Machine Learning on Distributed Platforms

# Introduction

With an increasingly use of drone in different areas such as agriculture, construction and parcel delivery there is also a growing number of use cases for machine learning algorithms. Due to reduced hardware costs and advantages that come with multiple drones working together – swarms of drones where the drones communicate with each other and deploy machine learning algorithms together are going to be used more and more in the coming years. The use of swarms can prevent the unprecedented growth of data collections as well as reducing the complexity of algorithmic solutions needed for solving complex problems.

However, before implementing complex machine learning algorithms on swarms there are several challenges that must be tackled as for example parallelisation of algorithms, statistical heterogeneity or the limited performance of the single drones.

In order to build a solid foundation for the implementation of other machine learning algorithms, the Fachprojekt “Digital Design for Machine Learning” focuses on the implementation of a machine-learning-framework. This framework will provide the basic functionality for the deployment of more complex algorithms on STM32 hardware (and other hardware) that can be used for well scaling drone swarms.

# Motivation (kann auch weggelassen werden als überschrift)

(Alongside the distribution of machine learning algorithms on distributed platforms come several challenges that must be considered.

* The ML-Algorithms need to have a high degree of parallelism, so we can make use of the distributed computational power.
* The performance of the hardware and the workload on each drone needs to be considered.
* Communication between the drones is highly resource expensive and therefore needs to be reduces to an absolute minimum.
* Data collected from different drones might show statistical heterogeneity if the collection is not well implemented.

Problems that normally come with distributed machine learning are privacy and system heterogeneity, but because we assume using the same hardware for all the drones, we can ignore problems that come with system heterogeneity. Privacy concerns are negligible as well because we are going to reduce the communication to the sending results of aggregation and updated weights only (see more in chapter Schneider) and therefore no sensitive data can be derived from this data. )

For testing the framework, we will implement random forest as machine learning algorithm in order to classify the data received.

# Theory

## How we tackle the problems

## Advantages of distributed machine learning

## Topology

## Random forest

## Federated learning for future projects

## The Swarm design

The first issue for the swarm design is to choose a suitable topology. One possibility is a decentralised circle or mesh topology, the other option is a star topology with a central node. The second issue is how to create a link between all nodes of the swarm.

(Here maybe some pro and cons for the different topologies)

For our swarm we decided to use the star topology with a Bluetooth link between the master node and the drones. This allows us to use the Master-Slave pattern for communication, so we don’t need to deal with the problems of a decentralised topology and on the other hand we need a central node for our approach of distributed machine learning.

Each drone runs two tasks, one is the collection and processing of data the other one is the communication with the master. In our case the data collection is simulated and processed by in random forest instance. After that the drone waits for the incoming link from the master.

To establish a link the master can search for a drone by his Bluetooth name or connects directly by the known Bluetooth address. According too the master-slave pattern, the master controls the communication with the drones by sending them a message when they should sent their collected am processed data. Ones the master has all data from the drones he need for the machine learning application he processed them and sent his results back to all drones.

## The communication-sequence

One the master established the link to the drones he can it’s possible to send and receive data from each drone. To collect the data, we need for the machine learning algorithm the master cycle the list with all drones and collects the data from each.

For that the master send a communication start massage to the drone, the answer of the drone is the data for the aggregation. The communication sequences ended with the end of this data massage.

## Structure of the Bluetooth-massages

The established link supplies an input and an output stream. To simplify the usage of the steam we designed a structure for our massages. A massage is structured like the following pattern:

#<length of the payload>|<payload>

With this pattern, characterized by his control characters, we know when a message starts and how long it is.

## Current solutions

# Implementation

## Platform for the Master & Software for master implementation

As a platform for the master, we used desktop computer, because of some developing issues. The majority of Bluetooth modules where the integrated modules in the notebooks. We use Java 10 SDK and IntelliJ IDEA as main Java developing software on a 64-bit Windows 10.

## Platform for the Drone

As platform for testing the implementation we used the STM32F407 Discovery board. It comes with a 32-bit Arm based, single core processor, gets powered by an external power supply of 3 or 5 Volts and is compatible to a variety of different sensors such as temperature, video and audio sensors. Already onboard are an accelerometer, LEDs that we used for debugging (read more at chapter Debugging) and full support for the STM32CubeIDE that is developed by STMicroelectronics as well.

For the communication we attached a Bluetooth module via a UART Interface to the STM32F407 Discovery board. We choose the HC-05 Bluetooth-Board because it’s very common and well documented.

## Software for drone Implementation

### IDE

As mentioned, the board is we used is compatible with the STM32CubeIDE, that we used for implementing the framework as well as the machine learning algorithm. The IDE is based on Eclipse and supports all the add-ons that come with Eclipse.

The STM32CubeIDE comes with a lot of helpful features such as:

* The built in STM32CubeMX that is used for configuring the Pinout, clock of the processor and peripheral hardware.
* A code generator that makes the definition and addressing of the different Pins easier.
* Advanced debug features like CPU fault analysis, register views and real-time tracing.

For this project we used the 64-bit version for Windows 10 and used the STM32CubeProgrammer in addition for erasing the memory before debugging.

### Debugging

There are different ways for debugging the STM32 boards:

* Using the serial cable that we also use to power the STM32F407
* Using the special STLink Debugging cable
* Using the built-in LEDs
* Reading values directly out of the memory

Unfortunately, we could not get the console input and output to run, therefore we used the built-in LEDs in order to debug our code part by part. The implementation for using the LEDs is quite simple as the right declaration and initialisation is already done by the built-in code generator. We simply used the pre-defined method HAL\_GPIO\_TogglePin() in order to turn on a LED.

### Programming Languages

As programming language for the drone, we used C++ because its performance advantages and our previous knowledge. There is also the possibility of using C as well as deploying TensorFlow and, hence using python. However, when using TensorFlow we could not have made use of the features that come alongside the IDE but would have access to a lot of useful libraries for more complex machine learning algorithms.

The master runs on a laptop, so we don’t have any restrictions for language by performance and hardware. Our first tests with C and Python failed by a bunch of issues. In Python we had problems with the library [PyBluez and in C we aren´t able to use the Bluetooth module of the laptop. So we ended by using Java.](https://github.com/pybluez/pybluez)

### Implementing the drone framework and random forest

When generating a new project within the STM32CubeIDE it auto-generates the main file consisting of the main method and the initialisation of the pins used.

#### Decision tree class

In addition to the main class we created another class called decisiontree.cpp that contains the implementation of the Table, Node and DecisionTree objects.

(Code einfügen?)

The DecisionTree object is later used to build up a decision tree on each drone. When building the tree the drone uses ….. to guess ….. . (Bitte ergänzen was der da genau macht, da hast du mehr durchblick)

(Code einfügen?)

#### Main method

The main method basically consists of two main parts, the Initialisation and the while loop.

Within the initialisation part we set up empty tables that are then used to learn and save the results. Therefore, we simply initialize the table objects table, obstable (observation table) and restable (result table). Furthermore, a Boolean variable called master is declared and set. This variable determines whether a drone is only there for building a decision tree and later sending it, or if the drone is the master drone/central drone that is responsible for aggregating all the results gathered from the other drones.

When entering the while loop, first the role of the drone is checked. If the drone does not need to do the aggregation, it simply builds the decision tree by calling the corresponding method and afterwards sends the leafs of the tree that was generated.

If on the other hand the drone needs to do the aggregation, it first of all receives all the results generated by the other drones and then builds the modus (in case of random forest with modus) of all the results. The result can then be sent back to the drones (This will give future projects the possibility to send updated weights when using “reenforced – das ist falsch” machine learning)

The while loop runs indefinite and repeats the same steps again for new data provided by the sensors.

## Implementation of the master

The implementation of the master mostly can be divided in three parts the main class, the client and communication section and the machine learning part.

The master is the main class which handles the dones, each as a separate client object, collect their data and passes them to the aggregation function of the machine learning algorithm. In the initiation phase the master creates for each drone a separate object by calling the client constructor with the Bluetooth address of the drone. If the Adress is unknown its possible to use the scanner to search by the device name. init ML? . Ones the initiation is done the main loop start. In the main loop the master connects to all drones after another and collects they’re processed data for the aggregation. After the aggregation the new weights for the random forest algorithm on the drones were sent to them.

To the client and communication section belong two classes, the first is the client class. The client class contains the constructures for creating a client object by his Bluetooth adress. The class also contains the methods to init, establish and close the Bluetooth link, as well as the methods to send to - and read from the drone.

The BluetoothScanner class contains the methods to search for a drone by the Bluetooth name.

The third section is machine learning section. Add stuff

## Possible upgrades for the future

* Send buffer in drones, so they can classify multiple objects and send a stack of information to the master
* Check if its is possible to establish a broadcast for the aggregation results
* Multithread the drone (Communication and AI application)