CS CAPSTONE REQUIREMENTS DOCUMENT

MAY 30, 2020

REDUCING PATIENT DOSE FROM DIAGNOSTIC IMAGING USING MACHINE LEARNING

PREPARED FOR

OREGON STATE UNIVERSITY

DR. STEVE REESE _______________ Signature _______ Date

PREPARED BY

RADIOLOGICAL COUNTING

IAN BROWN _______________ Signature _______ Date

SEAN GILLEN _______________ Signature _______ Date

YIHONG LIU _______________ Signature _______ Date

Abstract

The OSU capstone team will demonstrate a proof-of-concept reduction in spectroscopy classification time through the use of machine learning, following an algorithm developed collaboratively at OSU and Georgetown University. To do so, the team will use radiation counting data collected by an analogous detector set up at the Oregon State University Training, Research, Isotopes, General Atomics (TRIGA) reactor. Three different machine learning models will be trained and tested with this data to utilize their respective strengths. This document is covered under a Non-Disclosure Agreement (NDA) limiting access to its signers, the project's stakeholders, and OSU employees (including Teaching Assistants).

CONTENTS

1 INTRODUCTION

1.1 Purpose

The software will apply one or more machine learning algorithms to incoming radiation counting data during emission of radiation. The algorithm will process this data and attempt to determine when an optimal dosage of radiation has been reached. Once this occurs, or if a predetermined maximum threshold dosage is reached, the emitter will be signalled to stop or reduce radiation levels.

1.2 Scope

The first task required for the project will be to write software that manipulates existing data into a specific format. This software may leverage previous Python code done by OSU postgraduate students to perform the conversions. Specifically, the output data will be formatted for use by the Waikato Environment for Knowledge Analysis (Weka) software package. The goal of this task will be the visualization of data and machine learning algorithm performance.

The aforementioned data will contain radiation counting information to be used by machine learning algorithms. Usually, a spectrometer displays this information in real time using a graphical layout. Human operators are then able to visually identify which wave frequencies have a high occurrence of counts. This will appear as a peak in activity. Since a program will be doing this reading instead, the data is outputted in *list mode*. Instead of being displayed as a graph, data will be written to a file that contains an entry for every count occurrence on a wave frequency. A timestamp will also be included. Converting this data to the format used by Weka will allow us to utilize the software's visualization and statistical tools for implementing and evaluating our machine learning algorithms.

1.3 Product Overview

1.3.1 Product Perspective

The machine learning algorithm and relevant data will exist as part of a larger data analyzer unit. This will in turn constitute a single piece of a larger system. The system may vary depending on use cases, but can be categorized as either *active* or *passive*. Active instances will contain a radiation emitter while passive instances will not. The analyzer will receive radiation counting data in real time and will control the emitter if one is present. A display will also be connected to the data analyzer. This will provide feedback to operators during execution.

1.3.2 Product Functions

The product will provide an interface for the system operator to add parameters. When configured, the system must be able to start receiving data and return with real-time feedback on the imaging process using trained machine-learning models. When it determines it has reached the stopping point, it will signal the imagery machine to cease radiation emission and notify the operator of completion.

1.3.3 User Characteristics

Users of this system will be primarily spectroscopy experts who are evaluating the new technique. They may or may not have software development experience.

1.3.4 Limitations

The data available for machine learning is limited by the fidelity of the radiation detector. Because of the setup and safety precautions needed to test the system, the majority of the development and evaluation process will be virtual. Further work will be needed to verify findings and hone the solution for medical use.

Also, because machine learning algorithms are inherently based on probability, the program will not be 100% accurate. The project will begin with a required minimum certainty value of 95%. This may be lowered or raised based on experimental results. Ideally, a final product will make this configurable to fit the requirements of varying use cases.

Finally, the project will be limited by the simplistic methods of determining the presence or identification of radioactive elements. Certain materials may only be identified using highly sophisticated approaches. Due to the team's lack of knowledge in this field, these approaches will be outside the scope of the project.

2 Specific Requirements

The system will process incoming radiation data using three different machine learning methods: naive Bayes, decision tree, and neural network. These methods have different performance characteristics and thus will be used as complements to arrive at a result. Also, implementing the different machine learning methods will help to figure out the pros and cons of each one, to aid arriving at a single best algorithm.

2.1 Usability Requirements

The system will be a proof of concept, so it does not need to cater as much to non-experts. It will need to be understood by persons familiar with spectroscopy. However, setting the system up should be made as simple as possible.

2.2 Code Quality

The underlying code that will be running this application must be maintainable for teams in the future to improve upon. This means that the code must be homogeneous and well documented. These qualities will allow any portion of code to be understandable just from looking at its source and to have the reason for its existence documented.

2.3 Flexibility

The system should account for future changes to the algorithms used. The use of a version control system should allow for flexibility in implementing different features, with the ability to roll back easily to previous methods if needed.

2.4 Safety

Because the system developed by this team will be a proof of concept, the main safety concern is ensuring use of standard procedures when collecting radiation data, as determined by experts at the OSU Radiation Center. The users will be trained, and need to understand how to interact with the radiation sources used, safely. In the long run, derivative/continued works will need to match or exceed the safety of current diagnostic imagery systems. In order to maintain this baseline, the system will have constraints on its analysis. If the analysis fails entirely, the system reverts back to a module implementing current radiological counting methods. If the analysis suggests continuing past a point suggested by the module implementing current counting methods, the system will require user input and only continue to a pre-configured maximum dosage.

2.5 Performance Requirements

In the long run, the system will need to support one user: the radiologist controlling the imagery process. Once the model is trained, the processing needs to be real-time in order to provide continuous feedback on the progress of the imagery. As the spectrometer produces usable data on the millisecond scale, processing time should be as close to that as possible.

3 Gantt Chart