA$；

d）次可数可加性 $m^*(\bigcup_{i=1}^{\infty}A_i)\leqslant\sum_{i=1}^{\infty}m^*A_i$。

定理1　E 是可测的当且仅当对于任意 A，

$$m^*A=m^*(A\cap E)+m^*(A\cap E^C)$$

推论2　点集 E 可测的充要条件是 $\forall A\subseteq E,B\subseteq E^C$，有 $m^*(A\cup B)=m^*A+m^*B$。

定理2　有界集 E 是可测的当且仅当对于任意 ε，$\exists$ 开集 G 和闭集 F，$F\subseteq E\subseteq G$，使得 $m(G-F)<\varepsilon$。

性质2　如果 A、B 可测，则 $A\cup B$、$A\cap B$、$A-B$ 也可测。

定理3　$\{E_i|i\in I\}$是可测序列。

a）如果渐增，即 $E_i\subseteq E_{i+1}$，则 $E=\cup_{i\in I}E_i$ 可测，并且 $mE=\lim\limits_{i\to\infty}mE_i$；

b）如果是渐减序列，并且有界，即 $E_i\supseteq E_{i+1}$，$mE_1<\infty$，则 $e=\bigcap\limits_{i\in I}E_i$ 可测，并且 $mE=\lim\limits_{i\to\infty}mE_i$。

2　基于测度理论的变精度粗糙集

1993年 Ziarko[4] 引入了变精度粗糙集，变精度粗糙集放宽了分类条件，允许一定程度的错误分类率存在。类似于传统变精度的粗糙集的上下近似空间的构造，这里给出基于测度的无限集合上的变精度粗糙集，定义如下：

定义3　设 U 是一集合，R 是其上一等价关系，$[x]_R$ 为 x 的等价类，$1\geqslant\beta>0.5$。定义 U 的子集 X 的上下近似空间及边界为

a）$\underline{R}_\beta(X)=\cup\{[x]|\dfrac{m([x]\cap X)}{m[x]}\geqslant\beta,x\in X\}=$

$\{x|\dfrac{m([x]\cap X)}{m[x]}\geqslant\beta,x\in X\}$；

b）$\overline{R}_\beta(X)=\cup\{[x]|\dfrac{m([x]\cap X)}{m[x]}>1-\beta,x\in X\}=$

$\{x|\dfrac{m([x]\cap X)}{m[x]}>1-\beta,x\in X\}$；

c）$\mathrm{bnr}_\beta(X)=\overline{R}_\beta(X)-\underline{R}_\beta(X)$。

对给定的 X，随着 β 的变化，其上下近似空间也随之变化。因此下面对相关性质进行讨论。

性质3　对于任意 β 和 X，上述定义的两个算子是对偶算子，即

$$\overline{R}_\beta(X)=\sim\underline{R}_\beta(\sim X)$$

证明　对于任意 β 和 X，有

$$\frac{m([x]\cap(X\cup\sim X))}{m[x]}=\frac{m([x]\cap X)}{m[x]}+\frac{m([x]\cap\sim X)}{m[x]}=1$$

如果 $x\in\underline{R}_\beta(X)$，则

$$\frac{m([x]\cap X)}{m[x]}\geqslant\beta\Rightarrow\frac{m([x]\cap\sim X)}{m[x]}\leqslant 1-\beta\Rightarrow$$
$$x\notin\overline{R}_\beta(\sim X)\text{，即 }x\in\sim\overline{R}_\beta(X)\text{。}$$

同理，如果$\dfrac{m([x]\cap X)}{m[x]}>1-\beta$，则

$$\frac{m([x]\cap X)}{m[x]}>1-\beta\text{，即}\frac{m([x]\cap\sim X)}{m[x]}<\beta$$

故结论成立。

说明：当 U 测度为0时，则改为势（基数），为方便起见，统一采用记号 mX。将上下近似空间看做是 β 和 X 的二元函数，本文可以讨论其相关性质。

2.1　近似空间的单调性

由于$[x]\subseteq X\Rightarrow\dfrac{m([x]\cap X)}{m[x]}=1\geqslant\beta$，所以有

性质4　如果集合 $X\subseteq U$，$0.5<\beta\leqslant 1$，则 $\underline{R}(X)\subseteq\underline{R}_\beta(X)\subseteq\overline{R}_\beta(X)\subseteq\overline{R}(X)$，$\underline{R}_\beta(X)\subseteq X\subseteq\overline{R}_\beta(X)$不成立。

例如，$U=\{a,b,c,d\}$，$E_1=\{a,b,c\}$，$E_2=\{d\}$，$X=\{a,b\}$，$\beta=0.6$，$Y=\{a,d\}$，$\underline{R}_\beta(X)=E_1\not\subset X$，$Y\not\subset\overline{R}_\beta(Y)=E_2$。

推论3　如果集合 $X\subseteq U$，$0.5<\beta\leqslant 1$，则

a）$\mathrm{bnr}_R(X)\supseteq\mathrm{bnr}_\beta(X)$；

b）$\mathrm{neg}_R(X)\supseteq\mathrm{neg}_\beta(X)$；

c）$\underline{R}_\beta(\varnothing)=\overline{R}_\beta(\varnothing)=\varnothing$；

d）$\underline{R}_\beta(U)=\overline{R}_\beta(U)=U$。

本节以下均容易由测度的性质1可得，证明略。

定理4

$$\overline{R}_\beta(X\cup Y)\supseteq\overline{R}_\beta(X)\cup\overline{R}_\beta(Y)$$
$$\underline{R}_\beta(X\cup Y)\supseteq\underline{R}_\beta(X)\cup\underline{R}_\beta(Y)$$

推论4　如果 $X\subseteq Y$，则

$$\overline{R}_\beta(X)\subseteq\overline{R}_\beta(Y)$$
$$\underline{R}_\beta(X)\subseteq\underline{R}_\beta(Y)$$

这表明上下近似空间关于变量 X 是单调上升的。

定义4　如果集合 $X,Y\subseteq U$，$0.5<\beta\leqslant 1$，则

$$X\overset{\beta}{\subseteq}Y\Leftrightarrow\frac{m(X\cap Y)}{mX}\geqslant\beta$$

$$X\subseteq Y\Leftrightarrow\frac{m(X\cap Y)}{mX}=1$$

性质 5　如果集合 $X,A,B\subseteq U$, $0.5<\beta\leqslant 1$,则

a) 如果 $A\cap B=\varnothing$, $A\overset{\beta}{\supseteq}X$,则 $B\overset{\beta}{\supseteq}X$ 不成立;

b) 如果 $\beta_1\leqslant\beta_2$,则 $A\overset{\beta_2}{\supseteq}X$,则 $A\overset{\beta_1}{\supseteq}X$。

推论 5　$\underline{R}_\beta(A)\cap\underline{R}_\beta(\sim A)=\varnothing$

定理 5

$$\overline{R}_{\beta_2}(X)\supseteq\overline{R}_{\beta_1}(X)$$
$$\underline{R}_{\beta_2}(X)\subseteq\underline{R}_{\beta_1}(X)$$

定理 5 说明上近似关于 β 单调上升,而下近似关于 β 单调下降。

推论 6　$\mathrm{bnr}_{\beta_1}(X)\supseteq\mathrm{bnr}_{\beta_2}(X)$

2.2　近似空间的连续性

视上下近似空间 $R_\beta(X)$ 为二元函数,下面研究其连续性等性质,为此先给出极限等概念。

定义 5　对 $X\subseteq U$,

$$\lim_{\beta\to\beta_0}\overline{R}_\beta(X)\overset{\Delta}{=}\bigcap_{\beta>\beta_0}\overline{R}_\beta(X)$$
$$\lim_{\beta\to\beta_0}\underline{R}_\beta(X)\overset{\Delta}{=}\bigcap_{\beta>\beta_0}\underline{R}_\beta(X)$$
$$\lim_{\beta\to\beta_0}\mathrm{bnr}_\beta(X)=\lim_{\beta\to\beta_0}\overline{R}_\beta(X)-\lim_{\beta\to\beta_0}\underline{R}_\beta(X)$$
$$\overline{R}_{0.5}(X)\overset{\Delta}{=}\cup\{x\mid\frac{m([x]\cap X)}{m[x]}>1-\beta=0.5\}$$
$$\underline{R}_{0.5}(X)\overset{\Delta}{=}\cup\{x\mid\frac{m([x]\cap X)}{m[x]}\geqslant\beta=0.5\}$$
$$\mathrm{bnr}_{0.5}(X)=\overline{R}_{0.5}(X)-\underline{R}_{0.5}(X)$$

定理 6　$\lim\limits_{\beta\to 0.5}\overline{R}_\beta(X)\subseteq\overline{R}_{0.5}(X)$。

证明　$\beta'\leqslant\beta\Rightarrow\overline{R}_\beta(X)\supseteq\overline{R}_{\beta'}(X)\Rightarrow$

$$\bigcap_{\beta>0.5}\overline{R}_\beta(X)=\min_{\beta>0.5}\overline{R}_\beta(X)\supseteq\overline{R}_{0.5}(X)。$$

证毕。

注:等号不一定成立。

例 1　$U=\{a,b\}\cup\{c\}$, $X=\{a\}$,则

$$\underline{R}_\beta(X)=\varnothing,\ \sup_{\beta>0.5}\underline{R}_\beta(X)=\varnothing,\ \underline{R}_{0.5}(X)=\{a,b\}$$
$$U=\{a,b\}\cup\{c\},\ X=\{a\},则$$
$$\overline{R}_\beta(X)\underset{\beta>0.5}{=}\{a,b\},\ \inf_{\beta>0.5}\overline{R}_\beta(X)=\{a,b\},\ \overline{R}_{0.5}(X)=\varnothing$$
$$U=\{a,b,c\}\cup\{d\},\ X=\{a\},则$$
$$\overline{R}_\beta(X)\underset{\beta=2/3}{=}\{a,b\},\ \overline{R}_\beta(X)\underset{\beta>2/3}{=}\varnothing,\ \lim_{\beta\to 2/3}\overline{R}_\beta(X)=\varnothing$$

这说明上下近似算子关于 β 是不连续的,但可以证明是分段连续函数。

2.3　近似空间的对合性

性质 6　如果集合 $X\subseteq U$, $0.5<\beta\leqslant 1$,则 $\underline{R}_\beta(X)\overset{\beta}{\subseteq}X$, $\overline{R}_\beta(X)\overset{1-\beta}{\subseteq}X$。

证明　$\underline{R}_\beta(X)\overset{\beta}{\subseteq}X\Leftrightarrow\dfrac{m(\underline{R}_\beta(X)\cap X)}{m\,\underline{R}_\beta(X)}\geqslant\beta$

$$\frac{m((E_1\cup E_2)\cap X)}{m(E_1\cup E_2)}=\frac{m(E_1\cap X)+m(E_2\cap X)}{m(E_1\cup E_2)}=$$
$$\frac{m(E_1\cap X)}{mE_1}\times\frac{mE_1}{m(E_1\cup E_2)}+\frac{m(E_2\cap X)}{mE_2}\times\frac{mE_2}{m(E_1\cup E_2)}\geqslant\beta$$

另一部分类似可证。

性质 7　对于任意 $X=\cup\{[x]\mid x\in U\}$, $0.5<\beta\leqslant 1$,则 $\overline{R}_\beta(X)=\underline{R}_\beta(X)=X$。

证明　首先证明对任意 $x\subseteq U$,都有 $\overline{R}_\beta([x])=\underline{R}_\beta([x])=[x]$。

因为对于任意 $t\subseteq U$, $[x]\cap[t]=[x]$ 或 $\varnothing$,

所以 $\dfrac{m([x]\cap[t])}{m[t]}=1$ 或 0。

再证明对任意 $x_i\subseteq U$,都有

$$\overline{R}_\beta(\cup[x_i])=\underline{R}_\beta(\cup[x_i])=\cup[x_i]$$

对任意 $t\in U$,存在 i, $(\cup[x_i])\cap[t]=[x_j]$ 或 $\varnothing$,

$$\frac{m(([x]\cap[y])\cap[t])}{m[t]}=1\geqslant\beta\Rightarrow[t]\subseteq R_\beta([x]\cup[y])$$

否则不包含 $[t]$,命题得证。

推论 7　对于 U 的任意子集 X, $0.5<\beta\leqslant 1$,则 $\underline{R}_\beta(\underline{R}_\beta(X))=\underline{R}_\beta(X)$, $\overline{R}_\beta(\overline{R}_\beta(X))=\overline{R}_\beta(X)$。

推论 7 说明上下近似算子是对合的,即 $R^2=R$。

例 2　$U=[-1,1]\times[-1,1]\overset{\Delta}{=}K_1\cup K_2\cup K_3\cup K_4$, $R=\{(x,y)\mid x,y\in K_i\}$, K_i 为图 1 所示的第 i 个方格。

对 $\beta>0.5$, $X=[1/4,3/4]\times[-1,1]$, $\overline{R}_\beta(X)=K_2\cup K_4$, $\overline{R}_{0.5}(X)=\varnothing$。

有界无限集如图 1 所示。

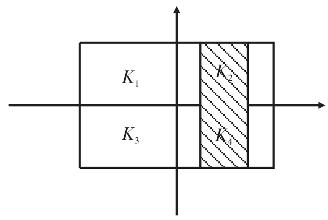

图1　有界无限集

2.4　近似空间的可辨别性

尽管引入 β 后, X 与其上下近似空间没有包含关系,但当其上下近似空间相等时,还是有许多的好的性质。为此给出可辨别的。

定义 6　对 X,β,如果 $\mathrm{bnr}_\beta X=\varnothing$,称集合 X 为 β 可辨别的。

下面两定理可由定理 5 推出。

定理 7　如果集合 X 为 $0.5<\beta\leqslant 1$ 是可辨别的,那么 X 对任何 $\beta'<\beta$ 也是可辨别的。

定理 8　如果集合 X 为 $0.5<\beta\leqslant 1$ 是不可辨别的,那么 X 对任何 $\beta'>\beta$ 也是不可辨别的。

推论 8　$\lim\limits_{\beta\to 0.5}\mathrm{bnr}_\beta X\neq\varnothing$,那么对任何 $0.5<\beta\leqslant 1$ 是不可辨别的。

可见讨论不可辨的 β 的极值是十分有意义的。

定义 7　$\mathrm{Ndis}(R,X)=\{\beta\mid 0.5<\beta\leqslant 1,\mathrm{bnr}_\beta X\neq\varnothing\}$。

定理 9　$\inf(\mathrm{Ndis}(R,X))=\min\{m_1,m_2\}\overset{\Delta}{=}\xi$,并且 $\xi\in\mathrm{Ndis}(R,X)$。

$$m_1=\inf\{\frac{m([x]\cap X)}{m[x]}>0.5\}$$
$$m_2=1-\sup\{\frac{m([x]\cap X)}{m[x]}<0.5\}$$

β 的可辨别区间如图 2 所示。

图 2　β 的可辨别区间 $(\xi,1]$

证明　令　$Y_1=\{x\mid 0\leqslant\dfrac{m([x]\cap X)}{m[x]}\leqslant 0.5\}$

$$Y_2=\{x\mid 0.5\leqslant\frac{m([x_1]\cap X)}{m[x_1]}\leqslant 1\}$$

如果存在 x_3 使得$\frac{m([x_3]\cap X)}{m[x_3]}=0.5$,那么对任何$\beta>0.5$,$\overline{R}_\beta(X)\supseteq[x_3]$,而 $x_3\notin\underline{R}_\beta(X)$,故 $\mathrm{bnr}_\beta X\neq\varnothing$。

设不存在 x_3,并且 $m_1=\frac{m([x_1]\cap X)}{m[x_1]}>\frac{m([x]\cap X)}{m[x]}$,$x_1$,$x\in Y_1$;

$m_2=1-\frac{m([x_2]\cap X)}{m[x_2]}\geqslant 1-\frac{m([x]\cap X)}{m[x]}$,$x_2,x\in Y_2$,$m_1=m_2=\beta>0.5$,$x_2\notin\underline{R}_\beta(X)=\overline{R}_\beta(X)$。

当 $m_1<m_2$ 时,对任何$\beta>m_1$,$[x_2]\not\subset\overline{R}_\beta(X)=\underline{R}_\beta(X)$;

当 $m_1>m_2$ 时,对任何$\beta>m_2$,$[x_2]\subseteq\overline{R}_\beta(X)=\underline{R}_\beta(X)$。

所以 $\sup(\mathrm{Ndis}(R,X))=\min\{m_1,m_2\}$。

如果不存在这样的 x_1、x_2,类似可证。

3 结束语

本文提出了一个基于有界无限集合的变精度粗糙集模型——基于 Lebesgue 测度的变精度粗糙集模型,进而对相关性质进行了充分的研究。后续有待解决的问题有很多,主要归纳如下:a)对于无界集合的上下近似空间如何定义?如自然数集合及其子集偶数集合;b)拓展模型的应用研究,如快速分类、属性约简等;c)变精度粗糙集模型的公理化以及基于 Lebesgue 测度的变精度粗糙集模型的公理化是今后要研究的重点。

参考文献:

[1] 王丽娜. 基于粗糙集的数据挖掘改进的属性约简算法研究[D]. 成都:电子科技大学,2015.

[2] 张人上,曲开社. 一种基于新的特征选择的海量网络文本挖掘算法研究[J]. 计算机应用研究,2014,31(9):2632-2634.

[3] Pawlak Z. Rough sets[J]. International Journal of Information and Compute Sciences,1982,11(5):341-356.

[4] Ziarko W. Variable precision rough set model[J]. Journal of Computer and System Sciences,1993,46(1):39-59.

[5] 刘妍琼,钟波. 变精度粗糙集模型中 β 参数范围的确定[J]. 湖南理工学院学报:自然科学版,2008,21(1):11-13.

[6] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社,2006:123-132.

[7] Liang Jiye,Wang Feng,Dang Chuangyin,*et al*. A group incremental approach to feature selection applying rough set technique[J]. IEEE Trans on Knowledge & Data Engineering,2014,26(2):294-308.

[8] Liang Jiye,Mi Junrong,Wei Wei,*et al*. An accelerator for attribute reduction based on perspective of objects and attributes[J]. Knowledge-Based Systems,2013,44(1):90-100.

[9] 苏永华,刘科伟,张进华. 基于粗糙集重心理论的公路隧道塌方风险分析[J]. 湖南大学学报:自然科学版,2013,40(1):21-26.

[10] Liu Guilong. Axiomatic systems for rough sets and fuzzy rough sets[J]. International Journal of Approximate Reasoning,2008,48(3):857-867.

[11] Liu Guilong. Using one axiom to characterize rough set and fuzzy rough set approximations[J]. Information Sciences,2013,223(2):285-296.

[12] 刘耀峰. 基于 Sugeno 测度的粗糙集模型[D]. 保定:河北大学,2010.

[13] 薛占熬,刘杰,朱泰隆,等. 基于覆盖的 Sugeno 测度粗糙集模型及其三支决策[J]. 计算机科学,2016,43(3):285-290.

[14] Yang Yanyan,Chen Degang,Dong Ze. Novel algorithms of attribute reduction with variable precision rough set model[J]. Neurocomputing,2014,139(9):336-344.

[15] Chen Degang,Yang Yanyan,Dong Ze. An incremental algorithm for attribute reduction with variable precision rough sets[J]. Applied Soft Computing,2016,45(8):129-149.

[16] Halmos P R. Measure theory[M]. [S. l.]:World Publishing Corporation,2007:100-152.

(上接第 1371 页)

参考文献:

[1] Resnick P,Iacovou N,Suchak M,*et al*. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proc of ACM Conference on Computer Supported Cooperative Work. New York: ACM Press,1994:175-186.

[2] Sarwar B,Karypis G,Konstan J,*et al*. Item-based collaborative filtering recommendation algorithms[C]//Proc of International Conference on World Wide Web. New York:ACM Press,2001:285-295.

[3] Linden G,Smith B,York J. Amazon. com recommendations: item-to-item collaborative filtering[J]. IEEE Internet Computing,2003,7(1):76-80.

[4] Hu Guangneng,Dai Xinyu,Song Yunya,*et al*. A synthetic approach for recommendation: combining ratings, social relations, and reviews[C]//Proc of International Conference on Artificial Intelligence. Palo Alto,CA:AAAI Press,2015.

[5] Goldberg K,Roeder T,Gupta D,*et al*. Eigentaste: a constant time collaborative filtering algorithm[J]. Information Retrieval Journal,2001,4(2):133-151.

[6] Titov I,McDonald R. A joint model of text and aspect ratings for sentiment summarization[C]//Proc of ACL-08 HLT. 2008:308-316.

[7] McAuley J,Leskovec J. Hidden factors and hidden topics: understanding rating dimensions with review text[C]//Proc of the 7th ACM Conference on Recommender Systems. New York:ACM Press,2013:165-172.

[8] Blei D M,Ng A Y,Jordan M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research,2003,3:993-1022.

[9] Wei Xing,Croft W B. LDA-based document models for Ad hoc retrieval[C]//Proc of the 29th Annual International SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press,2006:178-185.

[10] Alsumait L,Barbará D,Domeniconi C. On-line LDA: adaptive topic models for mining text streams with applications to topic detection and tracking[C]//Proc of the 8th IEEE International Conference on Data Mining. Washington DC:IEEE Computer Society,2008:3-12.

[11] Burns N,Bi Yaxin,Wang Hui,*et al*. A twofold-LDA model for customer review analysis[C]//Proc of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Piscataway:IEEE Press,2011:253-256.

[12] Lee K D,Han K,Myaeng S H. Capturing word choice patterns with LDA for fake review detection in sentiment analysis[C]//Proc of the 6th International Conference on Web Intelligence,Mining and Semantics. 2016:1-7.

[13] 张晓. 基于 LDA 主题模型的文本聚类研究[EB/OL]. [2012-02-28]. http://www. paper. edu. cn/releasepaper/content/201202-1066.

[14] 项亮. 推荐系统实践[M]. 北京:人民邮电出版社,2012.

[15] Jo Y,Oh A H. Aspect and sentiment unification model for online review analysis[C]//Proc of the 4th ACM International Conference on Web Search and Data Mining. New York:ACM Press,2011:815-824.

[16] Xu Jingnan,Zheng Xiaolin,Ding Weifeng. Personalized recommendation based on reviews and ratings alleviating the sparsity problem of collaborative filtering[C]//Proc of the 9th IEEE International Conference on e-Business Engineering. Washington DC: IEEE Computer Society,2012:9-16.