

基于多元特征的智能型生物识别模型

孙傲冰 张德贤 张 苗

(河南工业大学信息科学与技术学院 郑州 450001)

摘 要 随着生物识别技术的不断成熟,各种成形产品和设备不断进入市场,为生物识别手段代替传统的身份鉴别方法提供了基础。基于单一生物特征的识别模型由于过度依赖于一种识别模式,难以抵御针对型的欺骗方式。通过分析人类智能识别的行为,提出了一种基于多元特征的智能型生物识别模型。该模型能够同时捕捉目标的多种生物特征,通过选择可信度最高的识别模式进行组合识别,防止了针对型欺骗;通过引入多特征的交叉索引模型,提高了系统对复杂生物特征的检索效率;通过系统记录的历史数据的参与,实现了一种目标特征变化敏感识别方式,使系统在某一特征变化大于预设阈值时进行更高精度的生物识别。基于多元特征的生物识别模型增加了系统的欺骗代价,在保证系统的识别精度和效率的同时,降低了系统遭受恶意侵入的风险。

关键词 生物识别,多元特征,欺骗代价

中图分类号 TP391 文献标识码 A

Multiple Features Based Intelligent Biometrics Verification Model

SUN Ao-bing ZHANG De-xian ZHANG Miao

(School of Information Science and Technology, Henan University of Technology, Zhengzhou 450001, China)

Abstract With the development of biometrics verification technology, related devices and implements are put into market, which make it possible that the devices replace traditional means. But current single feature based verification devices depend on one single means and own very low cheat cost, which can not defense the specific cheat mode. We analyzed the intelligent verification of human beings and brought forward one multiple features based verification model. It can capture several biometric features at same time and select several features for verification according to need. The multiple features based cross index means can improve the search efficiency of complicated features. The joint of historic data in the model enables the system focuses on the changed features and provoke the verification with high precision. The multiple features based verification enhances the cheat cost and guarantees the verification accuracy to avoid the possible hostile access.

Keywords Biometric verification, Multiple features, Cheat cost

传统系统使用实物密钥(如钥匙盘、IC卡等)和虚拟密钥(如个人识别码/密码等)来鉴别用户身份。这些密钥均存在着被借用、冒用的可能,对系统安全带来极大的威胁^[1]。生物认证技术是一项新兴的安全技术,它将信息技术与生物技术相结合来验证用户身份。从统计意义上来说,人类的指纹、掌形、虹膜等生理特征都存在着唯一性,因而这些特征都可以成为鉴别用户身份的依据^[2]。与传统的身份鉴别技术相比,指纹、掌纹、虹膜等生物特征被复制或盗取的机会较小,因此基于生物特征的识别系统拥有相对更高的安全性,具有巨大的市场发展潜力^[3]。当前的生物识别系统普遍采用基于单一特征的生物识别模型和方法。但是基于单一生物特征的识别技术或者模型都有一定的缺陷。如当前的指纹识别、虹膜识别的欺骗方法,都是针对现有的识别系统过度依赖于一种识别模式的缺陷实施的^[4]。

1 基于单一特征的生物识别

基于单一特征的生物识别指系统仅基于一种生物特征构

建识别模型^[5]。在识别过程中,系统通过捕获设备获取识别目标的对应特征,并与系统存储的所有对象的特征进行对比;根据匹配结果进行比对,判断目标是否是系统准入的对象。基于单一特征的生物识别模型包括以下流程:

- 1) 捕获设备获取用户原始的生物特征信息 A (如指纹、虹膜图像等)。
- 2) 系统从原始生物特征信息 A 中提取生物特征信息 I_m ($I_m = \varphi(A)$) (如特征点分布、直方图信息等)。
- 3) 系统将 I 与特征库 L 中保存的其它用户的特征信息 $I_1, I_2, \dots, I_n$ 进行对比,分别计算其对应的相似度 $R = (R_1, R_2, \dots, R_n)$, $R_i = \psi(I_i, I_m)$ 。
- 4) 系统返回相似度大于阈值 y 且为所有相似度中最大值的特征信息对应的用户信息 U_i ($U_i = \text{MAX}(R), R_i > y$);如无,则报告为非法用户。

基于单一特征的生物识别系统,其识别精度受限于特征提取模型 φ 。当前捕获设备获取的用户原始生物特征信息 A

到稿日期:2009-03-17 返修日期:2009-05-26 本文受河南省创新型科技人才队伍建设工程项目(094200510009)资助。

孙傲冰(1978—),男,博士,主要研究方向为系统结构, E-mail: sunaobing@gmail.com; 张德贤(1962—),男,教授,主要研究方向为模式识别; 张 苗(1979—),女,讲师,主要研究方向为模式识别、信息安全等。

主要为图像信息。 φ 决定系统能否精确地提取图像的特征信息,保证其在识别过程中的惟一性(与惟一用户对应)、可操作性(易于计算相似度)、容错性(图像发生翻转等位置偏差是否仍然可以识别)。系统的识别速度受限于系统采用的相似度计算和对比方法 ψ 。如果 ψ 较复杂,则系统的识别速度会整体受到限制。当前通常通过建立索引的方法,缩小需要计算相似度对的范围,来提高 ψ 的速度^[6]。

在 20 世纪 90 年代,由于硬件价格普遍偏高,基于单一特征的生物识别模型得到了较多的采用。如基于指纹、掌纹、人脸、唇纹、虹膜、签名、语音等特征的识别系统已经在现实生活和工作中得到了广泛的应用。表 1 所列为当前常用的基于单一特征的生物识别方法。但是事实证明,基于单一生物特征的身份识别仍然有相当的缺陷。我们虽然证明了生物特征的惟一性,但是不能保证生物特征不会被复制。例如用户的指纹可以广泛地存在于他触摸过的任何地方,指纹的窃取也许比偷一把钥匙还要容易。而其它一些高级生物特征,如虹膜图像信息一旦失窃,就会使整个系统失效或不安全。我们可以因丢失钥匙更换门锁,但无论如何也换不了我们的虹膜^[7]。系统入侵者可以通过获取用户的指纹或虹膜图像,通过模具来复制出欺骗道具(如指纹手套、虹膜眼镜等),从而非法侵入系统。此外,当前的生物特征自动提取算法往往是基于图像信息提取特征点,识别速度较慢,软硬件代价较高;由于捕获的特征受污染(如声音特征)或发生形变(图像特征)等情况的发生,系统的误识率和拒识率较高^[3]。

表 1 常见的基于单一特征的生物识别方法

特征类型	识别精度	误识率	拒识率	识别速度
签名	较低	高	高	较快
唇纹	较低	高	高	较快
指纹	较高	低	较高	快
掌纹	较高	低	高	快
人脸	较高	较低	高	较慢
虹膜	高	较低	高	慢
声音	低	较高	较高	快

定义生物识别方法的欺骗代价(C)为一个常数,它与完成系统欺骗所要付出的代价成正比。一般情况下它与获取生物特征信息的难度 P 以及制造可以欺骗系统的道具所花费金钱的数量 M ^[8] 有关。

$$C = P + M = \sum_{i=1}^m p_i + \sum_{i=1}^n m_i \quad (1)$$

随着时间的推移和技术手段的进步,窃取生物特征以及制造这种道具的金钱花费会有所变化,因此 C 是一个可以变化的值。

2 人类的智能识别行为

人类本身不具备高精度识别生物特征的能力,在利用生物特征进行识别时主要依据日常积累的印象。但是人类之间一旦熟识,很少会发生误识的可能,这与人类识别的智能性是分不开的。人类的智能识别行为具有以下特点。

- 多特征综合:人类识别行为很少只基于一种特征,而是综合体形、语言、人脸等多种特征。只有各种特征完全吻合,才能做出判断。它甚至能结合一些智能的行为特征进行判断,如语言、举止习惯、风格等。

- 变化敏感性:人类的识别行为对于识别对象的特征变化是敏感的。在其它特征都吻合的条件下,人类往往会关注识别目标发生变化的特征,并判断变化是否在可以接受的范围内;如果变化不可接受,会引发对目标的重新识别。

- 历史承继性:如果识别对象的特征变化在可接受的范围内,用户会接受新的特征。然后用它们替代原有的特征,并用作下次判断的依据。这使一些简单的特征也可以作为判断用户身份的依据参与识别过程,如用户的面形、体形等。

人类的智能识别行为为建立更加准确和高效的生物识别系统提供了启发。我们可以通过引入多种生物特征的组合识别来增加系统欺骗的代价;通过引入变化敏感性,先对多种生物特征进行较低精度识别,使系统集中识别发生变化的特征(如对其调用更高精度的识别方法),降低其它特征的识别代价;通过引入历史承继性,将一些低精度的特征引入识别过程,如用户的身高、体重信息(随季节会发生变化,但是在一个区间内)。基于多元特征的生物识别模型需要系统同时采集多种生物特征,增加了硬件实现上的成本。随着硬件价格的下降,可以在有效地控制成本的前提下采用基于多元生物特征的识别模型,以提高识别的精度和速度^[9]。

3 基于多元特征的生物识别

3.1 模型定义

基于多元特征的生物识别指系统仅基于多种生物特征构建识别模型。在识别过程中,系统通过捕获设备获取识别目标对应的多种生物特征,依照各种生物特征的优先级和系统存储的对象的特征进行对比,判断目标是否是系统准入的对象^[10]。基于多元特征的生物识别模型包括以下主要流程:

- 1) 捕获设备获取用户多种原始的生物特征信息 $A = (A_1, A_2, \dots, A_n)$ (如指纹、虹膜图像等)。

- 2) 系统从原始生物特征信息 A 中提取生物特征信息集合 $I = (i_1, i_2, \dots, i_n) (i_i = \varphi_i(A_i))$, 其中 φ_i 为特征提取模型。

- 3) 系统将集合 I 与特征库 L 中保存的其它用户的特征信息 $I_1, I_2, \dots, I_n$ 进行对比,分别计算其对应的相似度 $R = (R_1, R_2, \dots, R_n)$, R 的计算公式如式(2)所示。

$$R_i = \prod_{j=1}^n \kappa(\varphi_i(i_j, i_j'), y_i) \cdot \sum_{i=1}^n \varphi_i(i_j, i_j') \cdot \theta_i \quad (2)$$

其中, φ_i 用以计算两个特征值间的相似度, y_i 为对应特征的控制阈值; $\kappa(a, b)$ 为控制函数,如果 $a > b$, 则返回 a , 否则返回 0; θ_i 为对应的优先级系数, $\sum_{i=1}^n \theta_i = 1$ 。

- 4) 系统返回相似度大于阈值 y 且为所有相似度中最大值的特征信息对应的用户信息 $U_i (U_i = \text{MAX}(R), R_i > y)$; 如无, 则报告为非法用户。

基于多元特征的生物识别,要求识别对象同时满足多种待识别生物特征的要求。这就极大地增大了系统的欺骗代价 C_{MF} , 它是各种基于单一特征生物特征识别的欺骗代价的乘积, 如式(3)所示。

$$C_{MF} = \prod_{i=1}^n C_i = \prod_{i=1}^n (\sum_{j=1}^m p_j + \sum_{j=1}^n m_j) \quad (3)$$

这也是基于多元特征生物识别模型的优势所在。这一点是很容易理解的。例如对于一个基于 4 种生物特征(包括脸形、虹膜、身高、体重)的识别系统,想要欺骗系统,需要有这样的一个欺骗者,他的身高、体重、脸形、虹膜与他要仿冒的对象完全相同。在一般情况下冒充者很难同时达到这些要求,或者获取全部的信息。

3.2 多元特征空间交叉索引

基于多元特征的生物识别系统,每一个用户关联多种生物特征 $A = (A_1, A_2, \dots, A_n)$, 如图 1 所示。通过特征值的关联,用户特征信息可以被关联在一个多维的空间中 $T = (T_1, T_2, \dots, T_n)$, 其中某一种类型的特征值可以构成空间的一个

特征值平面 $T_i=(a_1,a_2,\cdots,a_n)$,如图 2 所示。单对某一种类型的特征值建立索引进行检索,则分类后的集合中仍然有较多的用户特征值需要对比;基于多元特征值构建用户特征值空间,使得每一种类型的特征值都可以参与到特征值的筛选或分类中去,且多次筛选后需要进行相似度计算的集合中有较少的元素,从而大大提高了系统响应的速度。

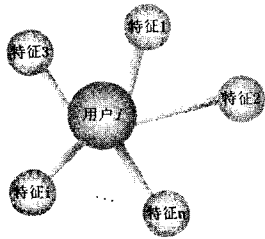


图 1 用户多元生物特征模型

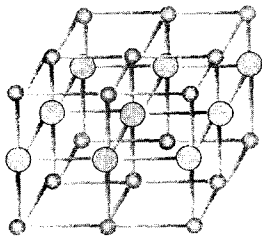


图 2 生物特征关联空间

定义基于一种生物特征的检索速时间为 t ,它与系统存储的识别对象的数量 n 成正比。对于一些复杂特征的识别, n 值较大时,系统识别速度会极大地降低。通过多特征的参与,可以把需要进行特征对比的对象数量极大地降低,这将极大地加速系统的特征检索。例如,对于一个以虹膜识别为主体的多元特征识别系统,通过引入身高、体重、性别特征的筛选,可以将需要进行虹膜特征对比的用户特征值数量缩小到最小,这将极大地提高系统的响应速度^[8]。

3.3 时间相关性

人的一生中绝大多数生物特征不会发生显著的改变,这也是生物特征可用作识别用途的优势之一。但是因为测量环境、时间、习惯的变化,系统实际测试的值会出现一定的偏差。一些简单的生物特征会随时间发生一些变化,如实测身高、体重随着月份的变化。此外,对于大多数生物特征来讲,因使用采集设备的习惯发生变化,采集到的特征值也会有一定的不匹配度。总体来讲,这种随时间变化的偏差可以概况为以下几种情况。

(1) 区间变化型: $\lambda=\gamma\pm|\Delta t|$,即实测特征值 λ 围绕一个固定值 γ 变化,变化值为 Δt 。设 t 为系统容许的最大变化量,若 $\Delta t>t$,则相似度为 0;对应的相似度为 $R=1-\Delta t/t$,如式(4)所示:

$$R=\begin{cases} 0, & \Delta t>t \\ 1-\Delta t/t, & \Delta t\leq t \end{cases} \quad (4)$$

(2) 时间变化型: $\lambda=\gamma(i)\pm\Delta t$,即实测特征与上次测得的特征值有一定的关系,并在其基础上进行微调。它是一个时间相关的变量, $\Delta t=\lambda(i_1)-\lambda(i_2)$ 。设 t 为系统容许的最大变化量(阈值),若 $\Delta t>t$,则相似度为 0;对应的相似度 $R=1-\Delta t/t$ 也如式(4)所示。

所有的生物特征的变化都可以归结为两种类型中的任一种。但区间变化型往往忽略生物特征随时间变化的特性。另外,为了使系统的拒识率降低,区间变化型的阈值 t 不可以定得过高。从统计学分析,所有即时取得的生物特征是随时间变化的,无论是身高、体重这些简单特征,还有一些复杂的图形特征(包括指纹、虹膜等),这和季节、天气、操作习惯的改变都会有很大的关系。通过引入时间变化量,捕捉最后一次认可的特征并以此作为下一次识别的依据,可以将阈值 t 进一步缩小,以此来提高系统的精度,降低系统的拒识率^[11]。

3.4 增量敏感与分级识别

人类进行识别的过程中,先以简单的一般印象作为识别人的基础。如果全部印证的话,也就确认了某人的身份。如

果有一项不符合,转而使用最为熟悉的特征再次来鉴定对方的身份,这里体现了识别过程的增量敏感与分级识别的思想。

定义一个二级的识别模型,选择一组特征 $A=(A_1,A_2,\cdots,A_n)$,其中 A_i 为选定的主特征类型(权值最大、精度最高的特征),它由低精度和高精度两级识别过程组成。首次识别时,使用与 A_i 对应的精度较低、速度较快的识别方法。设首次识别后各特征对应的增量为 $\Delta=(\Delta_1,\Delta_2,\cdots,\Delta_n)$,当存在 Δ_i 大于对应的阈值时,系统转而使用主特征 A_i 对应的高精度的识别过程进行识别。

从识别速度上讲,由于基于多元特征的识别可以依靠多种特征的共同认证,在绝大多数情况下不需要太高精度的识别过程,因此整体速度较快。而基于单一特征的识别,只能依靠一种特征做出判断,因此只能通过高精度的识别才能保证识别的准确率,因此速度较慢。

3.5 模型工作流程

图 3 所示为基于多元特征的生物识别的主要流程,它包括信息采集和生物识别两个主要流程。信息采集流程包括:①系统接收录入的用户身份及访问控制信息;②系统将信息保存到用户信息库中;③系统采集用户多种生物特征原始信息(如图像信息);④系统对多种原始特征信息进行预处理并获取对应的生物特征值;⑤系统将特征值存入特征信息库,链接到各索引库,并和用户信息关联。生物识别流程包括以下主要步骤:①采集设备同时捕捉识别对象的多种生物特征原始信息;②系统对多种原始特征信息进行预处理,获取对应的生物特征值;③系统按照建立的交叉索引检索到需要对比的特征值的对象集合;④系统计算提交特征与集合中对象的特征值间的相似度,若变化超出允许范围则转①;⑤系统按照相似度的计算结果,给出识别结果(即对应的用户信息),或报告为非法用户。

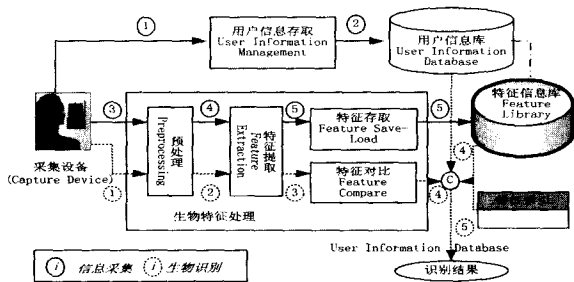


图 3 基于多元特征的生物识别流程

4 实验

本文通过一组基于单一特征和多元特征的生物识别实验来对比两种不同模型的性能,如表 2 所列。生物特征的组合可以有多种选取方法,例如人脸识别和虹膜识别的组合、指纹和掌纹的组合,因为这些特征可以在特征捕获时由一套设备同时或先后获取。此外,一些简单的生物特征,如身高、体重信息虽然较为简单,但是作为区间变化型的生物特征也可以参与到识别过程,并用于建立多特征交叉索引,这可以有效地加速特征识别的过程。仅基于单一特征的生物识别方法(如仅基于虹膜的图像特征分类),需要进行高精度的识别。此外,由于只有一种特征用于特征库检索,识别过程非常耗时。实验中将身高特征、体重特征、性别特征参与到特征索引,并实现高精度特征的分级识别,对系统识别速度的提高有很大的帮助。基于多元特征的生物识别在保证系统识别精度的同时,识别的速度也没有降低。本实验中主要采取了一种精度

较高的特征和其它低精度特征组合的方法。对于对精度有较高要求的系统,可以采取多种高精度特征组合的方法。

表2 基于单一特征和基于多元特征生物识别模型性能对比

模型类型	参与生物特征	识别时间(s)	精确度
基于单一特征	指纹	2.9	93.1%
基于单一特征	虹膜	3.1	97.2%
基于单一特征	签名	2.3	92.3%
基于单一特征	语音	2.7	91.4%
基于多元特征	掌纹、指纹、体重	1.1	96.1%
基于多元特征	人脸、虹膜、身高、体重	2.1	98.2%
基于多元特征	签名、身高、体重	1.3	96.3%
基于多元特征	语音、身高、体重	1.7	96.1%

结束语 本文针对基于单一特征的生物识别模型中存在的缺陷,参照人类智能识别行为提出了一种基于多元特征的生物识别模型。它同时捕获识别对象的多种生物特征进行识别,并将多种特征应用于特征交叉索引和分级识别,实现了模型的历史相关性;通过关注特征与时间变化的关系,使模型优先关注发生变化的特征,实现了模型的增量敏感性。基于多元特征的生物识别模型,在保证系统的识别精度的同时,提高了系统的识别速度,具有广阔的应用前景。

参考文献

- [1] 孙冬梅,裴正定.生物特征识别技术综述[J].电子学报,2001,29(12A):1744-1748
- [2] 胡瑾,田捷,陈新建,等.一种指纹合成的方法及其实现[J].软件

学报,2007,18(3):517-526

- [3] 田启川,潘泉,程咏梅,等.基于过零检测的虹膜特征提取算法[J].电子与信息学报,2006,28(8):1231-1234
- [4] 何国辉,甘俊英,李春芝,等.人脸与虹膜特征层融合模型的研究[J].电子学报,2007,35(7):1365-1371
- [5] 刘培强,朱大铭,谢青松,等.两元指纹向量聚类问题的复杂性与改进启发式算法[J].软件学报,2008,19(3):500-510
- [6] Ross A, Nandakumar K, Anil K. Handbook of Multibiometrics [M]. Berlin: Springer, 2006
- [7] Ross A, Govindarajan R. Feature level fusion using hand and face biometrics [C] // Proc. of SPIE Conference on Biometric Technology for Human Identification, March 2005:196-204
- [8] Tan X J, Bhanu B, Lin Y Q. Fingerprint Classification Based on Learned Features[J]. IEEE Trans. on Systems, Man, and Cybernetics, 2005, 35(3):287-300
- [9] Li J, Wang H. Fingerprint Indexing Based on Symmetrical Measurement [C] // Proc. of the 18th International Conf. on Pattern Recognition, Washington: IEEE Computer Society, 2006: 1038-1041
- [10] Meier-Ewert S, Lange J, Gerts H, et al. Comparative Gene Expression Profiling by Oligonucleotide Fingerprinting[J]. Nucleic Acids Research, 1998, 26(9):2216-2223
- [11] 陈宏,田捷.检验配准模式的指纹匹配算法[J].软件学报,2005,16(6):1046-1053

(上接第208页)

某车型类别的置信度。 T_1 , T_2 和 T_3 分别为待识别车型的显著性特征。

由于本文中测试图像的车型是唯一的,因此,通过计算上面的置信度,可以将置信度最大的结果识别为待测车辆的类型。

4 实验结果分析

为了说明特征显著性融合算法的优越性,作者对实际采集到的312帧可见光条件下含有车辆的图像进行实验,将本文算法与单独使用 T_1 特征、 T_2 特征和 T_3 特征得到的结果进行比较,本文融合算法的权值采用式(6)的加权方法,比较结果如表1所列。

表1 不同特征的车型识别结果比较

车型特征	实验图像总数	准确识别数目	识别准确率(%)
特征 T_1	312	275	88.1
特征 T_2	312	266	85.3
特征 T_3	312	251	80.4
显著性特征融合	312	295	94.6

注:显著性特征为三级特征

从表1可以看出,用特征 T_1 识别车型的准确率要高于其它两种特征的识别准确率,而当运用3种特征的融合算法时,识别准确率大大地提高了,可达到94.6%,因此其可以用于实际工程。

结束语 本文提出了一种快速、实用的基于特征显著性的车型识别方法。该方法提出了在目标识别中多特征之间的显著性分析问题。在实际识别过程中,根据各个特征的显著性不同,赋予了不同的权值,这样加大了有效特征的作用,提高了车型识别的准确率。上述显著性特征并行融合方式对于特征提取不到的环节不敏感,系统稳定性较高。但该方式对

所有的显著性特征都有使用,没有合适的终止规则,当置信度足够高时,系统仍然继续。而串行融合算法是将第 J 级的特征观测值与第 $J-1$ 级融合中心的判断结果组合起来传送到第 J 级融合中心,其优势在于当目标的置信度大于设定的阈值时,融合过程可以自动停止,而不需要用完所有的特征,在一定程度上,可以提高系统的实时性。

参考文献

- [1] Naito T, Tsukada T, Yamada K. Robust license-plate recognition method for passing vehicles under outside environment[J]. IEEE Trans. Veh. Technol., 2000, 49(6):2309-2319
- [2] 王年,任彬,等.基于神经网络的汽车车型图象自动识别[J].中国图像图形学报,1999,4(8):668-672
- [3] 魏小鹏,于万波,金一粟.奇异值方法用于汽车车型识别[J].中国图像图形学报,2003,8(1):47-50
- [4] 靳敏.基于小波分析的车型识别[J].电子测量与仪器学报,2003,17(2):31-34
- [5] 葛广英.基于SVM的车型检测和识别算法[J].计算机工程,2007,33(6):6-8
- [6] 宋丹,徐蔚鸿.基于模糊理论的车型识别[J].计算机技术与发展,2006,16(3):47-49
- [7] Chen Z X, Wang G Y. Small targets detection under complex background based on targets feature salience[J]. International Journal of Advances in Systems Science and Applications, 2006, 9(1):95-100
- [8] Chen X. An improved branch and bound algorithm for feature selection[J]. Pattern Recognition Letters, 2003, 24(12):1925-1933
- [9] Guyon I, Elisseeff A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research, 2003, 16(3):1157-1182