Project Proposal
Team: Hex HyperCity

Mason Lien
Matthew Penne
Mrinal Rawool
Ketemwabi Yves Shamavu

Contents

Abstract

This milestone proposes three ideas that the group may use for a final project. A natural language inference project to determine the relationship between a premise and a hypothesis. The second project uses a convolutional neural network to predict damage from freezing on images of wheat crops. The third project uses a convolutional neural network to classify chest X-ray images.

Chapter 1

Milestone 1: Project Ideas

1.1 Introduction

Project ideas were inspired from online datasets sharing platforms such as Kaggle and group members personal research. The group also looked at published literature for other curated datasets, including the papers by [11, 4, 10, 8, 7].

Since the data sets mentioned by [8] and [7] required extensive module-specific CITI training and given the limited timeframe for our project, we excluded these papers from the project ideas detailed below.

1.2 Project Idea 1: Natural Language Inference using Deep Learning

Introduction
Natural Language Processing (NLP) is a domain that deals with the field of computer science and linguistics to devise ways for humans to interact with machines using human language. NLP includes some low-level tasks whose objective is to learn linguistic context. Examples of such tasks are parts-of-speech tagging, named entity recognition, information extraction, relationship extraction etc. Such tasks fall under shallow learning as the machine does not seek to understand the meaning or context of the content. On the other hand, high-level tasks such as question-answering, reading comprehension, and inference require the use of reasoning and knowledge in order to achieve the intended objective. According to Oxford dictionary, inference is defined as the process of reaching a conclusion based on evidence and reasoning. Currently, research within the realm of NLI can be broadly classified in two categories; textual inference and plausible inference[9]. Textual inference tasks is characterised by a definite, concrete hypothesis for every premise-hypothesis pair while plausible inference tasks require abductive reasoning and external knowledge. In this project, our objective is to build a deep learning model to solve a 3-way textual inference

task.

Problem Statement
The inference task can be described as semantically determining whether a **hypothesis** statement is **true**, **false**, or **undetermined** given the **premise**. A true, false, and an undetermined statement is termed as $e$ntailment, contradiction, and **neutral**, respectively.

Examples:
Premise: A man inspects the uniform of a figure in some East Asian country
Hypothesis: The man is sleeping
Label: Contradiction
Premise: A soccer game with multiple males playing.
Hypothesis: Some men are playing a sport.
Label: Entailment
Premise: A black race car starts up in front of a crowd of people.
Hypothesis: A man is driving down a lonely road.
Label: Neutral

Applications
Natural Language Inference tasks are a stepping stone towards progress in question answering or semantic search tasks. Any task that requires the model to look beyond keywords and syntax and grasp the meaning and context of text uses some kind of inference-based technique to solve the task at hand. NLI is also used as an evaluation technique for machine translation task.

Approaches and Data set
The team plans to use a benchmark data set released in 2015 called Stanford Natural Language Inference corpus, also known as SNLI [5]. This data set is a collection of 570k human-written English sentence pairs manually labeled as entailment, contradiction, and neutral. The corpus includes a train, test, and development split with the last two having close to 10K examples each [2]. Additionally, we would be able to track our model performance by comparing it with the performance metrics of other models that use SNLI. This information is made available through a leaderboard [1].

1.3 Project Idea 2: Evaluating Freeze Injury in Winter Wheat with Deep Learning

Introduction
Freeze injury (FI) is an abiotic stress that significantly impacts normal growth in plants [6]. For winter wheat grown in the Midwest, the probability of experiencing FI is high, since it is planted in the fall where normal growth consists

of germination, emergence, and tillering [3]. Low temperatures are known to kill plants by damaging the crown of the wheat plant, which is the main point of growth. The severity of the injury is influenced by both low temperatures and the duration of low temperatures. When FI occurs, stem growth can be delayed or terminated, resulting in plants tillering at different times. This response causes moderate to severe yield reductions and prolongs harvest due to fields having a non-uniform dry down [3], reducing profits and food supplies.

In regard to the spring of 2020, most of the state of Nebraska experienced significant freezing temperatures between April 10-16th, where lows were below 20 degrees Fahrenheit for an extensive period of time. At the University of Nebraska-Lincoln Havelock research farm, current growth stages were at a tillering stage for the winter research plots being conducted for agriculture studies. At this growth stage, the research plots were moderatly susceptible to FI and had increased the risk of significant injury the closer to jointing. Typical injury observed during this growth stage is leaf burning, where leaf tissue turns from a healthy green to yellow and finally brown if the plant does not recover. One of the main objectives of the UNL small-grains breeding program is to develop wheat varieties that are tolerant to these types of environmental conditions experienced in The Great Plains. Traditionally these plots are subjectively screened visually for agronomic and performance traits that dictate the advancement into the breeding program. Visual scoring is subjective and introduces inter-rater bias and not the most robust method to produce quality data. Recent advancements in high-throughput phenotyping (H-TP) has allowed the use of imagery analysis to replace traditional scoring methods with more robust and repeatable methods of accurately quantifying agronomic traits like FI.

Objectively, H-TP paired with deep learning models like convolutional neural networks (CNN) can accelerate the rating of FI severity on research plots and identify cold-tolerant varieties that would be ideal for advancement into commercialization. Developing a deep learning method to evaluate FI in winter wheat is novel approach and the datasets are rare to collect due to the randomness of when freeze events occur. There is much interest in this work from the UNL agricultural engineering department as well as the UNL small-grains breeding program to better understand FI and develop advanced models for improving breeding methodologies.

Approaches and Data set

The FI dataset consists of 4,000 individual plot images ranging from healthy to severely injured winter wheat plots with annotated severity ratings (1-9), where 1 is healthy and 9 is severely injured.

The objectives of this study are the following:

1. Train a number of differing deep learning CNN models to predict FI.
2. Compare models and evaluate which generalizes the best for FI.
3. Suggest future work for further developing this research.

1.4 Project Idea 3: Clinical Diagnosis Based on the NIH Chest X-ray Dataset

Chest X-rays are a cost-effective and frequently medical imaging examination. But, it is hard to establish a clinical diagnosis using chest X-rays compared to other expensive medical imaging techniques such as CT imaging.

The goal of this project is to achieve clinically relevant computer-aided detection and diagnosis using chest X-rays with convolutional neural networks.

The dataset includes about 112,000 X-ray images with disease labels from about 30 thousands patients. The labels were created using text extraction from radiological reports [10]. Hence, the labels are not 100 percent accurate and are suitable for weakly-supervised learning.

The X-ray images can belong to one or more of the following 15 classes: Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening, Cardiomegaly, Nodule Mass, Hernia, and No findings.

This project is interesting because several intensive and emergency care units across the United States rely on X-ray imaging diagnosis to decide on the best treatment in a time-sensitive context.

Unfortunately, besides the low accuracy of the provided true labels, only about one thousand images come with predefined bounding boxes. The detection of regions of interest will be unguided for the most part, which can increase training time and impinge on the convergence.

1.5 Conclusions

The topics are presented in the order the group is interested in them. The natural language inference project seems to be the most interesting and has a complete dataset, but the group is unsure which model will be the best fit. Evaluating freeze injury in winter wheat has be partially done by one of the group members for their research. But the dataset is not a large as the group would like. The project may be expanded to additional crops to compensate. The third project is a more classical project in convolutional neural networks. There are multiple data sets available for this project and models that can be adapted for this project.

Bibliography

[1] Natural language inference on snli. `https://paperswithcode.com/sota/natural-language-inference-on-snli`. Accessed: 2021-02-10.

[2] Snli. `https://nlp.stanford.edu/projects/snli/`. Accessed: 2021-02-10.

[3] Douglas S. Alt. Managing risks of soft red winter wheat production: Evaluation of spring freeze damage and harvest date to improve grain quality. Accessed: 2021-02-11.

[4] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2015.

[5] Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*, 2015.

[6] Abdul Khaliq Umair Ashraf Shakeel A. Anjum Shengnan men Longchang Wang Hafiz A. Hussain, Saddam Hussain. Chilling and drought stresses in crop plants. *Front. Plant Sci. 9:393. doi: 10.3389/fpls.2018.00393*, 2018.

[7] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[8] Alexey Romanov and Chaitanya Shivade. Lessons from natural language inference in the clinical domain. *arXiv preprint arXiv:1808.06752*, 2018.

[9] Shane Storks, Qiaozi Gao, and Joyce Y Chai. Recent advances in natural language inference: A survey of benchmarks, resources, and approaches. *arXiv preprint arXiv:1904.01172*, 2019.

[10] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.

[11] Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78, 2014.