Capstone Project Report

Comprehensive AI Architecture to Monitor and Predict Heart Rate during Physical Exercises

Website: https://jakatakz.github.io/it4983capstone/

Team: Talia Brooks, Aaron Bemis, Alex Boyett, Lauren Bailey, Jack Morris

Course: IT4983 Capstone Project

Report date: 10.23.22

# Executive summary

The purpose of the project is to obtain enough exercise information to then use learning models and artificial intelligence to determine the person doing the exercise heart rate within several minutes of a certain timestamp. If a person is halfway through an exercise, is in a certain geographical location and certain weather conditions, we should be able to determine what their heart rate will be as they proceed through the rest of the exercise. Having the recorded data for the rest of the exercise will help us determine the accuracy of our predictive modeling. This has many uses. This type of predictive modeling could help us determine if the exercise is beneficial to the user. In other words, are they taking things too easier or pushing too hard. Predictive modeling can also help identify areas where something in the exercise spiked or steeply declined outside of the prediction. This can help identify anomalies in the data or perhaps some underlying health conditions that need to be addressed.

Throughout the project we have been working with a dataset provided by Endomondo that captures the data of over a quarter of a million exercises (253,020) from 1,104 users. Our group has taken this data and used the geographical coordinates, latitude, and longitude along with the timestamps in the exercises to determine the city, state, and zip location as well as the weather conditions temperature, humidity, and precipitation during the exercise.

For Milestone 2, our goal was to construct deep learning models to predict heart rate during exercises based on the generated data from milestone 1. We accomplished this through something called supervised learning which is a branch of Machine learning and artificial intelligence. We processed split off 10,000 points of data of 500 records each which amounts to 5 million records that was split off from our original data set and processed it for our models. We then constructed models for this data and ran the model types at 5 times each to predict the average error between the predicted and true heart rates. The models we ran are called One-Dimensional Convolutional Neural Network (1DCNN), Long Short-Term Model (LSTM), Gated Recurrent Unit (GRU) and Multi-head Attention (Attention).

Tables of contents

Executive Summary ……………….. 2

Table of Contents ……………………. 3

Background ………………………………. 4

Project Outcomes …………………….. 5

Project Planning ………………………… 6

Team Reflection …………………….. 8

Appendix ………………………………….. 11

Background

**Business and project background**

This project has been designed to take data from a heart rate monitor like a Fitbit and consider the environment to give a more accurate reading of your exercise. An issue with this though is there are a lot of data points that need to be accounted for and then combined. Cleaned and run through an API to produce accurate data. The general goal here is to create a program that will run through that data and make predictive results that are accurate as possible. We set the parameters for the models, run each set of parameters 5 times and average those results. The results give the average error rate of the full run from what the model predicted versus that actual heart rate. This helps us gauge which models are the most accurate and which models are the most consistent.

**Specific project scope, objectives, and deliverables.**

To break it down the scientific scope of the project would be able to create a program that would be able to analyze your exercise goals through heart rate while taking into account everything such as weather locations temperature, humidity, etc. Our objective is to create a program and device that would be able to do all of this in real time and do it efficiently. The objectives outline what the deliverables are in the form of the parsed Endomondo data set, the weather dataset we got from Meteostat, and the geographic data set for Milestone 1. These objectives were cleaned and prepared for modeling into a combined data set and another dataset that had specifically no NaN values for any column which are deliverables for milestone 1.

Milestone 2 objectives and deliverables included constructing deep learning models and obtain the average error between predicted and true heart rates. Some of our objectives included getting a sample of 5 million records or 10,000 data points of 500 records each from our original larger data set which we accomplished. Processing the data objective was completed and the average error rate was obtained after running each of the models their allotted amount thus satisfying the Milestone 2 deliverable.

**Technical background**

In milestone one, the technical environment was mostly working with the Meteostat API, the Endomondo dataset, and Nominatim API to be able to get the required data. The Meteostat API gives access to data for climate and weather using thousands of weather stations histories (Meteostat (n.d.)). The Nominatim API gives access to geographical data such as longitude and latitudes, countrys, etc (Nominatim.org (n.d.)). We obtained server access through Professor Linh Le with Kennesaw State University’s VPN to be able to connect to a Jupyter Notebook application where we read and wrote to data files and used Pandas, Python, and Numpy to clean and prepare the data for modeling. Jupyter Notebook is a web application which allows for “creating and sharing computational documents” (Project Jupyter. (n.d.)). Pandas is a software library for Python where we were able to perform “data manipulation and analysis”in Jupyter Notebook while also using the Numpy library for mathematical functions (pandas. (n.d.)). While the server hosts the files we are working with, we also are looking into forms of version control as a backup where we are integrating the Github vcs directly with Jupyter Notebook, however in the meantime we are uploading directly to Github and using Microsoft Teams for files as well.

For Milestone 2, our technical background needed to consist of being familiar with Python, TensorFlow, and Jupyter Notebook. We spent most of the time running models through Jupyter Notebook and to understand what was going on we had to understand aspects of TensorFlow such as how blocks of layers and neurons effected the model types. There was some python needed to get the sample data and understanding of the processing and model running programs.

Project outcomes and achievements summary

**Assessment of project outcomes**

It is unknown right now how successful this project will be due to it still being in the beta stages of development. However, as we have met our Milestone 1 deliverables and objectives so far as we understand them, the project is currently in a successful state. Though the deliverable we do want to provide is a small device and program that can adjust to account for your temperature, weather, and environment when reading your heart rate and determining your exercise rate. We will be able to have a better idea of how successful the project will be once we get into the creation stage. Each week, as we progress, we learn more about the struggles the project brings, and we are aware of what will work.

Adding to this, now that we have ran our learning models for milestone 2 we can access which models had the most accurate predictions as well as which models had predictions that were the most consistent over five runs. We will use the data in milestone 3 where we will compare our four milestone 2 models to more advanced models.

**Technical summary of the solutions**

* Architecture, process, methodology, etc.

For Milestone 1, a program was created in Python to download and extract the raw data that was stored in a .gz file. Once unzipped, another Python program was written to turn the raw data into a Pandas dataframe. Once in a dataframe was established, the parts of data needed for Milestone 1 were written out to a pipe delimited csv file.

* Summary of major system analysis and design.

The first bit of analysis was done to determine how to gather the data and what would need to be put in place for us to combine the data. The initial approach was to build out a database with tables to house each component of the data. There would be a table to capture the raw data, the geographical data, and the weather data. Our sponsor let us know that the modeling jobs will work perfectly fine with a single file containing all the combined data. However, the tables were not completely scrapped. They proved to be valuable in pull the exact data need to run the geographical job and the weather job. We could extract just the fields needed to accomplish those task and have fields to help join it all together.

What we did for Milestone 2 was something called supervised learning which is a branch of Machine learning and artificial intelligence. In our case, it was predicting labels for each data instance given the input features from our data from Milestone 1. It’s basically algorithms that learn from our Milestone 1 data to predict the average error between the predicted and true heart rates (Le, L 2022).

* Summary of actual development or implementations

Once a server and Github was setup and access by the entire team was established, sharing programs, datasets, backups and various other program was much easier to do.

We divide workouts into multiple short and overlapping time windows such as every 10 seconds.

We have input data for our models that include things such as environment data which is weather, gender, and sport and workout data which is coordinate, current heart rate, speed, elapsed training time. All of this is from our milestone 1 data but processed. Our output is the heart rate in the next timing window. Environment data is constant throughout the workout and workout data varies.

When we processed our milestone 1 data, the weather data was averaged, we dropped certain data because a lot of the data from milestone 1 did not include information about them or at least not enough information to be worthy of consideration such as precipitation, city, zip, and state. We also consolidated variables such as mountain bike into just bike.

We then process the data using models that accommodate both single row and sequential data in a neural network. Our models to do this were 1D convolutional neural network, recurrent neural networks which are long short-term memory and gated recurrent unit, and multi-head attention.

* Summary of documentations, including analysis, design, code, testing plan, etc.

We took samples of data that was missing expected data to be returned from the API and tried to determine if something was missing or wrong with the API. This helped us refine a part of the timestamp range on the weather data portion. Doing so allowed us to gather information for well over 16,000 records that had we left alone would simply be missing data. There were also steps taken by the team to look over the data as it was being capture and make sure that the data was clean. This review of the data led us to discover that there was a language setting that was missing in the geographical API. We then took all the geographical data that came in with strange characters and reran that data. This resulted in us being able to grab much cleaner data for thousands of foreign cities and states.

We obtained 10,000 points of data of 500 records each or 5 million records for processing to run on our models using Python and Jupyter Notebook. We ran some tests to see if the first processing job worked on our data, but we had to get the sponsor to update the processing job to fit our models which was successful the 2nd time.

* Summary of other content and documents created by the team

Most of the documents created for this team was divided into 5 categories. The weekly reports, the timelines, the scientific data reports, data reports, and finally summery reports. The weekly reports were rough summaries of what had been accomplished this week and how the workload was shared and organized., the timelines were things like the gnat chart that organized everyone's roles and responsivities into a neat and easy to read chart. The scientific report was what's being currently written to present our report with, and it is an overall view of the entire project, workload, and roles. Data reports were all the clean and edited data we had created from the original data sets. Finally, the summary reports were just an overall look at meetings and conversations.

Project planning and management summary

**Overview**

We followed some project management practices from previous classes we have taken in Information Technology such as IT 3223 Software Acquisition and Project Management. Several things were familiar for us such as the Gantt chart, the project plan, keeping things in scope, understanding our requirements, and figuring out what the critical path is for us. We kept the line for communication open for everyone and responded quickly to each other, our sponsor, and professor to ensure we were achieving our objectives and deliverables. We could have improved at having actual in-person meetings more often, but found we were just as effective through email and text as well as none of our schedules worked well with each other. Each member of the group began to work on part of the project after our first meeting with our sponsor. We were able to help each other when it was needed by just simply communicating with one another. We did not strictly break up the project as a group and get people to work on certain things, which made it a bit confusing to figure out who was going to do what and how that was going.

**Project process**

The project process really began at the kickoff meeting where we better understood what was required and expected of us. After that, we examined the Endomondo data and relevant API’s to retrieve the data that was required to combine with the Endomondo data. After discovering the relevant API’s we started collecting and parsing data, however the next big achievement for us was getting access to a Jupyter Notebook server given to us by the Sponsor. This allowed us to contribute to each other and be able to easily access the data between each other more easily. Throughout this process was getting more familiar with the tools such as Jupyter Notebook itself, Pandas a python library, Python itself, and a little bit of the Numpy library. The most difficult challenge was gathering the weather and geographical data as there turned out to be some unique circumstances where certain data was questionable. After parsing the data and combining it all into one accessible .csv or Pandas dataframe, we have achieved a major objective and are ready to move on to modeling if our next meeting and discussion about the data itself goes well.

**Team contribution summary**

Talia Brooks

Total Workload hours: 105 hours

Milestone 1: Communication - weekly reports, research paper, research power point, emailing sponsor and professor. Recording all meetings. From there coordinate with everyone where everyone was on separate projects and keep up to date with what needed to be submitted and what needed further editing.

Milestone 2: Set up and organized all the meetings with our sponsor, analyzed and looked over models and programs. Looked at the edited and cleaned data for errors or inconsistencies. Created and building the power point as well as edited and looked over the essay. Filled in the weekly report and coordinated team meeting for presentation.

Aaron Bemis

Total Workload hours: 250 hours

Milestone 1: Attended all sponsor meetings as well as team meetings. Ensured that minutes from the sponsor meetings are recorded and shared with those not in attendance. Parsed out raw Endomondo data into line-by-line files for the step where everything gets combined. Researched and worked on pulling the raw data using the Nominatim API along with a library in Python called Geopy. I took a sampling of latitude and longitude points from each exercise and looked up location data such as city, state, and zip code. Worked with Alex to help run all his weather code on the server. Mostly monitored the job and reported to him any missing data. We worked together to address a timestamp issue and were able to obtain even more missing data after working that out. Helping with documentation, making edits, and ensuring everyone participates. Getting the new weekly report setup each week for us to fill out.

Milestone 2: Attended all the sponsor meetings. Ask questions to our sponsor on behalf of members that could not attend. Researched the models to better understand what the parameter were for and how they would affect our runs. Ran the job that processed our combined data into an object that could be used by all four of our models. Ran 20 iterations of the 1DCNN model and 20 iterations of the LSTM model. This was several days of running five versions of models at a time. Collected the output of those runs and logged it into a spreadsheet. Setting up the new weekly reports at the start of each week. Keep the Gantt chart updated and converted to PDF for our submissions. Added to our PowerPoint slides for the presentation, built upon our final report and made suggestions on areas we could improve our reports and presentation.

Alex Boyett

Total Workload hours: 85 hours

Milestone 1: Worked primarily with Meteostat API. Python documentation for the API was limited, which meant testing the code was the only way to figure out how it worked. I was able to use just longitude and latitude, along with the timestamp, to pull hourly weather data for most of our points. We did have one error with setting an end time from the initial timestamp that had to be resolved, but after that most of our data pulled. One major issue was that some areas were not covered with the APIs historical data, so we were pull NaN values. The sponsor stated these would be addressed in the modeling phase for precipitation values where no data was reported. Data pulled from Meteostat was put into a data frame, which allowed us to pull out only the data needed for our data frames.

Milestone 2: In this part I worked mainly on checking output data with personal models to verify output was okay. I attempted to increase our data set, as the sponsor wanted 10,000 entries, by correcting the last of the weather data pulled from the Meteostat API, but Jack was able to increase the set from the existing data.

Lauren Bailey

Total Workload hours: 100 hours

Milestone 1: Communicated with the other group members to make sure that we were getting our work done. Researched different database types that were used by similar technology. Added code to the database code that Aaron began and handed over to me. Cleaned up the code and worked to make it as best as possible. Contributed to the documents, reports, PowerPoint, and so on.

Milestone 2: Met with the Sponsor to figure out what we needed to do in this next milestone. Created the PowerPoint presentation and essay for this milestone. Communicated with team members to figure out who needed to do what. Filled in the weekly report each week.

Jack Morris

Total Workload hours: 170 hours

Milestone 1: Analyzed parsed raw, weather, and geographical data and combined them into several dataframes/tables to use for modeling in milestone 2. One dataframe contains all data while, if necessary, we have another that contains 1539 500 no NaN datapoints. I work on any website updates or changes made. Contribute to most project management documents, the report draft, and PowerPoint.

Milestone 2: General communication with professor, sponsor, team members. Wrote some code to get 10,000 points of data with 500 records for each point of data and then communicated with the sponsor to get a processing file that could correctly process the data for the models. Split up the modeling runs with Aaron Bemis where I ran all the model types of Attention and GRU. This took between 1 and 6 hours of each run of a model type and we ran each model type 5 times, sometimes more if there was a convergence error. Averaged the results of all the model runs and confirmed that our results were satisfactory to our sponsor, and they were ready for Milestone 3. Created 2 pages of notes for power point model slides, summary, Milestone 2 objectives, and more. Added to this document, updated our IT Capstone website.

Team reflection on project experience

**Project success factors**

The most important factors for this project to succeed are the following, communication, time management, and efficient technology skills. Starting early on each milestone has proven to be critical to get clarification from our professor and sponsor on deliverables. It has also helped our team overcome obstacles well before the expected deadline. Having enough time to buffer our workload has paid off over the course of the project.

**Team collaboration and communication experiences.**

Most of the time we collaborated through an app known as GroupMe for day-to-day updates as well as negotiations and planning. We also would use Microsoft Teams to speak to each other in person as well as upload drafts of works to be edited, cleaned, and prepared for submission. Questions and ideas are addressed quicky and this has helped us to move forward any time there has been a bottleneck with the project or finding out what is unclear to other members. We come together and make sure everyone understands all the components that are needed for submission or a particular deliverable during a milestone.

Our team has found that collaborating via Microsoft Teams has deemed the most useful to us. Teams allows us to create different Microsoft files that are shared with the entire team. We can all collaborate on the same document and at the same time. It updates and saves edits in real time allowing us all to see what has been added or changed. We meet virtually through Teams with each other and with our sponsor. For everyday chats, we use GroupMe. GroupMe is an application that offers group messaging. This allows us to chat from anywhere at any time. It is much easier to speak to each other through GroupMe as we all have different schedules.

**General collaboration experiences**

We would collaborate through Microsoft Teams. Most assignments we collaborated on were the reports as we would create skeleton structures then go back through and edit it. A good practice we learned is to be timely but patient with the person's response as we all work very different schedules. We have a hard time getting everyone to meet all together at the same time through an online platform. We, therefore, did not have many virtual meetings because everyone has different things going on with different schedules.

Overall, we have had an easy time communicating with one another. GroupMe allows us to chat whenever we need to so meet every day virtually does not have to happen or be a challenge. We can ask questions or speak about what we are doing/how we are doing working on this project.

**Meeting arrangements and experiences.**

Our meetings were often limited to either when we meet with our sponsors or a quick update and chat now and again to confirm some information that could not be confirmed through text. We would have quick conversations and got to the point, but no issues or arguments.

Every Monday we have a meeting with our Sponsor. Choosing a day and time was the easiest so we did this before we met with him for the first time. We do not typically meet virtually just as a team, as we are all taking other classes and working. We can work on the project in our own time and get it completed.

**Collaboration system use**

We used two pieces of software for communication and collaboration. The first was GroupMe and that is an app that can be downloaded that creates small group chats for all users. We used it for day-to-day communications and to get quick and confident answers from each other. The other software was Teams, this software was created by Microsoft, and it involves video calls, file sharing and editing, and group chat. Teams is where we share all the documentation, Gannt chart and any documentation to prepare for weekly and milestone turn ins. Being able to share documentation in this fashion has proven to be crucial for completing tasks. Microsoft Teams has also been helpful when organizing and conducting meetings with our sponsor. We have been able to all work on the same documents at one time if needed because we used Microsoft Teams.

**Other experiences**

Initially collaborating on the code and data was difficult. We immediately reached out to our sponsor to see about securing a server where we could install and run Python as well as host the data as we collect and combine it. Having him provide a server for us to connect resources and share each step we are working on is what helped bring us together as a team.

Using GroupMe in a manner more like texting has been a tremendous help. It ensures questions get answered and ideas are shared. There are times when one of us will toss out an idea and move forward once a few members agree. Some of this discussion has led to coding improvement or prompted us to reach out to our professor or sponsor for clarification.

**Challenges**

The biggest challenge was figuring out how to acquire most of the weather and geographical data. A big challenge was figuring out who goes were on the team roles. We choose a leader at random, but the title doesn’t really mean much. Talia was stationed to do weather but ended up being far better at documenting and communicating. Lauren was supposed to do more of website but ended up helping edit data. We had to figure out everyone's strengths in a short amount of time and put it all together to make a clean and quick transition.

For Milestone 2, the largest challenge was getting the time to run the models as only one person could be running 5 models at a time otherwise, we would have kernel errors or it would take substantially longer. Thus, we had to work around who would do their model runs first so the other could start.

**Areas to improve**

We will ask our project owner for feedback in our upcoming meeting where we will deliver our first sample of Milestone 1 data. We are certainly welcome to feedback from our processor as well as our sponsor and will add more to this section before our final report. We can always get more used to Python, Pandas, and Jupyter Notebook because we expect it to be very relevant for Milestone 2.

Appendix

References

Jianmo Ni, Larry Muhlstein, Julian McAuley, "Modeling heart rate and activity data for personalized fitness recommendation", in Proc. of the 2019 World Wide Web Conference (WWW'19), San Francisco, US, May. 2019.

FitRec-project. FitRec-Project. (n.d.). Retrieved September 22, 2022, from https://sites.google.com/eng.ucsd.edu/fitrec-project/home‌

“Nominatim.” Nominatim.org (n.d.). nominatim.org/.

“The Weather’s Record Keeper.” Meteostat (n.d.). https://meteostat.net/en/.

Pandas. “Python Data Analysis Library — Pandas: Python Data Analysis Library.” Pydata.org, 2018, pandas.pydata.org/.

Jupyter Project Documentation#. Jupyter Project Documentation - Jupyter Documentation 4.1.1 alpha documentation. (n.d.). Retrieved September 22, 2022, from [https://docs.jupyter.org/en/latest/](https://docs.jupyter.org/en/latest/ )

Project jupyter. Project Jupyter. (n.d.). Retrieved September 25, 2022, from <https://jupyter.org/>

Pandas. pandas. (n.d.). Retrieved September 25, 2022, from <https://pandas.pydata.org/>

Brownlee, Jason. “1D Convolutional Neural Network Models for Human Activity Recognition.” Machine Learning Mastery, 20 Sept. 2018, machinelearningmastery.com/cnn-models-for-human-activity-recognition-time-series-classification/.

Brownlee, Jason. “A Gentle Introduction to Long Short-Term Memory Networks by the Experts.” Machine Learning Mastery, 19 July 2017, machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/.

By: IBM Cloud Education. (n.d.). *What is supervised learning?* IBM. Retrieved October 21, 2022, from <https://www.ibm.com/cloud/learn/supervised-learning>

Papers with code - multi-head attention explained. Explained | Papers With Code. (n.d.). Retrieved October 23, 2022, from <https://paperswithcode.com/method/multi-head-attention>

Kostadinov, S. (2019, November 10). Understanding GRU networks. Medium. Retrieved October 23, 2022, from <https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be>

Le, L. (2022, September 27). Modeling Instructions. Linh Le.