PROJECT REQUIREMENTS SPECIFICATION

VISUAL QUESTION ANSWERING ON STATISTICAL PLOTS

UE18CS390A – Capstone Project Phase – 1

Submitted by:

Sneha Jayaraman	**PES1201802825**
Sooryanath I T	**PES1201802827**
Himanshu Jain	**PES1201802828**

Under the guidance of

Dr. Mamatha H.R.
Professor
PES University

January-May 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

TABLE OF CONTENTS

1. Introduction

The aim of the project is to build a Visual Question Answering system which accepts statistical plots along with questions on the plot with respect to the elements of the plot (such as intersection of the curves, area under the curve, median value and few other varieties of such relational queries) and provides answers to the questions posed.

The system should discover relationships between elements of a plot and provide relational reasoning to answer questions on the plot.

Given an image of a statistical plot and a corresponding question, the model must be able to generate a representation of the image, parse and understand the query, and generate a suitable reply. Therefore, it involves an understanding of image and the query language to be able to provide for visual reasoning.

1.1. Project Scope

Purpose and Benefits

Statistical charts are an intuitive and simple way to represent data. Deep Learning focuses on emulating human intelligence to develop new models that can reason figures and understand relationships that are intuitive to humans. Therefore, it is useful to build a model that can reason visual data in statistical plots. It is one step towards improvement in machine reasoning capabilities.

Visual plots are commonly found in research papers, scientific journals, business records e.t.c. Therefore, automation of plot analysis through the means of question-answering aids an individual to draw statistical inferences quickly from them.

The most important benefit is that visual question answering models on charts will help data analysts question and reason plots on a large scale, and automate the decision-making capabilities in several sectors such as the financial sector.

Goal

The system should be able to answer questions on Vertical and Horizontal Bar Graphs, Pie Charts, and Line Plots.

Limitations

The system will not be able to answer questions on the smoothness or the roughness of the plots. It can only answer relational queries – queries with respect to the other elements of the plot.

2. Product Perspective

Visual question answering specific to statistical plots, has less prevalent work done. It is one step towards the improvement of machine reasoning and pattern identifying capabilities.

2.1. Product Features

1. Input: The input to our model is an image of a statistical plot and a corresponding question on the plot.
2. Model: Our model will take in the input, process the visual image and parse the query, concatenate the inferences from both and produce an output. Thus, it combines the technicalities of image processing and query language understanding.
3. Output: The output of the model is the answer to the question.

The product consists of the model, and an interface to the model.

2.2. Operating Environment

The model will be made available as a web application; hence it does not depend on the underlying Operating system and its versions. It only requires a browser support of HTML5 and above. Use of the model can be done via the internet.

On the server side, the model can be saved and loaded when necessary. Therefore, the platform must be reliable and available.

2.3. General Constraints, Assumptions and Dependencies

The project focuses on model building and providing for accurate and reliable answers to the questions posed on the statistical charts. Hence, there is less focus on the security considerations. However, our solution will consist of an interface to the model that facilitates image upload and questions on the image.

2.4. Risks

Operational risk in terms of management and support for the product is a possible risk case.

3. Functional Requirements

Question Answering System for Charts take in statistical charts and questions related to them as input. Visual features from the charts have to be extracted and preprocessed. This can be done using techniques like image processing and Optical Character Recognition (OCR). The questions provided as an input need to be pre-processed using NLP techniques. Important details from the image and the questions need to be mapped and the corresponding answer should be predicted. Deep learning methods and architectures will be employed to predict the output.

Inputs are validated by the system to recognize only statistical charts. Any other images will be rejected by the system. The system will be trained on huge amounts of data and the results will be validated against the true answers. The parameters for the model will be tuned to provide the most optimal answer as the output.

4. External Interface Requirements

4.1. User Interfaces

The user interface is a web application which will take a chart and user-specific question. Users will be allowed to upload an image from their local drive. A text box will be provided that will accept the questions. The model will take the images and questions as the input and run in the backend. It will return the most probable answer and display it on the screen. Additional data like the prediction accuracy can also be displayed. A trained model will be deployed on the web server, and hence, the output should be produced within a few seconds.

4.2. Hardware Requirements

Deep learning tasks are compute intensive. The hardware requirements to build and train the model will require a minimum of 8GB RAM. Optionally, the model can be trained on the cloud to avoid physical hardware limitations. Once the model is trained, the testing phase can be done on any commodity hardwares. Web application development places no hard constraints on the hardware requirements.

4.3. Software Requirements

Question Answering Systems are built using the deep learning models and NLP techniques. Python (Version: 3.6 or more), NLP libraries, image processing tools and deep learning frameworks are required to build the model. Web frameworks like React and Flask are required to build and deploy the web application. The system can be built on any operating system like Windows or Ubuntu. Versioning of the model will be done using GitHub. Updated versions will be merged to the main branch and any testing and feature engineering will be done separately.

4.4. Communication Interfaces

Communication interfaces are not required to build the question answering system as the model is tested and run on the local machine.

5. Non-Functional Requirements

5.1. Performance Requirement

- **Usability** : The trained model must be available at ease to use it ,just by choosing an image input from the local drive or file-system . Similar to drag and drop or attaching files. The user must be able to navigate through the interface even with minimal exposure towards computing technologies.

- **Reliability** : As the product is developed by following practices in deep-learning , machine learning and imaging , there is no certain fixed level of reliability that can be set for the product. Reliability is not completely independent of the inputs to the model , hence reliability varies with respect to the context and type of inputs passed in.

- **Maintainability** : As the product is just a model , maintaining the model isn't a difficult task , it only requires a machine to reside on and a browser / interface to access.

- **Performance** : The model is expected to draw statistical inferences based on the question and the input passed with a good and acceptable accuracy level.

- **Robustness** : The model must be robust enough to classify any of the graph images and questions related to its scope of operation. The model must yield good performance for any input from the wide range of possible inputs pertaining to the scope of the model.

5.2. Safety Requirements

Safety requirements pertaining to this product are minimal as it isn't deployed onto a live physical environment . The decision process for users , such as passing image

inputs or attaching them to gain access to model /product and data must be based on the need-to-know principle, which is that access to covered data must be necessary only to the designers of the model and shouldn't be exposed to the end users.

5.3. Security Requirements

Data Sharing : As there are lots of graphical images to be collected , the data and statistics storage will be done to maintain the correct functioning of the model and to reconstruct what went wrong in case of any system - failures by constructing checkpoints. The datasets if artificially generated need to be secured locally and not be made available for commercial usages.

Model security : The model trained needs to be protected on a local machine or online if deployed onto the cloud . Modification of firewall rules or enabling regular scans may help in securing the build model.

6. Other Requirements

The necessary and sufficient requirements have already been covered in the preceding sections in detail.

Appendix A: Definitions, Acronyms and Abbreviations

ACRONYMS/TERMS	EXPANSIONS / DESCRIPTION
OCR	Optical Character Recognition
NLP	Natural Language Processing
GitHub	Source code versioning system

Appendix B: References

Title	Version Number	Date	Publishers	Reference
Plot QA	1.0	12/04/2020	Mitesh Khapra Nitesh Methani , Pritha Ganguly and Pratyush Kumar	[3]