1 | Table of contents

2 | Introduction

The following is a recommendation on how to use Artificial Intelligence (AI) to monitor greenhouse plants or to establish the optimal circumstances. The recommendation is made in cooperation with Sigrow, which creates sensors tailored to greenhouses. The data collected from monitoring plants can recursively be used to optimize growth. This is where AI can be of great use.

Our group of five people have been given the chance to cooperate with Sigrow and utilize their data to implement techniques that can assist the cultivation of cucumbers, tomatoes and other produce. We have been given the challenge to eventually implement an AI method to calculate the optimal circumstances for greenhouse plants.

The first part of the assignment consists of this recommendation, which aims to provide Sigrow with useful information about the use of AI for their data. In this text, the measurements of the given data will be discussed. Furthermore, multiple implementations will be stated and explained. This will be done in a somewhat ranking matter, where the pro's and con's of the methods are noted. The relevance of a certain method to the assignment given by Sigrow, will also be set out. Important to note is that the different implementations will be given in two groups. One group consists of methods which can be used to estimate the optimal features of plants after the process of cultivation. The other group consists of techniques used for creating the optimal environment for plants during cultivation. So in essence, these groups are different in the purposes they fulfil. Therefore we decided to make the distinction between these two types.
All in all, this text will discuss, in order: the measurements of greenhouse plants; the AI implementations for after cultivation; AI methods during cultivation; and a summarizing conclusion.

3 | About the measurements

Sigrow is currently collecting thirty features, including humidity, light, temperature and CO2. Literature research suggests that these features are the most important features for optimizing plant growth. The problem with the thirty features currently being collected is that there is little data available regarding the plants. In order to efficiently predict the plant growth, data regarding the plants is essential.

There are multiple features which could be measured to get more data regarding the plants. These features consist for example of the height, stem diameter, number of leaves or weight. Sigrow is currently measuring the weight of the plant only. The more features there are about the plants, the better the results will be.

To optimize the plant growth, output features are essential. If Sigrow can measure the aforementioned features, it is possible to use artificial intelligence. Artificial intelligence consists mostly of using data to make accurate predictions. This means that not only the output features are important, but the amount of data is significant as well. The more data there is, the more accurate predictions using artificial intelligence will be.

If Sigrow wants to optimize the plant growth in order to optimize production, it is also necessary to measure the production. Sigrow can use a similar approach as in the greenhouse challenge, in which the cucumber production is measured by dividing the cucumbers in three classes. The first class is A, which contains the cucumbers with a weight above 375 grams and has no defects. Class B contains the cucumbers with a weight between 300-374 grams or has some defects. At last class C contains all the cucumbers with a weight below 300 grams. Of course, in order to successfully classify the plants, each kind of plant has its own threshold values. Below is an overview of the thresholds for the cucumber distribution.

Cucumber distribution among three different classes, like with the greenhouse challenge.

Class	Weight	Defects
A	> 375 grams	None
B	300-374 grams	Some
C	< 300 grams	A lot

4 | Feature value optimization

Introduction to Time Series Analysis

Time series datasets differ from regular machine learning datasets in that they have a time component. Time series analysis is the analysis of a time series dataset in order to gain insights, such as the importance of the different parameters on the result. In Sigrows case, this could be the importance of for example CO_2, humidity or amount of light on the growth of the plants. Two suitable methods for time series analysis will be clarified below.

Support Vector Regression (SVR)

https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/

Support vector regression (SVR) is a form of regression that is concerned with fitting a line through data points by dividing the points into two sections. The line is drawn in between these points with the two groups on opposite sides. So essentially SVR decides on data points to group together and then plots a line on the imaginary border. Generally, SVR is used for categorization problems. Which seems logical with regard to the definition. A figure is shown below for clarity.

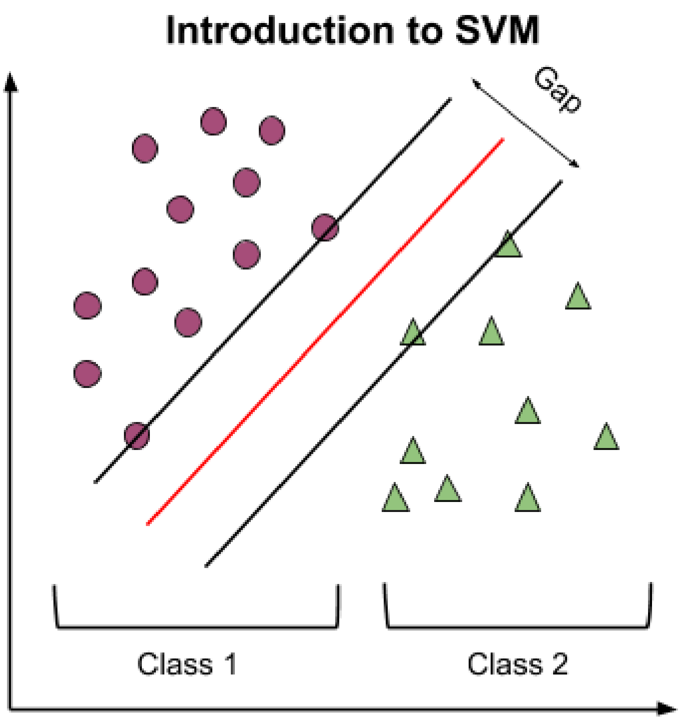


Figure 2: A simple example of a classification problem tackled with the use of Support Vector Machine (SVM).

However, the technique is also used for regression problems. This will be the case for Sigrow. So like the Polynomial Regression strategy, certain conclusions can be drawn from the final fit. Estimation of values that are not included in the dataset for instance. This is a great implementation for scarce data. See figure 2 for an example of what a SVR method of solving a regression problem would look like. It is clear that the technique is similar to the SVM method of classifying.

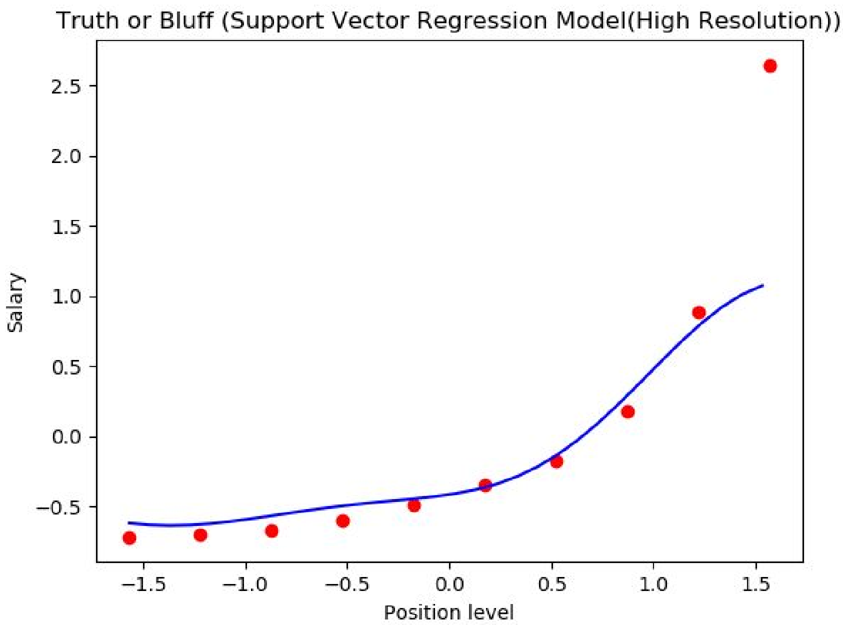


Figure 2: An example of a regression problem tackled with the use of Support Vector Regression (SVR)

Pro's:

- Relatively easy to implement
- Obtain results from little data

Con's:

- Less useful results

Relevance:

Like Polynomial Regression, this technique is relatively easy to implement. It does not require a lot of data and/or runtime to acquire results. However, the results can be unstable or less informative. A regression approach to the Sigrow Challenge is a trade-off of difficulty and quality. SVR methods can perform better with a lot of data.

Polynomial Regression

Regression uses the relationship between variables to find the best fit line that can be used to make predictions. Since plants do not grow linearly, a nonlinear model has to be fitted to the data. Otherwise the data will be under fitted and thus not plausible.

Polynomial regression is a relatively effortless method of modeling. It is a type of regression analysis, which is concerned with fitting functions on data points. For this form, the chosen function is an nth degree polynomial. So by using known data points, a polynomial function can be plotted to reveal the relations between the points. From these relations, certain conclusions can be drawn. For instance, what the optimal value for a variable/parameter is. See below for an insightful figure. Polynomial regression is a great technique to see the correlation between features and helps with the estimation of values between data points. Since the amount of data is scarce, the estimation of data is really valuable.

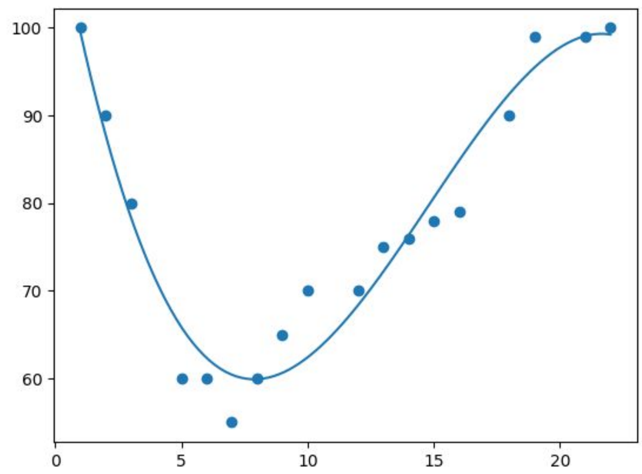


Figure 3: A graph of a fitted polynomial, showing the relations between a handful of data points.

Polynomial regression is an overall unchallenging and direct method of creating a helpful view on data. On the other hand, there are most certainly negative aspects to be noted.

Pro's:

- Relatively unchallenging implementation
- Easy to understand
- Uncomplicated way of drawing conclusions
- Does not require a lot of data
- Overall short runtime

Con’s:

- Results depend on degree choice
- Prone to underfitting/overfitting
- Prone to outliers
- Change of suboptimal results

Relevance:

Polynomial regression is a rather easy way of modeling data. Not much computing power, advanced know-how and time is needed to gather results from this implementation. For Sigrow this could be a reasonable option. The acquired datasets that we have now do not contain much information about the plants. It is difficult to implement certain techniques that require vast amounts of such data. The regression technique can acquire results with limited data. These results can be unstable, but they can be of use nevertheless. Therefore polynomial regression is a solid pick for working with such limited datasets.

5 | Explanations of AI methods

A challenge has been fulfilled by several international teams that focussed on the creation of an autonomous AI optimizer for plant growth. Five teams participated in this challenge and they all implemented different strategies. Below is an overview of these methods with an explanation. Noting these techniques is of importance to Sigrow, since the teams developed rather useful ways of growing plants optimally. The methods are shortly discussed.

Introduction to Time Series Forecasting

Time series forecasting is making predictions about the development of a time series, such as the data from Sigrows sensor in conjunction with the plant's growth. The factors that are most important for accurate time series forecasting are the amount of data, how far in time you want to predict the time series and the complexity of the problem that is modelled.

Plant growth is a very complex system since there are many factors that influence it and there are many parts that are not well understood about plant growth. Plant growth is for example influenced by the amount of water, CO_2, nutrients, light and the temperature. These factors are all captured by Sigrow's sensors, but for example the spectrum of the light or what type of nutrients are in the soil are not. More types of data about the plant's growing environment will make the model more accurate.

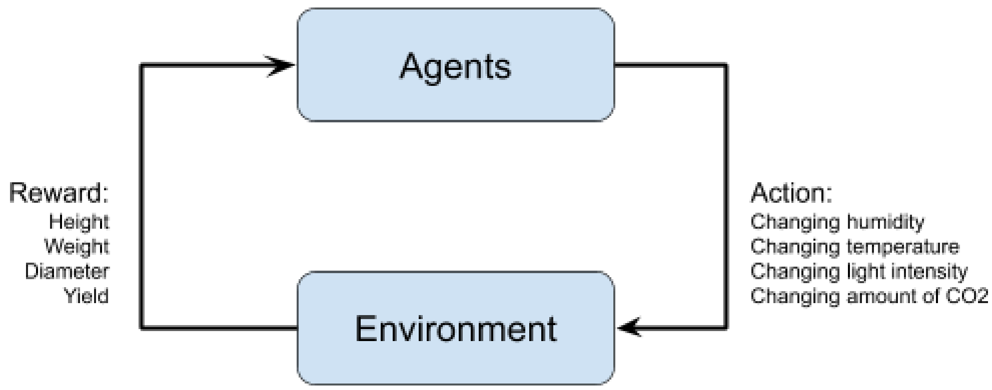
Because of the complexity of a plant's growth, a lot of data is needed to determine the effect each parameter has on the growth, even more so when the data is far away in the future, such as at the end of a plant's life cycle. The models can be updated when more data becomes available, so as to make them more accurate. Below we will exemplify AI techniques that are suitable for forecasting plant growth using Sigrow's sensor data.

Reinforcement Learning (RL)

Reinforcement learning is one the three main paradigms of machine learning, the other two being supervised and unsupervised learning. Reinforcement learning models the way humans learn from experience. Reinforcement learning problems can be seen as an agent trying to find the best action in a given state.

Inherent in this type of machine learning is that an agent is rewarded or penalized based on their actions. If an action or combination of actions leads to the target outcome, the agent is rewarded (reinforced). A real life example of this would be training a dog to sit on command. When the dog sits it gets rewarded with a treat and when it does not listen it will get punished by withholding the treat. This set-up allows the dog or agent to recognize the action which maximizes the reward obtained.

The same idea can be used by Sigrow to optimize the growth of crops. The agent will be rewarded if its actions contribute to a higher plant growth and will be punished when its actions inhibit plant growth. Other factors can be introduced as well, like penalizing the agent when its actions require a lot of resources like energy and water. There are however a few obstacles that need to be overcome. The first being the fact that any agent cannot directly influence the growth of the plants. It only has control over some properties of the environment of the crops. This creates some delay in the effects of the agents actions and the feedback it receives. This can be fixed by taking this delay into account when calculating the rewards.



The second obstacle is the amount of training data that is needed for such an algorithm. A team in the AI greenhouse challenge used artificial data to train such an algorithm because the amount of data needed is inhibitively large.

Another obstacle is the fact that in the Sigrow challenge the action space, that is the possible actions the agent can take, is continuous. This means that in any given situation there are theoretically infinite possible actions the agent can take. For example, the agent can set the amount of light, water and CO_2 received by the plants to any number. In practice the options are limited to the precision of the equipment used. In any case the amount of possible actions is still immense. Which means that it is computatively infeasible to compute the best possible action., which involves iterating through every possible action. The following technique introduces a solution to this obstacle.

Deep Deterministic Policy Gradient (DDPG)

Deep Deterministic Policy Gradient or DDPG for short is an algorithm that uses a learned policy to decide on actions in any given state. A policy maps the states enountered to policies taken in that state. In more layman terms: A policy decides what an algorithm should do at any time. This algorithm is an example of an off-policy reinforcement learning algorithm. Being an off-policy algorithm means that the algorithm learns the optimal policy concurrently but independently of the actions it should take in a state.

Normal reinforcement algorithms struggle with continuous action spaces because it is computationally impractical to calculate every possible action in a state. DDPG solves this problem by using a gradient based learning rule instead to approximate the maximum instead of exact calculations.

Pro’s:

- Works in continuous action spaces
- Can solve complex problems with enough training

Con’s:

- Large amount of training data is needed
- Like supervised learning algorithms, reinforcement learning algorithms need feedback from their actions.

Relevance:
Reinforcement learning is a very computing expensive technique since significant computing power is needed. However, the main problem for Sigrow is the amount of training data needed. Deep reinforcement learning would be an excellent technique for learning the optimal environment for plant growth, assuming the changes reported about the measurements are made. For the Sigrow challenge the agent would need to find the correct action, setting the humidity at the correct level for example, in the current state, the observed environment.

6 | Results

Since the provided data was insufficient for a lot of machine learning techniques we had to choose a method to get results with even a small amount of data. The chosen method is polynomial regression. Like mentioned in the explanation of polynomial regression, the regression method makes estimation of data possible and since the amount of data is scarce this seemed like a great method. The given maximum is/maxima give information about the overall optimal value for a feature.

Where to find and change the results

The results can be found on the web page https://readmees.github.io/polynomial_fit.html. More than 45 graphs are plotted to understand certain features. The results are made with Python in a Jupyter Notebook. Jupyter Notebook is a web application, which is a very useful tool if data needs to be visualized by Python code. We tried to make this Notebook as user friendly as possible, so if all cells are run the user will be asked for input, like the directory of the datasets and if the data needs to be divided by weekly data (to create more usable data).

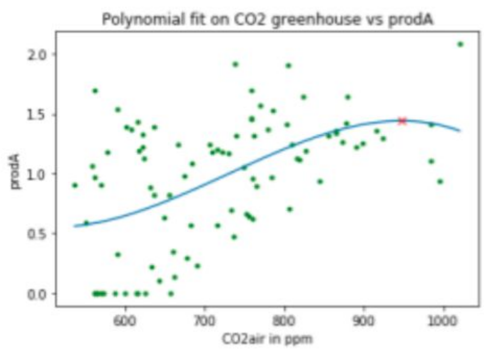
The Jupyter Notebook can be used on more data as well. So every time the growth of a plant is measured the data can simply be added in the directory where the other data is stored (for example the data of the 5 teams of the greenhouse challenge). Please make sure the directory you add with data has the same directories and features as the one in the greenhouse challenge.

Evaluating the results

Before going in depth it is important to realize that the results are found on a very scarce amount of data, so these results are not reliable, but may help getting results in the future.

Carbon Dioxide concentration

In 'The optimal CO2 concentrations for the growth of three perennial grass species' (Zeng et al., 2018) Zeng et al. examine the growth of three perennial grasses in growth chambers with different CO2 ppm values. They found an ideal total biomass value of 915ppm (see figure 2). As shown in figure 2 Zeng et al. found this maximum by fitting 2nd degree polynomials (formula: $y=ax^2+bx+c$, a-c are parameters) throughout their measurements




```
The optimal value(s) according to a polynomial
with a degree of 3: 947.475264901824.
```

Figure 1: Automonous greenhouse challenge data, devided in production per week

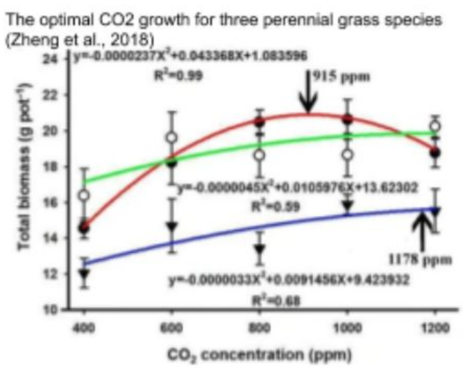


Figure 2: Zheng, Y., Li, F., Hao, L., Shedayi, A. A., Guo, L., Ma, C., ... & Xu, M. (2018). The optimal CO2 concentrations for the growth of three perennial grass species. BMC plant biology, 18(1), 27.

of 400, 600, 800, 1000 and 1200 ppm CO2. With the autonomous greenhouse data a 2nd degree polynomial is heavily underfit, if however, the data is fitted by a 3rd degree polynomial (formula: $y=ax^3+bx^2+cx+d$, a-d are parameters), an overall maximum production value (of class A) is found with a CO2 concentration of +-947 ppm (see figure 1). All teams started with a relative low concentration of CO2 and increased it while the plants grew as shown in figure 3.

Temperature

According to www.gardening.cornell.edu a cucumber crop should stay in an environment with a temperature above 70 F, so above +-22℃ at day (and above 60 F at night, +-16℃). In ‘CO2 enrichment of strawberry and cucumber plants grown in unheated greenhouses in Israel’ (Enoch, Rylski, & Spigelman) they keep the temperature in greenhouses under 28℃. The perfect temperature depends on the state of development of the plant and the

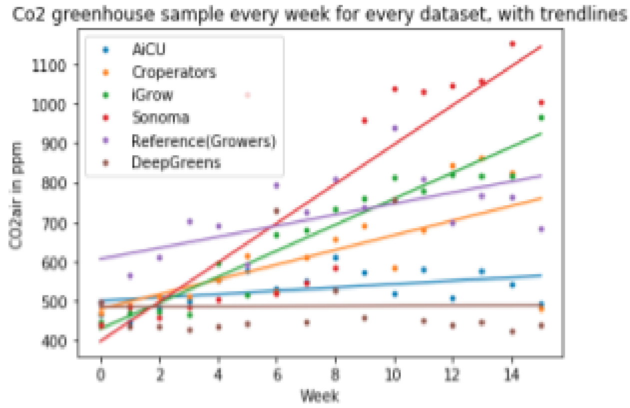
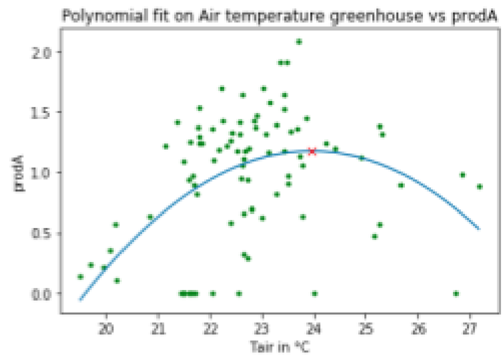


Figure 3




```
The optimal value(s) according to a polynomial
with a degree of 2: 23.961013195348002.
```

Figure 4

strain of the plant. However, it would be fair to assume based on these sources that a good temperature is between 22-28℃, which is exactly what our current model found, see figure 4. Based on the production per week and the features in that week, we found an optimal temperature of +-24℃ (see figure 4).

7 | Conclusion

In conclusion, Sigrow proposed a challenge aimed at improving crop growth. With the data gathered, AI methods can be used to achieve this challenge. However, there is a broad variety of implementations to choose from. Each with their respective positive and negative aspects. For an AI engineer, it is important to make a well-argued decision. This text serves as an informative recommendation to obtain insights from.
Many AI techniques have been discussed so that a good view on the possible approaches to the challenge is established. These techniques have been divided into two groups: methods of optimization after cultivation and methods of optimization during cultivation. The first group consisted of Support Vector Regression and Polynomial Regression. The second group consisted of Time Series Forecasting, Reinforcement Learning, Deep Deterministic Policy Gradient, Convolutional Neural Networks and Recurrent Networks. Explaining these methods has hopefully given a structured idea on what implementations are useful for what circumstances. The given information can eventually be used for creating algorithms that optimize the environmental features of greenhouse plants. Resulting in better crops and more production. A polynomial regression model has been implemented by our team. The results of this have also been discussed. It is a hopeful prospect, but more elaborated work is needed to successfully fulfil the Sigrow Challenge. This recommendation has been an attempt at aiding future work in this process.