

一种基于信息论模型的入侵检测特征提取方法

宋 勇^{1,2}, 蔡志平²

(1. 湖南民族职业学院工程技术系 湖南 岳阳 414000; 2. 国防科技大学计算机学院 长沙 410073)

【摘要】在网络入侵检测中, 由于原始数据特征维度高和冗余特征多, 导致入侵检测系统的存储负担增加, 检测分类器性能降低。针对该问题本文提出了一种基于信息论模型的入侵检测特征提取方法。它以具有最大信息增益的特征为搜索起点, 利用搜索策略和评估函数迭代调整数据集分类标记、已选取特征子集和候选特征三者之间的相关度, 最后通过终止条件确定选取特征子集。以入侵检测样本数据集为实验数据, 将该方法选取的特征向量运用到支持向量机分类算法中, 在特征维度大幅度降低的情况下, 检测精度变化很小。实验结果证明了本方法的有效性。

关键词 特征选择; 信息熵; 入侵检测; 互信息; 半监督

中图分类号 TP393.08 文献标志码 A doi:10.3969/j.issn.1001-0548.2018.02.017

An Intrusion Detection Feature Extraction Method Based on Information Theory Model

SONG Yong^{1,2} and CAI Zhi-ping²

(1. Department of Engineering Technology, Hunan Vocational College for Nationalities Yueyang Hunan 414000;

2. College of Computer, National University of Defense Technology Changsha 410073)

Abstract In the network intrusion detection, because of the high dimensionality and redundant features of the original data, the storage burden of the intrusion detection system is increased, and the performance of the classifier is reduced. Aiming at this problem, this paper proposes an intrusion detection feature extraction method based on information theory model. The method starts with the feature of maximum information gain, and then iteratively adjusts the correlation among the classification mark of the data set, selected feature subset and candidate feature by search strategies and evaluation functions. Finally, the feature subset is determined by terminating conditions. In the experiment, we chose sample dataset for intrusion detection as the experimental data, and apply feature vector selected by the method to the support vector machine classification algorithm. It is found that the detection accuracy is almost unchanged, in the case that the dimension of the feature is greatly reduced. The results show the validity of the method.

Key words feature selection; information entropy; intrusion detection; mutual information; semi-supervised

随着网络技术的快速发展, 计算机网络在各个领域得到广泛应用, 网络安全也变得日益重要。入侵检测是一种通过采集和分析被保护系统的信息从而发现入侵行为的技术。入侵检测的数据源通常来自网络中的数据包和系统的审计日志等, 这些原始数据通常包含多达几十个特征, 如果直接将它们应用到检测算法中, 将导致检测速度缓慢, 响应不及时。因此, 从大量特征中提取可替代的特征子集是入侵检测数据预处理需要解决的重要问题。

如何选择一度量对量化特征进行测度是在进行特征选取前的一项重要工作。信息论中的熵是对

随机信息中不确定性的度量。熵作为一种良好全局测度, 在特征选取时, 它能够帮助我们寻找到更有作用的特征。本文在深入分析网络流量数据集的基础上, 提出了一种改进的面向入侵检测的基于信息论模型的特征提取方法。

1 相关工作

特征提取作为数据挖掘技术中的一项重要预处理步骤, 在机器学习领域具有重要的研究价值。根据样本集是否有标记, 特征提取方法可以分为3种: 无监督学习算法^[1-2]、半监督学习算法^[3-4]和有监

收稿日期: 2017-01-13; 修回日期: 2017-10-24

基金项目: 国家自然科学基金(601379145)

作者简介: 宋勇(1975-), 男, 副教授, 主要从事网络与信息方面的研究。

督学习算法^[5-6]。有监督模式特征提取算法是在计算类别与特征之间的相关度时,依据标记的类别信息,选择类别区分力最大的特征子集。该算法的优点是分类性能比较高,但缺点是必须在样本集中标记类别,故通用性不强。无监督模式特征提取算法是在没有前面的经验指导下,通过特征间的内在联系,利用样本数据的方差或分离性从特征集中提取重要的特征。该算法的优点是无需具有标记的数据样本,具有一定的通用性,但性能不如有监督模式特征提取算法,并且算法的时间和空间复杂度也高于前者。结合无监督学习算法和有监督学习算法,出现了半监督模式特征提取算法,该算法为了提高大量未标记类别样本数据的特征提取,先对少数样本数据标记类别,并以之为指导。

根据与学习算法结合的方式不同,一般将特征提取算法大致分为4类: Embedded、Filter、Wrapper和Hybrid模型^[7]。Embedded模型将特征提取与学习过程同时进行,在学习训练的过程中选择合适的特征,学习过程完成后,算法所涉及到的特征即为提取的特征结果。该模型的典型算法是决策树算法,如文献[8]提出的ID3和C4.5算法。Filter特征提取算法的评估标准直接由数据集求得,独立于学习算法。该算法以特征之间相互独立为条件,依据每一个特征对于分类的区分能力,选择其中区分能力最强的 n 个特征进行组合。与Filter模型特征提取算法不同,基于Wrapper模型的特征提取算法使用分类器的分类评价来确定特征子集,即首先利用分类器对不同的特征子集组合进行评价,然后对不同的特征子集评价对比得出最优的特征子集。但是该算法一个明显的缺点是算法的时间和空间复杂度比较高。Hybrid模型特征提取算法结合Filter模型和Wrapper模型确定最后的特征子集。它一般先采用前者除掉与类别无关的特征,然后将剩下的特征和类别作为后者的输入,进一步获取最优的特征子集。

2 特征提取方法

特征选择是一种典型的搜索寻优问题。大小为 n 的特征集合,根据选择或不选择两种情况,可以形成 2^n 种集合空间,然后从这个空间中搜索出最优的结果。但是在特征数目多的情况下,如果采用穷尽式的选择算法,那么将会导致特征选择过程占用大量的计算资源、花费大量的时间,因此许多研究人员致力于用一种智能化的搜索算法来寻找最优解。一般特征选择算法必须确定以下4个方面^[9]: 1)搜索

起点; 2) 搜索策略; 3) 特征评估函数; 4) 终止条件。本文从这4个方面出发,根据入侵检测中数据集的特点,提出了一种结合信息熵中的信息增益和条件互信息概念的特征提取方法。

2.1 问题模型及相关定义

设入侵检测训练数据集为 D , $|D|$ 表示其样本的容量,即样本个数。设数据集中的类别标记为 $C=\{C_1, C_2, \dots, C_k\}$, 在本文中将入侵检测数据集类别标记划分为normal和abnormal两类,即将所有正常数据标记为normal,所有攻击数据均标记为abnormal,因此 $|C_i|=2$ 。数据集中类别标记集合 F 是类别 C 的一种组合。设数据集的特征集合为 $T=\{t_1, t_2, \dots, t_m\}$, $|T|$ 表示其特征的个数, T' 为特征集合 T 的一个子集,即 $T' \subset T$ 。特征选取方法的基本思想是从原特征集合 T 中选择出它的一个子集构成新的特征空间。

对于某一个特征 t_x ,其可能取值集合记为 S_{t_x} 。对于 $x \in S_{t_x}$,其概率为 $P(x)$,根据Shannon信息理论, t_x 的信息熵定义为:

$$H(t_x) = -\sum_{i=1}^m p(x_i) \log p(x_i) \quad (1)$$

在信息论里面,式(1)是信息不确定性的一个度量,值越大则表示信息的不确定程度越高。对于训练数据集,可以得到数据集类别标记集合的信息熵为 $H(F)$,也是数据集的经验熵。

当训练数据集中特征 t_y 已知时,特征 t_x 中剩余的不确定性用条件熵来定义:

$$H(t_x | t_y) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log p(x_i | y_j) \quad (2)$$

对于训练数据集中两个随机特征 t_x 和 t_y 的统计依存关系用互信息来定义:

$$MI(t_x; t_y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (3)$$

互信息越大,这两个随机特征之间的联系越紧密;当互信息趋近于零时,这两者之间相互独立。

在机器学习领域,信息增益作为信息论统计中的一个重要概念被广泛应用。对分类系统来说,信息增益是某一特征项 t 在系统中出现与否的信息量之差,它定义为:

$$\begin{aligned} IG(F, t) &= H(F) - H(F | t) = \\ &= -\sum_{i=1}^m p(c_i) \log p(c_i) + p(t) \sum_{i=1}^m p(c_i | t) \log p(c_i | t) + \\ &= p(t') \sum_{i=1}^m p(c_i | t') \log p(c_i | t') \end{aligned} \quad (4)$$

式中, $p(c_i)$ 表示 c_i 类在样本标记集合 F 中出现的概率; $p(t)$ 和 $p(t')$ 表示数据集中特征 t 所有取值的概率; $p(c_i|t)$ 和 $p(c_i|t')$ 表示数据集中特征 t 取某一值时属于 c_i 类的概率。在特征选择中, 某个特征的信息增益越大, 说明它的作用越大, 对 F 也越重要。因此, 在进行特征选择时, 信息增益值大的特征将会是候选特征。但通过式(4)可以看出: 给定一个数据集后, 它的 $H(F)$ 是固定的, 因此 $IG(t)$ 的大小取决于 $H(F|t)$ 。当特征 t 的取值都不相同时, $H(F|t)$ 将会等于0, 这时 $IG(t)$ 将最大, 显然这种情况的特征意义不大。因此根据C4.5^[8]算法, 为了解决这个问题, 一般采用信息增益率, 定义如下:

$$IGR(F, t) = \frac{IG(F, t)}{H(t)} \quad (5)$$

式中, $H(t)$ 表示特征 t 的信息熵。

在入侵检测数据集中, 类别标记是由离散型数据{normal, abnormal}表示, 但有些样本特征的值是连续型的数据, 如duration、src_bytes等特征的值。在评判这些连续特征与类别标记的相关性时, 需要对连续特征进行离散化。由于入侵检测数据集的类别标记只有normal和abnormal两类, 所以对于连续型特征的值域只需要将它们划分为两个区域, 分别对应前面的两种标记类型。具体划分点的计算方法如下。

对于连续特征 T , 对它的样本数据集中不重复的值经排序后为 $v_1 < v_2 < \dots < v_n$, 并按下面的步骤进行计算:

- 1) 设 $i=1$;
- 2) 令 $v=(v_i+v_{i+1})/2$; 设样本数据集中特征 T 的值 $\leq v$ 时为0, $>v$ 时为1;
- 3) 令 $I_i=I(T; C)$, 设 $i=i+1$, 重复步骤2), 直到 $i \leq n$;
- 4) 输出 $\{I_1, I_2, \dots\}$ 值最大对应的 v 值。

v 值就是特征 T 值域的划分点, 利用该值将特征 T 的值域分成两个区域, 按照约定分别将在这两个区域的值指定为0或者1。

2.2 搜索起点

被选取特征子集的初始状态称为算法的搜索起点, 它对搜索结果有十分重要的影响。当被选取特征子集为空时, 可以按照一定方法逐个地向被选取特征子集加入特征; 当被选取特征子集为全集时, 可以按照一定方法不断地删减特征; 当被选取特征子集从特征集的某个子集开始时, 那么搜索策略一般采用随机或者启发式。

本文根据最优被选取特征集中一定包含了对数

据集类别信息增益最高的特征这一原则, 搜索起点为: $\max\{IGR(t_1), IGR(t_2), \dots, IGR(t_n)\}$ 。IGR函数为某一特征的信息增益率, 具体如式(5)所示。

2.3 搜索策略

被选取特征集搜索算法一般采用顺序搜索和随机搜索。顺序搜索算法采用顺序的方式, 从特征集中选取特征到被选取特征集, 从而逐步扩展搜索, 如顺序前向搜索和顺序后向搜索等。随机搜索算法包括模拟退火、遗传算法和集束式搜索等。本算法采用顺序搜索的方式, 基本步骤如下:

- 1) 初始化: 提取特征子集 $T'=\max\{IGR(t_1), IGR(t_2), \dots, IGR(t_n)\}$, $T'=\{\text{除}T'\text{外所有的属性}\}$;
- 2) $\forall t \in T$, 按特征评估函数计算特征 t 的值;
- 3) 计算 T 集合中所有特征的值, 获得具有最大值的特征 t_{\max} , 设置 $T=T-\{t_{\max}\}$, $T'=T' \cup \{t_{\max}\}$;
- 4) 重复步骤2)和步骤3)直至满足终止条件;
- 5) 输出特征集 T' 。

2.4 特征评估函数

特征评估函数是特征选择的评价标准, 即可以利用单独的特征独立进行评价, 也可以利用某个特征子集整体进行评价。

在文献[10]中, 针对利用信息增益思想提取特征时存在的一些问题, 分别提出了增加了频度、集中度和分散度3项测试指标的方法。在MaxMI^[11]特征选择算法中, 提出了最大化互信息的基本思想, 即提取的特征子集应当尽可能多地提供关于类别的信息, 也就是最大化 $I(C; T')$ 。这两个算法最大的问题就是只考虑到了各特征与类别之间的关系, 而没有考虑到提取的特征之间的相互影响。在PG-HMI^[12]算法中, 虽然评估函数即考虑到了特征选取准则既与属性能够提供的新信息量 $I(C; f|S)$ 相关, 也与属性和类别标记的相关度 $I(C; f)$ 相关, 但是没有考虑到新选择的特征 f 可能会对已提取的特征子集产生冗余。

结合前面的研究成果和信息论模型, 提出一种基于信息论模型的特征选择算法, 该算法即考虑到在已选取特征子集确定的条件下, 候选特征与类别之间的关系, 又将候选特征与已选取特征子集之间的相关度作为该特征是否选取的重要依据。

设原始特征集为 $T=\{T_1, T_2, \dots, T_n\}$, 选取的特征子集为 $T'=\{T'_1, T'_2, \dots\}$, 候选特征为 t 。在已选取特征子集确定的条件下, 候选特征与类别之间的相关性根据式(2)和式(3)得到:

$$\begin{aligned}
& \text{MI}(F;t|T') = \text{MI}(F;t|(T'_1, T'_2, \dots)) = \\
& \text{MI}(F, (T'_1, T'_2, \dots); t) - \text{MI}((T'_1, T'_2, \dots); t) = \\
& H(F, (T'_1, T'_2, \dots)) + H(t) - H(F, (T'_1, T'_2, \dots), t) - \\
& H(T'_1, T'_2, \dots) - H(t) + H((T'_1, T'_2, \dots), t) = \\
& H(F, T'_1, T'_2, \dots) + H(T'_1, T'_2, \dots, t) - \\
& H(F, T'_1, T'_2, \dots, t) - H(T'_1, T'_2, \dots) \quad (6)
\end{aligned}$$

式中, $H(F, T'_1, T'_2, \dots)$ 表示样本数据集分类标记集合与当前已选取特征子集数据集的联合熵; $H(T'_1, T'_2, \dots, t)$ 表示当前已选取特征子集数据集与候选特征数据集的联合熵; $H(F, T'_1, T'_2, \dots, t)$ 表示样本数据集分类标记集合、当前已选取特征子集数据集和候选特征数据集的联合熵; $H(T'_1, T'_2, \dots)$ 表示当前已选取特征子集数据集的联合熵。根据条件互信息的性质, 式(6)的值大于等于零。当值等于零时, 表示候选特征与样本数据集分类标记不相关。值越大表示在当前已选取特征子集固定的条件下, 两者的相关性越强。

候选特征与已选取特征子集之间的相关度定义为:

$$\begin{aligned}
& \text{MI}(t; T') = \text{MI}(t; (T'_1, T'_2, \dots)) = \\
& H(t) - H(t | (T'_1, T'_2, \dots)) = \\
& H(t) - H(t, T'_1, T'_2, \dots) + H(T'_1, T'_2, \dots) \quad (7)
\end{aligned}$$

式中, $H(t)$ 表示候选特征数据集的信息熵; $H(t, T'_1, T'_2, \dots)$ 表示候选特征数据集与当前已选取特征子集数据集的联合熵; $H(T'_1, T'_2, \dots)$ 表示当前已选取特征子集数据集的联合熵。式(7)的值等于零时, 表示候选特征与已选取特征子集不相关。但是根据入侵检测数据集的分析, $\text{MI}(t; T')$ 为零的可能性几乎很小, 所以必须设置一个阈值控制候选特征与已选特征子集之间的相关度。

综合上述得到本算法的搜索评估函数为:

$$\max \{ \text{MI}(F; t_i | T') > 0 \cap \text{MI}(t_i; T') < \varepsilon \} \quad (8)$$

该函数表示在已选取特征子集确定的条件下, 从所有候选特征中选取与已选取特征子集的互信息小于阈值 $\varepsilon \in [0, 1]$ 、并且与样本标记的互信息值最大且大于零的特征。

2.5 终止条件

算法的终止条件可以选择固定的迭代次数和特征数目来进行, 也可以自己定义终止条件函数。本算法的终止条件为:

$$\begin{aligned}
& \{t_i | \text{MI}(F; t_i | T') > \beta\} = \emptyset \\
& \text{or } \{t_i | \text{MI}(t_i; T') < \varepsilon\} = \emptyset \quad (9)
\end{aligned}$$

上述终止条件表示当所有候选特征在已选取特

征子集确定的条件下, 与样本标记的互信息都小于阈值 $\beta \in [0, 1]$ 时终止, 或者在前面表达式不成立的条件下, 所有候选特征与已选取特征子集的互信息大于阈值 $\varepsilon \in [0, 1]$ 。

3 实验结果及分析

3.1 实验数据

为了检验本文提出的特征提取算法的有效性, 选用了KDD99数据集作为实验数据集。KDD CUP 1999数据集是由麻省理工学院的林肯实验室采用1998年美国国防部高级规划署的入侵检测数据集建立起来的。数据提取了41个特征, 包括34个连续型特征和7个离散型特征。

由于原始数据集过于庞大, 因此只选取KDD99中的kddcupdata10percent数据集。该数据集包含494 021条连接记录, 其中攻击连接记录396 743条, 正常连接记录97 278条, 攻击类型22个。为了验证本算法提取的特征集对未知攻击类型的有效性, 在训练数据集中只包含12种攻击类型, 记录条数4 000条, 其中攻击记录2 000条。测试数据集中包含22种攻击类型, 记录条件10 000条, 其中攻击记录3 000条。

3.2 实验设置

在实验过程中, 本方法的主要参数是搜索函数和终止条件函数的两个阈值。其中通过阈值 β 可以控制选取特征子集与样本标记的相关性程度。如果 β 太大, 它们的相关程度变高, 入选特征数少, 可能一些较重要的特征无法选取, 从而导致分类算法精确度不高; 若 β 值太小, 入选特征数变多, 不利于分类算法的性能。通过多次实验, 确定 β 为0.35。阈值 ε 主要控制被提取特征之间的相关性程度, 较大时导致被提取特征之间的冗余度较高, 因此 ε 值设为0.05。

实验环境为CPU Inter Core i7 2.50 GHz, 内存8.00 GB, 操作系统Windows8 64位, 开发工具Matlab2010。

3.3 特征选取结果和分析

根据上面的实验数据和实验设置, 利用本文提出的特征提取算法, 得到特征子集的属性ID为1、3、5、6、12、23、24、32、33、36、37、40。为了验证特征选取的有效性, 本文使用支持向量机分类算法(support vector machines, SVM)对测试数据集进行两组实验, 分别是原始特征全集和本文提出的方法所选出的特征子集, 并进行入侵检测性能评估。本文采用检测时间、检测率和误报率3个评价指标进行

性能评估, 其中:

$$\begin{aligned} DR &= \frac{DC}{AC} \\ FPR &= \frac{MIC}{NIC} \end{aligned} \quad (10)$$

式中, DR表示检测率; DC表示检测的异常记录个数; AC表示真实入侵记录个数; FPR表示误码报率; MIC表示正常记录误报为异常的个数; NIC表示正常记录个数。实验结果如表1所示。

表1 特征提出前后的效果比较

特征数	检测时间/s	检测率/%	误报率/%
原始41个特征	23.2	87.5	1.64
提取12个特征	11.4	86.7	1.71

从表中可以看出, 使用本文的特征提取算法, 特征维度降低了70.7%, 但是检测率和误警率影响不大, 算法的处理速度提高了50.8%。

4 结束语

本文面向入侵检测提出了一种基于信息论模型的特征提取方法, 本方法依据最优被选取特征集中一定包含对数据集类别信息熵增益最高的特征这一原则, 首先从所有特征中选取信息增益最大的特征作为搜索起点, 然后充分考虑样本数据集分类标记、已选取特征子集和候选特征三者之间的信息相关性基础上顺序搜索最合适的特征, 最后在搜索终止条件下完成所有特征的选取。通过在KDD99数据集上的实验表明, 本文提出的特征选择方法在保证检测准确率的前提下, 大大降低了入侵检测数据的特征维度, 从而减轻了检测系统的存储负担, 提高了入侵检测分类器的性能。后续将在真实系统架构^[13]中验证本文方法在大数据流量情况的性能, 并考虑在Spark等大数据处理平台^[14]上的实现方法。

参 考 文 献

[1] CHIMIENTI M, CORNULIER T, OWEN E. The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data[J]. *Ecology & Evolution*, 2016, 6(3): 1948-1952.

[2] 杨国亮, 谢乃俊, 王艳芳, 等. 基于低秩稀疏评分的非监督特征选择[J]. *计算机工程与科学*, 2015, 37(4): 649-656.

YANG Guo-liang, XIE Nai-jun, WANG Yan-fang, et al. Unsupervised feature selection based on low rank and sparse score[J]. *Computer Engineering and Science*, 2015, 37(4): 649-656.

[3] ZHENG Zhao, LIU Huan. Semi-supervised feature selection via spectral analysis[C]//Proceedings of the 7th SIAM International Conference on Data Mining. [S.l.]: DBLP, 2007: 1193-1201.

[4] 史彩娟. 网络空间图像标注中半监督稀疏特征选择算法研究[D]. 北京: 北京交通大学, 2015.

SHI Cai-juan. Research on semi-supervised sparse feature selection for image annotation in web space[D]. Beijing: Beijing Jiaotong University, 2015.

[5] YU Lei, LU Ling. Feature selection based on loss-margin of nearest neighbor classification[J]. *Pattern Recognition*, 2009, 42(9): 1914-1921.

[6] 郑莹斌. 有监督的视觉特征提取算法研究[D]. 上海: 复旦大学, 2013.

ZHENG Ying-bin. Research on supervised visual feature extraction algorithms[D]. Shanghai: Fudan University, 2013.

[7] MANSOORI E G, SHAFIEE K S. On fuzzy feature selection in designing fuzzy classifiers for high-dimensional data[J]. *Evolving Systems*, 2015, 7(4): 1-11.

[8] QUINLAN J R. Programs for machine learning[M]. San Mateo, CA: Morgan Kaufmann, 1993.

[9] 张丽新, 王家钦, 赵雁南, 等. 机器学习中的特征选择[J]. *计算机科学*, 2004, 31(11): 180-184.

ZHANG Li-xin, WANG Jia-qin, ZHAO Yan-nan, et al. Feature selection in machine learning[J]. *Computer Science*, 2004, 31(11): 180-184.

[10] 李玲, 刘华文, 徐晓丹, 等. 基于信息增益的多标签特征选择算法[J]. *计算机科学*, 2015, 42(7): 52-56.

LI Ling, LIU Hua-wen, XU Xiao-dan, et al. Multi-label feature selection algorithm based on information gain[J]. *Computer Science*, 2015, 42(7): 52-56.

[11] 唐亮, 段建国, 许洪波, 等. 基于互信息最大化的特征选择算法及应用[J]. *计算机工程与应用*, 2008, 44(13): 130-133.

TANG Liang, DUAN Jian-guo, XU Hong-bo, et al. Mutual information maximization based feature selection algorithm in text classification[J]. *Computer Engineering and Applications*, 2008, 44(13): 130-133.

[12] WANG H, SUN H B, ZHANG B M. PG-HMI: Mutual information based feature selection method[J]. *Pattern Recognition & Artificial Intelligence*, 2007, 20(1): 55-63.

[13] CAI Zhi-ping, WANG Zhi-jun, ZHENG Kai, et al. A distributed TCAM coprocessor architecture for integrated longest prefix matching, policy filtering, and content filtering[J]. *IEEE Trans Computers*, 2013, 62(3): 417-427.

[14] 方峰, 蔡志平, 肇启佳, 等. 使用Spark Streaming的自适应实时DDoS检测和防御技术[J]. *计算机科学与探索*, 2016, 10(5): 601-611.

FANG Feng, CAI Zhi-ping, ZHAO Qi-jia, et al. Adaptive technique for real-time DDoS detection and defense using spark streaming[J]. *Journal of Frontiers of Computer Science and Technology*, 2016, 10(5): 601-611.