流量异常检测中的直觉模糊推理方法

范晓诗* 雷英杰 王亚男 郭新鹏
(空军工程大学防空反导学院 西安 710051)

摘 要：针对网络流量特征属性不确定性和模糊性的特点，将直觉模糊推理理论引入异常检测领域，该文提出一种基于包含度的直觉模糊推理异常检测方法。首先设计异常检测中特征属性的隶属度与非隶属度函数，其次，给出基于包含度的强相似度计算方法并生成推理规则库，再次给出多维多重式直觉模糊推理规则，最后建立异常检测中的直觉模糊推理方法。通过对异常检测标准数据集KDD99的实验，验证该方法的有效性，与常见经典异常检测方法对比，该方法具有更良好的检测效果。

关键词：网络；信息安全；直觉模糊集；异常检测；直觉模糊推理；特征属性；包含度
中图分类号：TP393.08 **文献标识码**：A **文章编号**：1009-5896(2015)09-2218-07
DOI：10.11999/JEIT150023

Intuitionistic Fuzzy Reasoning Method in Traffic Anomaly Detection

Fan Xiao-shi Lei Ying-jie Wang Ya-nan Guo Xin-peng
(*Air and Missile Defense College, Air Force Engineering University, Xi'an* 710051, *China*)

Abstract: Aiming at the characteristics of uncertainty and fuzziness of the network traffic attribute, an Intuitionistic Fuzzy Reasoning Theory (IFRT) is introduced to the anomaly detection field. A method of IFRT detection based on the inclusion degree is proposed. Firstly, the membership and non-membership functions of attributes in anomaly detection are designed. Secondly, the intensity similarity measure method based on the inclusion degree is presented and the rules library is generated. And then, the FMP rules of the IFRT are presented. Finally, an anomaly detection based on the IFRT is constructed. The validity is checked by experiment on the standard detection dataset KDD99, compared with other traditional theory, the IFRT anomaly detection method performs better than others.

Key words: Network; Information security; Intuitionistic fuzzy sets; Anomaly detection; Intuitionistic Fuzzy Reasoning (IFR); Feature attribute; Inclusion degree

1 引言

直觉模糊集合(Intuitionistic Fuzzy Set, IFS)作为模糊集合的泛化理论，因其引入犹豫度参数对客观事物不确定性表述的优势，被学者们广泛研究。直觉模糊推理(Intuitionistic Fuzzy Reasoning, IFR)是模糊理论应用最为广泛的工具，文献[1]将直觉模糊推理与认知图结合，应用于决策支持，文献[2]利用模糊神经网络规则推理对非线性系统进行控制，文献[3]将直觉模糊推理理论应用于目标识别领域，得到了很好的识别效果，文献[4]首次提出了Ⅱ型模糊集理论在模式识别中的应用问题，文献[5]和文献[6]分别将模糊推理运用于故障诊断与图像处理方面，文献[7,8]研究了威胁评估中的直觉模糊推理方法。由此可见，直觉模糊推理理论是一种十分有效的智能信息处理方法。

相似度和包含度是直觉模糊集合间关系的度量，能够有效处理直觉模糊计算问题，是精确问题和模糊问题相互转换的桥梁，许多文献也做了相关研究，文献[9]和文献[10]分别研究了直觉模糊相似度的建模和模式识别应用问题，文献[11]提出了基于包含度的直觉模糊推理方法，并证明给出的公式满足相关公理化定义。这些研究都为直觉模糊推理的应用和推广提供了理论基础，并促进相关领域的创新和发展。

异常检测是网络安全面对的重要问题，网络异常检测通常分为基于统计的检测和基于特征的异常检测，前者通用性能好，但准确度不够理想，后者的优点是准确度高，缺点是特征匹配方法往往效率较低，维护特征数据库的系统开销较大。流量异常检测已有许多基于不同智能计算理论的研究，如统计学原理、免疫计算[12]、支持向量机[13]等，以及一

2015-01-06收到，2015-03-18改回，2015-06-18网络优先出版
*通信作者：范晓诗 fan_xs@126.com

些结合模糊理论的方法，如遗传模糊系统[14]、模糊粒子群算法[15]。传统异常检测方法受到计算复杂度和数据规模的限制，只能选取若干特征属性作为检测指标，对网络流量全局特征的刻画能力有限，尤其是针对连续特征属性，传统精确数据的处理方法，为了提高检测准确率，通常采用模式规模扩充，或者对特征属性进行更细致的划分，这些方法都是以牺牲系统资源为代价，不利于计算方法的进一步优化。直觉模糊理论可以很好描述系统的不确定性和模糊性，通过直觉模糊化将精确数据映射到直觉模糊集中，降低规则库规模，是解决网络流量分类、异常检测问题的新思路。

本文将直觉模糊推理理论应用于网络流量异常检测，充分考虑流量特征描述不确定性和模糊性，并通过实验证明其有效性。

2　异常检测的直觉模糊推理方法

直觉模糊推理的一般过程包括系统输入变量直觉模糊化，推理规则的建立，推理规则合成，输出结果等步骤。运用直觉模糊推理方法进行异常检测时，首先对网络流量特征属性直觉模糊化，然后建立相应的推理规则库，根据推理规则合成，将检测数据输入系统，最后得到输出结果。

2.1 数据直觉模糊化

根据网络流量特征属性，确定直觉模糊系统的隶属度和非隶属度函数。该方法实际上是一个集合映射的过程，将每一个特征属性定义为一个直觉模糊变量，根据特征属性类型确定函数。异常检测KDD99 数据集[16]中属性包括离散型和连续型，由于离散型数值各自互斥，没有明显的相关性和相似性，因而采用严格三角隶属度。本文涉及的主要特征属性参数说明如表 1 所示。

假设某一特征属性 A 有 N 个离散属性值，定义第 i 个属性值对应隶属度函数为 $(i-1)/n$，令犹豫度 $\pi_A(x)=0$，则非隶属度函数为 $\gamma_A(x)=1-\mu_A(x)$。例如，KDD99 数据集中 protocol_type 特征属性包括 3 个离散变量 TCP, UDP 和 ICMP，根据上述定义，可以计算得到其特征属性函数分别为<0, 1>, <0.333, 0.667>和<0.667, 0.333>。

为了合理描述网络流量分布特性，对于连续型变量，采用高斯型隶属度函数，即

$$\left.\begin{aligned}\mu_A(x)&=\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\\ \gamma_A(x)&=\delta_A(x)-\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)\\ \delta_A(x)&=1-\pi_A(x)\end{aligned}\right\}\tag{1}$$

根据式(1)，首先对特征属性的论域进行划分，得到特征属性子集，其次确定参数 σ 和 c，得到各个特征变量函数，计算特征属性值对应各个直觉模糊子集的相关输入函数参数，具体步骤为：

步骤 1　根据属性特征划分特征子集 x 变化范围,记做 $I_1, I_2, \cdots, I_n$；

步骤 2　设特征属性子集 A_i 的定义域为[A,B]，确定对应的值域为[C, D]，记映射函数为 $f(x, a, b, c)$；

步骤 3　计算映射参数，$D=\sigma+c$，其中 c 为中心，σ 为宽度，$a=A+(B-A)/2$，将定义域和值域带入映射函数 $f=c+(x-a)/b$，计算得到 b；

步骤 4　多次代入数值检验 x 输出分布是否均匀，调整参数和区间划分；

步骤 5　根据式(1)计算隶属度与非隶属度函数。

例如对数据包 byte 这一特征属性进行直觉模糊化，由于网络中字节数这一指标是非均匀分布，存在大量空数据包，而小数据包变化单位为字节级，而大数据包变化尺度为百字节级甚至千字节级，平均划分论域不能很好描述数据特征。因此，本文对全局论域变尺度划分，得到数据包由大到小分别为 $B_1=[10240,\infty)$, $B_2=[4096, 10240)$, $B_3=[1024, 4096)$, $B_4=[256, 1024)$, $B_5=[1, 256)$, $B_6=0$。相应

表 1　KDD99 数据集主要特征属性参数说明

特征属性	参数说明	类型	取值
duration	连接持续时间(s)	连续	[0, 58329]
protocl_type	协议类型，包括 TCP, UDP, ICMP	离散	3
service	目标主机的网络服务类型	离散	70
flag	连接正常或错误的状态	离散	11
src_byte/dst_byte	从源(目的)主机到目标(源)主机的字节数	连续	[0, 1379963888]
urgent	加急包的个数	连续	[0, 14]
hot	访问系统敏感文件和目录的次数	连续	[0, 101]
count	与当前连接相同的目标主机连接数	连续	[0,511]

的直觉模糊子集可划分为 I_1=[0, 0.1), I_2=[0.1, 0.3), I_3=[0.3, 0.5), I_4=[0.5, 0.7), I_5=[0.7, 0.9), I_6=1，隶属度函数如图 1 所示，特征属性 Byte 函数参数值如表 2 所示。根据以上步骤，可以得出特征属性 byte 输入函数为

$$b=\begin{cases}0, & x\geq 20480\\ (x-10240)/51200, & 10240<x\leq 20480\\ (x-7168)/30720+0.2, & 4096<x\leq 10240\\ (x-2560)/15360+0.4, & 1024<x\leq 4096\\ (x-640)/3840+0.6, & 256<x\leq 1024\\ (x-128.5)/1275+0.8, & 0<x\leq 256\\ 1, & x=0\end{cases} \tag{2}$$

同理可以分别得到特征属性 duration, service, flag, urgent 的函数，将特征属性 service 和 flag 直接线性映射在[0,1]区间，其余连续函数如式(3)和式(4)所示。

$$d=\begin{cases}(x-30000)/20000, & 10000<x\leq 50000\\ (x-5500)/4500+0.25, & 1000<x\leq 10000\\ (x-550)/4500+0.5, & 100<x\leq 1000\\ (x-50)/500+0.75, & 0<x\leq 100\\ 1, & x=0\end{cases} \tag{3}$$

$$u=\begin{cases}(x-12.5)/4.5005+0.333, & 10<x\leq 14\\ (x-7.5)/7.5+0.667, & 5<x\leq 10\\ -x/15.015+1, & 0<x\leq 5\end{cases} \tag{4}$$

最后，定义输出论域U'。将流量检测结果分为

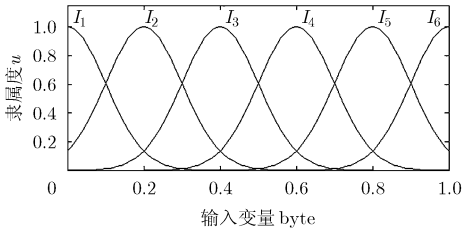

图 1 变量 byte 隶属度函数

5 类，normal, Probe, DoS, U2R 和 R2L，分别对应 5 个直觉模糊子集[0,0.2], [0.2,0.4], [0.4,0.6], [0.6,0.8]和[0.8,1.0]。

对于 U2R 和 R2L 之类的攻击，其数据包与正常连接没有明显区别，所以选择若干基于连接内容和连接时间的特征属性如 hot, count。经过特征属性直觉模糊化后，得到异常检测参数变量 D(duration), S(service), F(flag), B(byte), U(urgent), H(hot), C(count)，则异常检测推理系统推理规则数 $N=N_d\times N_s\times N_f\times N_b\times N_u\times N_h\times N_c=4\times 70\times 11\times 6\times 3\times 3\times 3=498960$。这样的推理规则数量过于庞大，可以进行再次直觉模糊化，例如 service 特征属性包含 70 个变量，而 csnet_net, ctf, discard, daytime 等均对应 neptune 攻击，将这些可以推理出相同分类结果的服务类型进行聚合，最后得到 12 个新的直觉模糊子集，约简后推理规则 $N'=4\times 12\times 11\times 6\times 3\times 3\times 3=85536$ 条，可见利用属性约简的方法可以大大降低推理规则数量，另一方面，这里的 N' 是理论规则库，并非所有规则都需要生成，通过前期对样本数据的训练，得到理论规则库的一个子集，可进一步缩减规则库的规模，达到提高效率的目的。

2.2 推理规则建立

包含度和强相似度是刻画直觉模糊集合关系的度量，能够有效反应直觉模糊集数据间关系，是直觉模糊推理的基础，本文拟在相关理论的基础上构建基于直觉模糊推理的异常检测方法。

下面给出直觉模糊包含度的相关定义[11]。

定义 1 设直觉模糊集上 IFS×IFS→[0,1]的映射 θ 满足条件：(1) $A\subseteq B\Rightarrow\theta(A,B)=1$, (2) $\theta(A,\phi)=0$, (3) $A\subseteq B\subseteq C\Rightarrow\theta(C,A)\leq\min\{\theta(B,A),\theta(C,B)\}$。则称 $\theta(A,B)$ 为 A 在 B 中的包含度，θ 为 IFS 上的包含度函数。

定义 2 设 R 是直觉模糊蕴含算子 $R(a,b)$关于 a 的非增函数，关于 b 的非减函数，则

$$\theta(A,B)=\frac{1}{n}\sum_{i=1}^{n}\left\{\lambda R\left[\mu_A(x_i),\mu_B(x_i)\right]+(1-\lambda)\cdot R\left[1-\gamma_A(x_i),1-\gamma_B(x_i)\right]\right\},\ \lambda\in[0,1] \tag{5}$$

表 2 特征属性 byte 函数参数值

特征属性	变量	[A,B]	[C, D]	a	b	c	σ
byte	b	[10240, ∞)	[0, 0.1)	/	51200	0	0.1
		[4096, 10240)	[0.1, 0.3)	7168	30720	0.2	
		[1024, 4096)	[0.3, 0.5)	2560	15360	0.4	
		[256, 1024)	[0.5, 0.7)	640	3840	0.6	
		[1, 256)	[0.7, 0.9)	128.5	1275	0.8	
		0	1	/	/	1	

为 IFS 包含度函数，λ是蕴涵算子权重系数，一般取值较小，这里取$\lambda=0.1$。包含度函数选择不唯一，通常选择满足定义并易于计算的函数。

定义 3　设$\theta(A,B)$为 IFS 上的包含度函数，则

$$\delta(A,B)=\theta(A\cup B,A\cap B) \tag{6}$$

是 A 和 B 的强相似度量，相关公理化定义可参阅文献[10]。

根据以上定义和函数公式，表 3 给出多维多重式规则的基于包含度的直觉模糊推理形式。

表 3 中，$i(D)=1,2,\cdots,N_d$，$i(S)=1,2,\cdots,N_s$，$i(F)=1,\ 2,\cdots,N_f$，$i(B)=1,2,\cdots,N_b$，$i(U)=1,2,\cdots,N_u$，$i(H)=1,2,\cdots,N_h$，$i(C)=1,2,\cdots,N_c$，CF_i为直觉模糊推理可信度因子，d, s, f, b, u, h, c 是输入特征属性变量，z是输出变量，D,S,F,B,U,H,C是语言前件，即$<d,\mu_{D_i},\gamma_{D_i}>,\ d\in D;\ <s,\mu_{S_i},\gamma_{S_i}>,\ s\in S;\ <f,\mu_{F_i},\gamma_{F_i}>,f\in F\ ;<b,\mu_{B_i},\gamma_{B_i}>,b\in B\ ;<u,\mu_{U_i},\gamma_{U_i}>,u\in U;<h,\mu_{H_i},\gamma_{H_i}>,h\in H;<c,\mu_{C_i},\gamma_{C_i}>,c\in C$。$U'$为推理后件，即输出论域的直觉模糊子集，$<z,\mu_{\mathrm{U}'_j},\gamma_{\mathrm{U}'_j}>,z\in U'$。

表 3 基于包含度的直觉模糊推理形式

规则：IF d is D_i and s is S_i and f is F_i and b is B_i and u is U_i and h is H_i and c is C_i, Then z is U'_j (CF_i).
输入：d^* is D_i and s^* is S_i and f^* is F_i and b^* is B_i and u^* is U_i and h^* is H_i and c^* is C_i.
输出：z^* is U'_j (CF_i).

2.3 推理规则合成

根据以上定义，可以构建出直觉模糊推理方法，具体步骤为：

步骤 1　根据式(5)选取λ和直觉模糊蕴含算子R，若 X 表示规则特征属性变量，X^*表示检测数据特征属性，求得包含度$\theta(X,X^*)$，进而根据式(6)求得强相似度$\delta(X,X^*)$。

步骤 2　利用 Mamdani 算子 $R_{\mathrm{c}}(\mathrm{A}\rightarrow\mathrm{B})$推导输出结果 z。

$$R_u=\bigcup_{i_d,i_s,i_f,i_b,i_u,i_c;j=1}^{N_d,N_s,N_f,N_b,N_u,N_c;N_{u'}}\cdots\bigcup R\left(D_i,S_i,F_i,B_i,U_i,H_i,C_i;C_j\right)$$

$$=\bigcup_{i_d,i_s,i_f,i_b,i_u,i_c;j=1}^{N_d,N_s,N_f,N_b,N_u,N_c;N_{u'}}\cdots\bigcup R\left(D_i\cap S_i\cap F_i\cap B_i\cap U_i\cap H_i\cap C_i\cap C_j\right) \tag{7}$$

$$(\mathrm{CF}_i,\ i=1,2,\cdots,N')$$

$$\mu_R=\bigvee_{i_d,i_s,i_f,i_b,i_u,i_c,i=1;j=1}^{N_d,N_s,N_f,N_b,N_u,N_c,N;N_{u'}}\cdots\bigvee\left(\mu_{D_i}\wedge\mu_{S_i}\wedge\mu_{F_i}\wedge\mu_{B_i}\wedge\mu_{U_i}\wedge\mu_{H_i}\wedge\mu_{C_i}\wedge\mu_{z_i}\wedge\mathrm{CF}_i\right) \tag{8}$$

$$\gamma_R=\bigwedge_{i_d,i_s,i_f,i_b,i_u,i_c,i=1;j=1}^{N_d,N_s,N_f,N_b,N_u,N_c,N;N_{u'}}\cdots\bigwedge\left(\gamma_{D_i}\vee\gamma_{S_i}\vee\gamma_{F_i}\vee\gamma_{B_i}\vee\gamma_{U_i}\vee\gamma_{H_i}\vee\gamma_{C_i}\vee\gamma_{z_i}\vee\mathrm{CF}_i\right) \tag{9}$$

则$Z=<\mu_R,\gamma_R>,z\in U'$。

步骤 3　当$\sum_{i=1}^{N}\delta_i\neq 0$时，根据推理规则$\delta_i(X,X^*)$的推理结果为

$$Z^*=<\frac{1}{N}\mu_R\sum_{i=1}^{N}\delta_i,\frac{1}{N}\gamma_R\sum_{i=1}^{N}\delta_i>\Big/z_i,\ z_i\in U' \tag{10}$$

式中 N 是特征属性维数。根据上述推理过程，将检测数据在规则库中进行匹配，选择最大强相似度输出直觉模糊集作为推导结果。

3　实验和分析

为验证本文提出的直觉模糊推理异常检测的方法，利用 KDD99 实验数据集的 10%的训练样本集生成一个直觉模糊推理规则库，接着对 corrected 测试数据集直觉模糊化，得到输入向量，最后经过推理系统得到输出结果，以分类准确率验证方法性能。

3.1 训练规则库

训练样本集的部分数据如表 4 所示，每条数据由 7 维特征属性和 1 个类别标签构成，根据变量直觉模糊化方法得到相应的推理规则库，部分规则如表 5 所示。规则推理求解如图 2 所示，从图中可以看出，前 4 条规则可以合并为 1 条，由此可见直觉模糊化具有化简规则库的作用。

3.2 推理步骤

实验 1　本文首先验证推理方法的有效性，以 corrected 测试数据集随机选取的 9 条数据为例，如表 6 前 8 列所示，首先对检测数据进行直觉模糊化，得到如下输入向量：

$\boldsymbol{I}_1=[1,\ 0.9130,\ 0.1818,\ 0.7820,\ 1,\ 0,\ 0.0020]$, $\boldsymbol{I}_2=$

表 4 训练样本数据集

duration	service	flag	byte	urgent	hot	count	label	duration	service	flag	byte	urgent	hot	count	label
0	http	SF	181	0	0	8	normal	0	ecr_i	SF	1032	0	0	316	DoS
0	http	SF	239	0	0	8	normal	0	ecr_i	SF	1032	0	0	511	DoS
0	http	SF	235	0	0	6	normal	0	ecr_i	SF	1032	0	0	509	DoS
0	http	SF	212	0	1	8	normal	⋮			⋮				⋮
⋮			⋮				⋮	0	private	REJ	0	0	0	1	Probe
0	smtp	SF	1551	0	0	1	normal	⋮			⋮				⋮
0	smtp	SF	1367	0	0	1	normal	0	ftp_data	SF	334	0	0	1	R2L
⋮			⋮				⋮	⋮			⋮				⋮
184	telnet	SF	1511	0	3	1	U2R	0	http	REJ	0	0	0	1	normal
305	telnet	SF	1735	0	3	1	U2R	0	domain_u	SF	30	0	0	1	normal
⋮			⋮				⋮	⋮			⋮				⋮

表 5 推理规则库

duration	service	flag	byte	urgent	hot	count	label	duration	service	flag	byte	urgent	hot	count	label
1.0000	0.0833	0.1818	0.8414	1	0	0.0160	0.2	1.0000	0.7470	0.1818	0.3005	1	0	0.6320	0.6
1.0000	0.0833	0.1818	0.8867	1	0	0.0160	0.2	1.0000	0.7470	0.1818	0.3005	1	0	0.9980	0.6
1.0000	0.0833	0.1818	0.8836	1	0	0.0120	0.2	1.0000	0.7470	0.1818	0.3005	1	0	0.9941	0.6
1.0000	0.0833	0.1818	0.8656	1	0.01	0.0160	0.2	⋮			⋮				⋮
⋮			⋮				⋮	1.0000	0.9130	0.5454	1.0000	1	0	0.0020	0.4
1.0000	0.3320	0.1818	0.3343	1	0	0.0020	0.2	⋮			⋮				⋮
1.0000	0.3320	0.1818	0.3223	1	0	0.0020	0.2	1.0000	0.4980	0.1818	0.9609	1	0	0.0020	1.0
⋮			⋮				⋮	⋮			⋮				⋮
0.4187	0.1660	0.1818	0.3343	1	0.03	0.0020	0.8	1.0000	0.0833	0.5454	0	1	0	0.0020	0.2
0.4456	0.1660	0.1818	0.3463	1	0.03	0.0020	0.8	1.0000	0.3320	0.1818	0.7234	1	0	0.0020	0.2
⋮			⋮				⋮	⋮			⋮				⋮

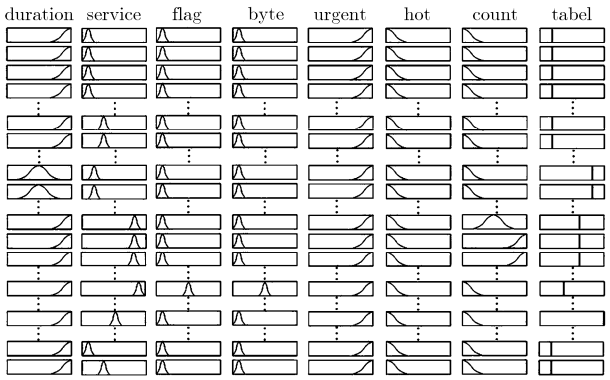

图 2 规则推理图

[1, 0.9130, 0.1818, 0.7820, 1, 0, 0.0040], $\boldsymbol{I}_3$= [1, 0.7470, 0.1818, 0.3005, 1, 0, 0.9785], $\boldsymbol{I}_4$= [1, 0.7470, 0.1818, 0.3005, 1, 0, 0.9941], $\boldsymbol{I}_5$= [1, 0.7470, 0.1818, 0.3005, 1, 0, 0.9980], $\boldsymbol{I}_6$= [1, 0.7470, 0.1818, 0.7141, 1, 0, 0.0020], $\boldsymbol{I}_7$= [0.7880, 0.1660, 0.1818, 0.9586, 1, 0, 0.0040], $\boldsymbol{I}_8$= [1, 1, 0.5454, 1, 1, 0, 0.0040], I_9= [1, 1, 0.5454, 1, 1, 0, 0.0020]。

接着将向量输入规则库，分别求出与每条规则的强相似度，取最大强相似度输出结果 Z^*，得到推理结果 O_1=0.2, O_2=1.0, O_3=0.6, O_4=0.6, O_5=0.6, O_6=0.6, O_7=0.4, O_8=0.8, O_9=0.8，最后反推出所属类别 label，结果如表 6 第 9 列所示。

经过与原始标签对比，9 条数据分类结果正确，通过实验 1 可以看出，基于该直觉模糊推理系统的异常检测方法是有效的。

3.3 实验结果与分析

实验 2 对比本文方法与其他相关方法，测试数据集 corrected 包括约 300000 条数据，为测试直觉模糊推理方法的分类效果，从中随机选择 10000 条数据，其中包括 5227 条正常数据和 4773 条异常数据，数据分布情况如表 7 第 1 列所示。分别对测试数据进行 8 次实验取平均值。前 3 次实验得到分

表 6 检测数据集

No.	duration	service	flag	byte	urgent	hot	count	label
1	0	private	SF	105	0	0	1	normal
2	0	private	SF	105	0	0	2	R2L
3	0	ecr_i	SF	1032	0	0	501	DoS
4	0	ecr_i	SF	1032	0	0	509	DoS
5	0	ecr_i	SF	1032	0	0	510	DoS
6	0	ecr_i	SF	18	0	0	1	Probe
7	69	telnet	SF	331	0	0	1	R2L
8	0	other	REJ	0	0	0	2	U2R
9	0	other	REJ	0	0	0	1	U2R

类结果及 8 次平均正类正确分类率 TP 和正类错误分类率 FP 值如表 7 所示。

实验分类正确率结果与相关经典方法比较，根据文献[17]的实验部分，对比 Wenke Lee 的异常检测方法，SVM, BP network，免疫算法[12]和遗传模糊系统(基于 Michigan 算法)[14]方法，取平均分类正确率如表 8 所示。

通过表 8 可以看出，基于包含度的直觉模糊推理方法在网络流量异常检测中具有良好的表现，由于 DoS 攻击具有明显的特征属性，各种异常检测方法均能达到良好的检测结果，对于 Probing 攻击，本文提出的方法仅较 SVM 和 BP network 略有逊色，仍然具有较高的检测率，R2L 和 U2L 攻击特征属性并不明显，不同方法特征提取的不同导致检测结果差异较大，而基于直觉模糊推理的方法有规则库作为支持，对该类攻击检测结果均能达到 90%以上，从而说明本文方法的有效性和泛化能力。

4 结束语

本文针对网络流量异常检测问题，提出了基于直觉模糊推理的异常检测方法，将直觉模糊理论针对不确定性和模糊性描述能力强的特性与网络流量特征属性相结合，进一步提高了网络流量的刻画能力。通过 KDD99 标准数据实验，验证了本文方法的可行性，同时与其他相关方法比较，对 R2L 和 U2L 等特征属性不明显的攻击取得较好的检测效果，对直觉模糊理论在网络信息安全领域应用进行

表 7 异常检测结果

类别	数据分布	正确分类结果	TP	错误分类结果	FP	平均 TP	平均 FP
Normal	5227	5156/5155/5150	0.9860	4/2/5	0.0004	0.9637	0.0004
Dos	2547	2481/2450/2498	0.9723	32/25/34	0.0119	0.9735	0.0132
Probing	1241	1219/1132/1228	0.9613	31/31/38	0.0269	0.9621	0.0273
R2L	851	832/836/823	0.9757	58/42/54	0.0603	0.9714	0.0652
U2L	134	126/128/124	0.9403	11/12/9	0.0796	0.9392	0.0783

表 8 算法对比结果(%)

类别	Dos	Probing	R2L	U2L	平均
Wenke	97.00	79.93	75.00	60.00	77.98
SVM	98.57	99.11	64.00	97.33	89.75
BP network	92.71	97.47	48.00	95.02	83.30
IIDV	97.33	93.70	97.50	96.25	96.19
GFS	92.90	68.20	79.40	44.00	71.12
本文	97.35	96.21	97.14	93.92	96.16

了有益的探索。直觉模糊理论在异常检测方面的研究目前还停留在线下检测，如何应用于线上即时检测还需要进一步深入研究。

参 考 文 献

[1] Dimitris K and Elpiniki I. Intuitionistic fuzzy reasoning with cognitive maps[C]. Proceedings of the IEEE International Conference on Fuzzy Systems, Taipei, China, 2011: 821–827.

[2] Chen Cheng-hung. Compensatory neural fuzzy networks with rule-based cooperative differential evolution for nonlinear system control[J]. *Nonlinear Dynamics*, 2014, 75(1): 355–366.

[3] Lei Yang, Lei Ying-jie, and Kong Wei-wei. Technique for target recognition based on intuitionistic fuzzy reasoning[J]. *IET Signal Processing*, 2012, 6(3): 255–263.

[4] Mitchell H B. Pattern recognition using type-II fuzzy sets[J]. *Information Sciences*, 2005, 170(2/4): 409–418.

[5] Hong Peng, Jun Wang, Mario J P J, *et al*.. Fuzzy reasoning spiking neural P system for fault diagnosis[J]. *Information Sciences*, 2013, 235: 106–116.

[6] Luigi L and Larbi B. Using multiple uncertain examples and adaptative fuzzy reasoning to optimize image characterization[J]. *Knowledge Based System*, 2007, 20(3): 266–276.

[7] 雷英杰, 王宝树, 王毅. 基于直觉模糊推理的威胁评估方法[J]. 电子与信息学报, 2007, 29(9): 2077–2081.
Lei Ying-jie, Wang Bao-shu, and Wang Yi. Techniques for threat assessment based on intuitionistic fuzzy reasoning[J]. *Journal of Electronics & Information Technology*, 2007, 29(9): 2077–2081.

[8] 雷英杰, 王宝树, 王毅. 基于直觉模糊决策的战场态势评估方法[J]. 电子学报, 2006, 34(12): 1275–1279.
Lei Ying-jie, Wang Bao-shu, and Wang Yi. Techniques for battlefield situation assessment based on intuitionistic fuzzy decision[J]. *Acta Electronica Sinica*, 2006, 34(12): 1275–1279.

[9] Hwang C M, Yang M S, Hung W L, *et al*.. A similarity measure of intuitionistic fuzzy sets based on the Sugeno integral with its application to pattern recognition[J]. *Information Sciences*, 2012, 189: 93–109.

[10] Boran F E and Akay D. A biparametric similarity measure on intuitionistic fuzzy sets with applications to pattern recognition[J]. *Information Sciences*, 2014, 255: 45–57.

[11] 王毅, 刘三阳, 张文, 等. 基于包含度的直觉模糊相似度量推理方法[J]. 系统工程与电子技术, 2014, 36(3): 497–500.
Wang Yi, Liu San-yang, Zhang Wen, *et al*.. Intuitionistic fuzzy similarity measures reasoning method based on inclusion degrees[J]. *Systems Engineering and Electronics*, 2014, 36(3): 497–500.

[12] 严宣辉. 应用疫苗接种策略的免疫入侵检测模型[J]. 电子学报, 2009, 37(4): 780–785.
Yan Xuan-hui. An artificial immune-based intrusion detection model using vaccination strategy[J]. *Acta Electronica Sinica*, 2009, 37(4): 780–785.

[13] Kuang F J , Xu W H, and Zhang S Y. A novel hybrid KPCA and SVM with GA model for intrusion detection[J]. *Applied Soft Computing*, 2014, 18(5): 178–184.

[14] Abadeh M S, Mohamadi H, and Habibi J. Design and analysis of genetic fuzzy systems for intrusion detection in computer networks[J]. *Expert Systems with Applications*, 2011, 38(6): 7067–7075.

[15] Karami A and Zapata M G. A fuzzy anomaly detection system based on hybrid PSO-Kmeans algorithm in content-centric networks[J]. *Neurocomputing*, 2014, 149(3): 1253–1269.

[16] Hettich S and Bay S D. KDD cup 1999 data[OL]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, 1999.

[17] Guo S Q, Gao C, Yao J, *et al*. An intrusion detection model based on improved random forests algorithm[J]. *Journal of Software*, 2005, 16(8): 1490–1498.

范晓诗： 男，1988 年生，博士生，研究方向为网络信息安全.

雷英杰： 男，1956 年生，博士，教授，研究方向为网络信息安全与智能信息处理等.

王亚男： 女，1988 年生，博士生，研究方向为智能信息处理.