

机器学习 (/tags/#机器学习)

TalkingData开源智能设备情景感知框架“Myna”

Posted by 俞多 on December 7, 2016

阅读：130次

什么是情景

简单地说，就是与用户相关的信息：

什么人 + 在什么时候、地点 + 做什么 = 情景

- “什么人”指的是相对静态的用户属性，比如时尚辣妈、运动狂人、宅男等
- “什么时候、地点”就是用户所处的环境，包括时间、地点、天气、光感等
- “做什么”主要是用户的行为或状态，比如走路、跑步、休息或开车等等

针对不同的情景，用户需要的是不同的服务内容。比如保险领域中的UBI，基于手机传感器的数据，判断司机是否有急刹车、超速、快速变道等比较危险的架势行为；还有O2O领域，比较常见的就是精准推送，比如在上班的时候，推荐一杯星巴克的咖啡券，或者在外出旅游的时候，可以推荐一些景点及周边美食。

很多应用都存在对情景感知的需求，很多应用也都在做类似的尝试，相信这些应用在结合了情景感知后，不仅使核心功能变得更加深入、目标更加聚焦，也会给用户带来更加良好的体验。

在2016年5月召开的I/O开发者大会上，Google向我们介绍了一些与地理位置和情境相关的开发者功能。其中Awareness API提供了统一的情景查询和围栏接口，比如当前在什么地方，天气怎么样，正在做什么等，同时可以提供环境触发能力，比如如果气温高于43度，可以触发应用的回调。在新API的支持下，应用开发人员将能够利用起当前设备的情境信息（比如时间、地理位置以及行为）以便向用户提供动态、个性化的体验。

Google提供的情景感知功能都很不错，但是很可惜，这些功能都依赖于Google Play，而Google Play在中国无法正常使用。对于苹果来说，提供的状态检测接口也不太好用，因为所有的调用系统都会提示用户说有应用想要访问健身数据，并询问“是否同意”，这一步会导致不少的转化失败，所以现在有些公司也在发展自己的情景感知能力。同时我们也希望能为用户们提供这样的服务。我们一直在帮助客户采集、加工和分析数据，通过各种数据输入，多方数据汇聚在一起通过多种模型进行计算，得出对应的人群标签、环境属性以及行为识别结果。

TalkingData现在有12大类、800多个人群标签，包括手机环境、地理位置等通用标签、也包括领域相关的标签，比如金融、游戏、地产等。人群标签回答的就是“什么人”的问题，这是相对静态的数据，衰退周期比较长。环境属性解决的是“在什么时候、地点”的问题，我们覆盖全国80个城市4200万POI数据，可以识别手机在什么地方，比如是在星巴克还是在麦当劳。当然这些数据都是脱敏的，无法对应到个人。另外，也包含天气、温度、光感等信息，描述周围的环境。最后是行为识别，对应的是“做什么”的问题，主要是判断静止、走路、跑步、驾驶等状态。这是通过专门的情景感知的SDK实现的，通过多种算法投票来判断，包括SVM、随机森林等。

我们不仅希望可以给用户提供服务，而且希望和更多对情景感知有兴趣的人共同探索，在技术与智慧的碰撞中不断进步。所以我们开源了情景感知框架——Myna。

Myna 简介



Myna是基于智能设备的情景感知框架，目前暂时只支持Android平台。

Myna为以下两类用户提供服务

- 开发者可以直接使用Myna在Android上进行基于传感器数据的行为识别
- 一些算法研究者或数据科学家可以在Myna中添加新的识别算法和训练新的模型

Myna 和 Google Awareness API保持兼容

Google 将 Google Play Service 中和用户场景识别相关的服务和功能整合在一个统一的 API 下，为开发者从兼顾内存占用和电量消耗方面提供更高效率的方案。

我们可以通过

`com.google.android.gms.awareness.Awareness.SnapshotApi.getDetectedActivity` 方法获取最后一次获取到的用户行为。Myna 兼容 Awareness API，开发者可以在初始化的时候选择使用 Awareness API 或者 Myna 的识别算法，当 Myna 检测到当前运行的设备不支持 Google Play Service 的时候，会自动切换到 Myna 的识别算法。

开发者如何使用？

Myna 项目中包含一个测试 Demo 工程：demo-myna, 将该工程和 Myna 项目本身导入到 Android Studio 中，就可以开始调试了。

目前 Myna 可以识别下面三种行为类型：

1. On_Foot

2. In_Vehicle
3. Still

Myna中已经内置了一个训练好的模型文件，会在识别算法运行过程中加载。模型的ROC为：



如果开发者在应用中集成，只需要关注接口部分内容即可。

初始化

在应用自定义的 Application 派生类或者某个 Activity 的 onCreate 方法中调用下面的接口进行初始化：

```
@Override
public void onCreate() {
    super.onCreate();
    context = this;
    MynaApi.init(this, new MyInitCallback(), new MyCallback(), MynaApi.TALKINGDATA);
}
```

初始化的时候，需要传入一个实现了接口 MynaInitCallbacks 的类的实例作为回调，这样将可以在 Myna 初始化成功或者失败时做不同的处理。接口 MynaInitCallbacks 的定义为：

```
/**
 * Define resultCallback methods to handle different initialization results.
 */
public interface MynaInitCallback {

    /**
     * Called when Myna is successfully initialized.
     */
    void onSuccessed();

    /**
     * Called when Myna failed to initialize.
     */
    void onFailed(MynaResult error);
}
```

MynaResultCallback 用来返回识别结果：

```
public interface MynaResultCallback<R extends MynaResultInterface> {
    void onResult(@NonNull R var1);
}
```

通过下面的接口可以获取 Myna 的初始化状态：

```
/**
 * Get the status of Myna initialization
 */
public static boolean isInitialized()
```

开始和停止

初始化后，就可以调用 start 接口开始识别算法的运行并获得识别结果，也可以调用 stop 接口以停止识别算法的运行。

```
/**
 * Stop all background tasks
 */
public static void stop(){
    MynaHelper.stop();
}

/**
 * Start to recognizes
 */
public static void start(){
    MynaHelper.start();
}
```

如果希望使用 Google Awareness API 提供的实时行为识别能力，可以通过Myna调用，具体方法请参考集成文档：Myna快速集成文档

(<https://github.com/TalkingData/Myna/blob/f27f19785625b3b8d24801dec159589fd54fab02/QuickStart.md>)

数据科学家如何在Myna中添加自己的算法？

数据科学家可以根据对识别的行为对应的传感器数据的需求，订阅不同类型的传感器数据，设置采样的时间间隔和采样点的个数，具体方法可以参考上面的集成文档。

根据定制的数据集的格式与类型，实现ClassifierInterface 接口，在其中的recognize方法中实现具体的识别算法。详细步骤可以参考我们使用随机森林算法实现的RandomForestClassifier。

实现行为识别的步骤

- 确定要实现哪种行为的识别：
 - 走路、跑步、开车等
 - 根据自己的需求来确定需要的传感器数据的类型
- 数据采集：

- 可以设置采集人员的基本信息，如男、女、高、矮、胖、瘦等
- 设置采样时间及频率
- 注意采集数据的质量（对模型的准确度和泛化能力有很大影响）
- 数据清洗：
 - 清洗采集的原始数据集，去除明显的噪音数据
- 特征抽取：
 - 抽取对应行为数据的特征，如峰值、方差、平均值、频域特征等，并与行为标签进行绑定
- 训练模型：
 - 通过机器学习算法训练行为识别的模型
- 测试模型：
 - 使用新的测试数据对训练好的模型的识别准确率进行验证
 - 如果模型准确率没有达到预期，找到原因（如数据集质量低），重新调整后，再重新验证

已测试过的算法

我们已经测试过Random Forest、kNN、SVM算法在移动端上的性能与识别准确率，经过对比最后选择了其中表现最好的随机森林算法。





Q&A

关于Myna的情景感知，也有开发者们提出了相关的问题：

Q：像Myna这样一直采集传感器数据，并使用算法识别，会不会增加很多手机的耗电量？

A：针对耗电量，我们做过性能测试，目前如果一直使用Myna进行实时行为识别，根据多台手机的对比测试，每小时大概在1%左右。

Q：怎样才能保证原始数据集的可用性？

A：需要采集数据人员的准确的配合，比如在采集running标签的数据使，采集人员并没有进行跑步，就会很容易导致这次数据不可用。

Q：对于同一种行为，而手机处于不同状态时是否可以准确识别？

A：这个问题对于行为识别来说是一个难题，对模型的泛化能力要求很高，需要采集大量的数据样本训练模型。比如用户把手机放在衣服口袋、拿在手里、放在背包中等不同状态，对应的走路、跑步、开车等行为的数据都需要考虑到。

总结和展望

Myna的三个阶段目标：

1. 开发者可以使用Myna进行行为识别，并兼容Google Awareness API。
2. 处理收集的传感器数据的格式，可以让数据科学家无需关心Android平台传感器数据相关知识，就可以在Myna中添加新的算法，训练新的模型。开源训练模型的代码和数据集，并添加更多的行为的识别能力。
3. 添加更多的机器学习算法来实现行为识别，并移植Tensorflow的CNN到Android端。

目前第一阶段目标已经实现，并已经在github上开源：<https://github.com/TalkingData/Myna>
(<https://github.com/TalkingData/Myna>)。

Myna目前已经在github上开源，开发者们已经可以使用Myna进行行为识别。形象地说，Myna现在更像是一个时代的新生儿，我们希望能和广大开发者和数据科学家们一起培养Myna长大，不断的推进Myna走向目标的最终阶段，添加更多的行为种类，支持更多的算法及模型，让场景识别可以为更多的开发者服务。

•

分享到：

-
-
-

微信
微博
豆瓣

PREVIOUS

使用 TENSORFLOW 实现神经网络
(/2016/11/20/NEURAL_NETWORKS_USING_TENSORFLOW/)

NEXT

机器学习模型错误的4个原因（以及如何修复它）
(/2016/12/22/ML_MODEL1/)

FEATURED TAGS (/tags/)

调研 (/tags/#调研)

iOS (/tags/#iOS)

机器学习 (/tags/#机器学习)

技术 (/tags/#技术)

FRIENDS



(https://zhuanlan.zhihu.com/talkingdata)



(http://weibo.com/TalkingData)



(https://github.com/TalkingData)

Copyright © voyagelab 2017

Theme by voyagelab (http://leopan.cn/) |

Star 5