

硕 士 研 究 生 读 书 报 告



题目 手势识别的机器学习算

法比较读书报告

作者姓名 谢雅婷

作者学号 21851401

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 二○一八年十月

Report of Hand Gesture Recognition：A Comparison Between Machine Learning Algorithms

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Li Qilei

By

Xie Yating

Zhejiang University, P.R. China

2018

摘要

作为基本的人机交互方式，手势识别技术正变得越来越流行，用户甚至可以以免提方式与各种电子设备交互，这在用户的手太脏，潮湿或不方便触摸设备的情况下特别有用。如何实现更高识别准确率的无触摸手势识别技术就成为了我们面临的问题。手势识别可以分为静态和动态手势识别两种，本文重点关注静态手势识别，确定手势语义的静态手势识别被认识到是一个分类问题。

本篇读书报告主要参考文献为2018年SIGGRAPH会议论文，它的主要工作是比较了决策树、KNN和SVM三种不同的机器学习算法应用于手势识别的有效性，并取最佳的算法完成了一个实时手势交互的应用。精读本篇论文，希望得到能够拓展手势识别技术实现的启迪。

**关键词**：手势识别， 机器学习，方法比较

Abstract

As a basic HCL method, gesture recognition technology is becoming more and more popular, users can even interact with various electronic devices in a hands-free manner, which is especially special when the user's hand is too dirty, wet or inconvenient to touch the device. How to achieve higher recognition accuracy of touchless gesture recognition technology has become a problem we face. Gesture recognition can be divided into static and dynamic gesture recognition. This paper focuses on static gesture recognition. Static gesture recognition to determine gesture semantics is a classification problem.

The main reference for this reading report is the 2018 SIGGRAPH conference paper, which compares the effectiveness of three different machine learning algorithms, decision tree, KNN and SVM, for gesture recognition. They have completed a real-time gesture interaction application with the best algorithm. Intensive reading of this paper, I hope to be able to expand the realization of gesture recognition technology.

**Keywords**: gesture recognition, machine learning, method comparison

1背景与问题描述

作为基本的人机交互方式，手势识别技术正变得越来越流行，用户甚至可以以免提方式与各种电子设备交互，这在用户的手太脏，潮湿或不方便触摸设备的情况下特别有用。如何实现更高识别准确率的无触摸手势识别技术就成为了我们面临的问题。手势识别可以分为静态和动态手势识别两种，本文重点关注静态手势识别，确定手势语义的静态手势识别被认识到是一个分类问题。因此，随着计算机视觉和机器学习算法的发展，有望将机器学习应用到手势识别上，以此获得比传统识别算法更高的命中率，并将手势识别技术应用于VR游戏等更多的领域。

本篇读书报告主要参考文献为2018年SIGGRAPH会议论文，它的主要工作是比较了决策树、KNN和SVM三种不同的机器学习算法应用于手势识别的有效性，并取最佳的算法完成了一个实时手势交互的应用。本篇最终解决的问题是，利用决策树算法实现了97%命中率的特定手势识别。

**2 算法简介**

**评估方法** 为了比较三种机器学习算法应用于手势识别的优劣，首先要确定一个评估方法，这项工作中评估方法采用的是“留出法”。直接将数据集D划分为两个互斥的集合，其中一个集合作为训练集S（30%的数据集D），另一个作为测试集T（剩下70%的数据集D），在S上训练出模型后，用T来评估其测试误差，作为对泛化误差的估计。

**数据提取** 为了得到有关特定语义手势的训练数据，使用了Unity中的Leap Motion API，可以获得关于手指的许多不同变量组合。在这项工作中，选择的主要变量是5个手指尖端的归一化空间位置和相邻手指之间的4个角度的组合，因此训练算法的输入数据为[(p1,p2,p3,p4,p5),(θ1，θ2，θ3，θ4)，y]，其中y为手势类型。同时，本项工作中，测试了当使用超过8400个样本（每个手势1200个）时开始发生过拟合，因此随后使用了每个类型手势1200个学习样本的文件计算不同分类算法的手势预测的准确性。

**机器学习算法** 为了完成特定语义的静态手势识别，这里规定了7个手势类型，因此这是一个有监督分类问题。决策树、KNN和SVM三种不同的机器学习算法都可以用于多类分类中，本文没有罗列具体算法，本文使用scikit-learn和python训练数据。scikit-learn是一个使用python 编程语言的免费软件机器学习库，是用于数据挖掘和数据分析的简单而高效的工具，建立在 numpy、scipy和 matplotlib 上。它具有各种分类、回归和聚类算法, 包括支持向量机、随机森林、梯度提升、决策树、最邻近分类等。对于阅读文献中涉及的三种机器学习方法，进行了扩展阅读和补充，现在简要介绍如下：

决策树：决策树是一个预测模型；他代表的是对象属性与对象值之间的一种映射关系。树中每个节点表示某个对象，而每个分叉路径则代表某个可能的属性值，而每个叶节点则对应从根节点到该叶节点所经历的路径所表示的对象的值。一般的，一棵决策树包含一个根结点、若干个内部结点和若干个叶结点; 叶结点对应于决策结果，其他每个结点则对应于一个属性测试；每个结点包含的样本集合根据属性测试的结果被划分到子结点中；根结点包含样本全集.从根结点到每个叶结点的路径对应了一个判定测试序列。决策树学习的目的是为了产生一棵泛化能力强，即处理未见示例能力强的决策树，其基本流程遵循简单且直观的"分而治之"策略。决策树的生成是一个递归过程。

支持向量机（SVM）：支持向量机在高维或无限维空间中构造超平面或超平面集合，其可以用于分类、回归或其他任务。直观来说，分类边界距离最近的训练数据点越远越好，因为这样可以缩小分类器的泛化误差。假设某些给定的数据点各自属于两个类之一，而目标是确定新数据点将在哪个类中。对于支持向量机来说，数据点被视为 p 维向量，而我们想知道是否可以用 (p-1) 维超平面来分开这些点。这就是所谓的线性分类器。可能有许多超平面可以把数据分类。最佳超平面的一个合理选择是以最大间隔把两个类分开的超平面。因此，我们要选择能够让到每边最近的数据点的距离最大化的超平面。如果存在这样的超平面，则称为最大间隔超平面，而其定义的线性分类器被称为最大间隔分类器，或者叫做最佳稳定性感知器。值得注意的是，SVM仅直接适用于两类任务。因此，必须应用将多类任务减少到几个二元问题的算法

K近邻（KNN）：k 近邻 (k-Nearest Neighbor，简称 kNN)学习是一种常用的监督学习方法，其工作机制非常简单: 给定测试样本，基于某种距离度量找出训练集中与其最靠近的 k 个训练样本，然后基于这 k 个"邻居"的信息来进行预测。通常，在分类任务中可使用"投票法" 即选择这 k 个样本中出现最多的类别标记作为预测结果，本次手势识别即选用投票法；在回归任务中时使用"平均法" ，即将这 k 个样本的实值输出标记的平均值作为预测结果;还可基于距离远近进行加权平均或加权投票，距离越近的样本权重越大。与前面两种机器学习方法相比，它是"懒惰学习" (lazy learning)的著名代表，此类学习技术在训练阶段仅仅是把样本保存起来，训练时间开销为零，待收到测试样本后再进行处理;相应的，那些在训练阶段就对样本进行学习处理的方法，称为"急切学习" (eager learning)。

**3 思考与拓展建议**

阅读该文献受益颇多。首先，最重要的一点是，确定语义下的手势识别被认识到是一个分类问题，而只要能给出特征足够的数据，机器学习能够很好地完成分类任务并实现预测。其次，本文全文展示了一个非常完整的机器学习算法比较的实例，从确定评估方法，到可视化比较结果。最后，本文利用socket通信实现了实时的静态手势识别。

基于本文的阅读有三点拓展性的思考。一是继续本文基于leap motion的思路的动态手势识别的研究，这一点在原文的未来研究展望中也有提及。另一点，更为引人注目的是，机器学习在基于声学的动态手势识别中的应用。声波手势识别的好处也在于摆脱了基于视觉感知的光照条件的约束。而现有声波手势识别系统，要么对设备需求极高，要么仍需要用户手持设备。并且现有声波手势识别主要基于TOF（飞行时间）测距或者多普勒频移原理，仅能够实现厘米级经度。通过本文的启发，未来的研究也许可以基于声波数据，比较各种机器学习算法进行手势分类的优劣，然后实现声波手势识别的精度的提高，以此来克服基于视觉感知的手势识别面临的困难。同时，声音信号的特征又不同于该文献中的输入，如何找到适合声音信号数据分类的算法将是我们的问题与挑战。

**4 附录：文献翻译**

基于leap motion的手势识别：机器学习算法之间的比较

**摘要**

在本文中，我们比较了各种机器学习算法用于实时手势识别的有效性，以便找到最佳的识别静态手势的方法，以及在算法训练期间使用的最佳样本量。

在我们的框架中，Leap Motion和Unity用于提取数据。 然后使用Python和scikit-learn训练数据。 利用关于手和手的标准化信息，我们使用决策树分类器获得了97％的命中率。

**CCS 概念**

·以人为本的计算→手势输入

·计算方法→分类监督学习

**关键词**

虚拟现实，leap motion，动作捕捉，机器学习

**1 介绍**

人类以各种方式与机器互动。为此，人类已经开发出许多形式的HCI（人机交互）。虽然鼠标和键盘的使用很普遍，但新的HCI方法也在发展。一个例子是通过手势识别进行人机交互，手势识别是一个由于VR（虚拟现实）技术的发展而在HCI领域受到极大关注的主题，也是一种控制机器人的方法。

手势是一种非言语表达形式。它包括手，脸和身体的其他部位。手势可以分为两类：静态和动态手势。本文将重点关注静态手势。

根据输入数据的类型，手势识别也可以分为两类：基于外观和基于3D模型的算法。基于外观的算法使用从输入图像的轮廓或轮廓获取的数据，同时基于3D模型的算法使用体积或骨架数据，或甚至两者的组合。

在这项工作中，我们提出了三种不同机器学习算法的比较，包括决策树，KNN（K-最近邻分类算法）和SVM（支持向量机）。

**2 框架**

为了训练机器学习算法，有必要获取关于手指的数据。为此，选择使用Leap Motion记录一系列手势，为了简化录制过程，使用了Unity中的Leap Motion API（应用程序编程接口）。可以获得关于手指的许多不同变量组合。在这项工作中，选择的主要变量是5个手指尖端的归一化空间位置和相邻手指之间的4个角度的组合。

为了生成必要的数据，有必要为每个机器学习类做至少一个记录会话，每个记录会话代表不同的手势。对于每个记录会话，在以不同方式定位手以表示相同手势的同时，Unity中的脚本将基于Leap Motion的必要数据以及类名记录到.csv文件。包含数据和类名的行每0.05秒记录一次，根据需要在总记录时间上变化。

在这个.csv文件中，30％行用于训练，而剩余的70％行用于评估准确性。然后，Python算法返回正确预测的百分比。

在这项工作中，使用了七个类：“Open”, “Close”, “Thumb”,“Two”, “Four”, “Cool”, “Love”，每个类别代表一个手势。.csv文件被分析，每个包含比例相同的7个手势。当使用超过8400个样本（每个手势1200个）时开始发生过拟合，因此随后使用了特定的文件。图2显示了使用每个类型手势1200个学习样本的文件的情况下不同分类算法的手势预测的准确性。

总的来说，决策树分类给出了更好的结果，因此，它被用于进一步的测试。 值得注意的是，最糟糕的结果是“Close”类的结果，命中率为97％，并且“Four”类能够100％次预测正确。

**3 SOCKET通信和UNITY实时测试**

基于上述结果，决策树分类被用于离线基础分类和实时样本输入训练。通过离线训练，可以进行实时预测。

这是通过运行在在主机上的一个循环来完成的，这个循环由Unity和Python创建的socket之间形成，如图3所示。因此，当前状态在Unity的界面中每0.02秒更新一次，然后Unity通过 socket，发送一个向量到scikit-learn（Python）中的决策树分类器。通过socket接收数据后，scikit-learn运行predict()函数（使用决策树分类）并返回结果。例如：“Open”。 最后，Unity从socket接收结果并根据需要创建效果。

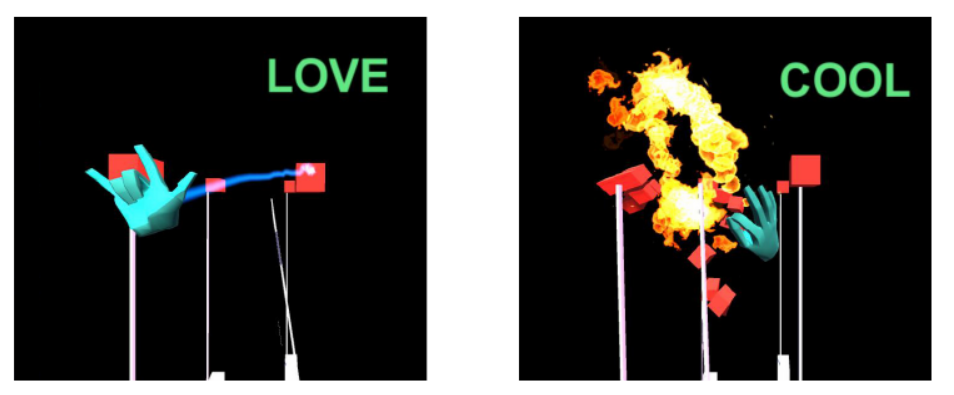
在图1所示的示例中，Unity接收预测结果并相应地显示特殊效果。 如果socket回复是“Cool”，它会产生一个火焰魔法，而如果它是“Love”，它会产生一个闪电魔法。

**4结论和未来的工作**

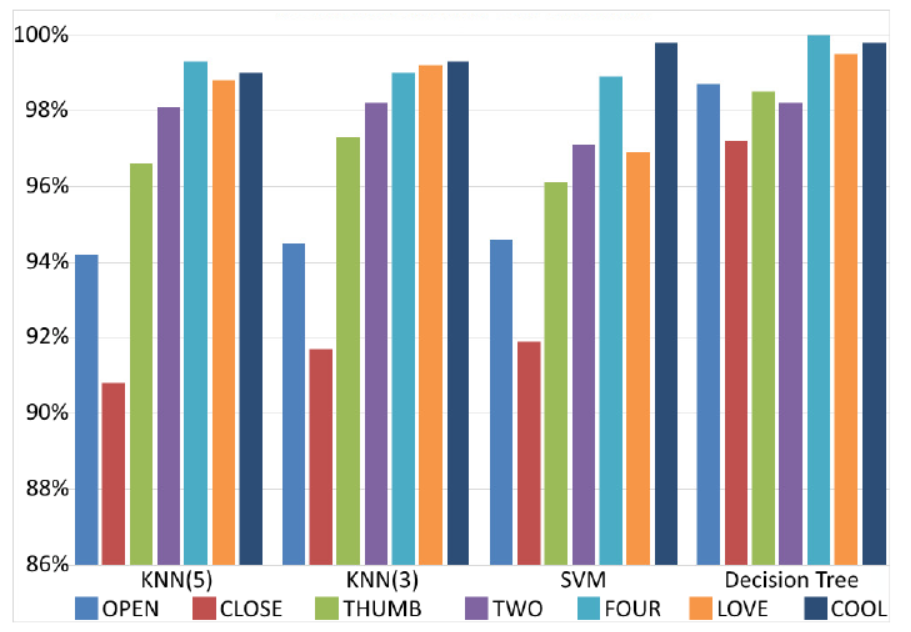
在本文中，我们每类使用了1200个样本，并对手指和它们之间的角度的数据进行了归一化。有了这个，我们使用决策树分类器实现了超过97％的命中率。

我们的研究指出，这种与Unity引擎和Leap Motion API结合使用的应用程序可用于任意手势的实时应用程序。

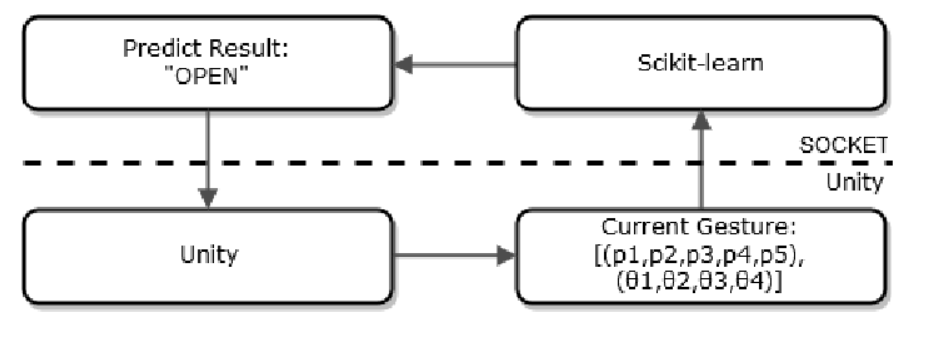
考虑到所有因素，进一步研究动态手势如何通过规范化表现将会很有趣，因为动态和静态手势都可以用于各种应用，例如VR游戏。



图一 Unity中的手势识别



图二 7种手势的机器学习算法比较



图三 Python与Unity间的数据循环

参考文献

[1] Ivo Aluízio Stinghen Filho , Estevam Nicolas Chen , Jucimar Maia da Silva Junior, Ricardo da Silva Barboza，Gesture Recognition Using Leap Motion: A Comparison Between Machine Learning Algorithms [J]: SIGGRAPH, SIGGRAPH '18 ACM SIGGRAPH 2018 Posters Article No. 65, Vancouver, British Columbia, Canada — August 12 - 16, 2018

[2] 周志华，机器学习[M]，中国，清华大学出版社，2016.01.01

[3]Wikipedia contributors. "SVM" Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 10 Dec. 2018. Web. 21 Dec. 2018.

[4] Wikipedia contributors. "decision tree" Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 28 Oct. 2018. Web. 21 Dec. 2018.