Measuring Possible Future Selves: Using Natural Language Processing for Automated Analysis of Posts about Life Concerns

Thesis

Presented in Fulfillment of the Requirements for Graduation with Honors Research Distinction in the School of Communication in the Undergraduate School of

The Ohio State University

By

Birkan Gokbag

Undergraduate Program in Computer Science and Engineering

The Ohio State University

2020

Thesis Committee

Silvia Knobloch-Westerwick, Ph.D., Advisor

Jeremy Morris, Ph.D, Co-Advisor

Abstract

Individuals have specific perceptions regarding their lives pertaining to how well they are doing in particular life domains, what their ideas are, and what to pursue in the future. These concepts are called possible future selves (PFS), a schema that contains the ideas of people, who they currently are, and who they wish to be in the future. The goal of this research project is to create a program to capture PFS using natural language processing. This program will allow automated analysis to measure people's perceptions and goals in a particular life domain and assess their view of the importance regarding their thoughts on each part of their PFS.

The data used in this study were adopted from Kennard, Willis, Robinson, and Knobloch-Westerwick (2015) in which 214 women, aged between 21-35 years, viewed magazine portrayals of women in gender-congruent and gender-incongruent roles. The participants were prompted to write about their PFS with the questions: "Over the past 7 days, how much have you thought about your current life situation and your future? What were your thoughts? How much have you thought about your goals in life and your relationships? What were your thoughts?" The text PFS responses were then coded for mentions of different life domains and the emotions explicitly expressed from the text-data by human coders.

Combinations of machine learning techniques were utilized to show the robustness of machine learning in predicting PFS. Long Short-Term Memory networks (LSTM), Convolutional Neural Networks (CNN), and decision trees were used in the ensemble learning of the machine learning model. Two different training and evaluation methods were used to find the most optimal machine learning approach in analyzing PFS.

The machine learning approach was found successful in predicting PFS with high accuracy, labeling a person's concerns over PFS the same as human coders have done in The Allure of Aphrodite. While the models were inaccurate in spotting some measures, for example labeling a person's career concern in the present with around 60% accuracy, it was accurate finding a concern in a person's past romantic life with above 95% accuracy. Overall, the accuracy was found to be around 83% for life-domain concerns.

Vita

Education:

The Ohio State University, Columbus, Ohio USA 2016-2020

B.S. Computer Science and Engineering, Artificial Intelligence Specialization, with Honors in Engineering and Honors Research Distinction in Communication

Minor: Mathematics, with Honors

Research Experience:

Student Researcher – Biomedical Informatics, Wexner Medical Center, Ohio State University

PI: Dr. Lijun Cheng

Student Researcher – SEMI-ME Research Lab, School of Communication, Ohio State University

PI: Dr. Silvia Knobloch-Westerwick

Student Researcher – High Performance Computing and Software Lab, Ohio State University

PI: Dr. Xiaodong Zhang, worked together with PhD candidate Simon Zhang

Presentations:

Spring Denman Research Forum at Ohio State University, 2020: Measuring Possible Future Selves: Using Natural Language Processing for Automated Analysis of Posts about Life Concerns. Category: Psychology and Neuroscience

Ohio Supercomputer Center - Statewide User Groups Conference, 2019: Systematic Detection of Enzalutamide Resistance Mechanism in Castration-Resistant Prostate Cancer

Spring Undergraduate Research Festival at Ohio State University, 2019: Agenda Setting Theory Applied in a Selective-Exposure Environment on Controversial Topics

Table of Contents

List of Tables

Introduction

Literature Overview

I. Possible Future Selves Concept

Possible Future Selves (PFS) was first defined as a representation of goals in the article Possible by Markus and Nurius (Markus, & Nurius, 1986). According to Markus and Nurius (1986), “possible selves derive from representations of the self in the past and they include representations of the self in the future. They are different and separable from the current or now selves, yet are intimately connected to them. They represent specific, individually significant hopes, fears, and fantasies” (p. 954). PFS represents a schemata of what a person wishes to become or not become in the future, creating a link between cognition and motivation. In the thought process for thinking about PFS, people visualize the aspects of their lives that they wish to improve, things to watch out for in the future, and the person they wish to be. The goals can be regarded as a collection of domains that include relationships, appearance concerns, health, and other topics the person believes as important. The created PFS around the specified goals create a representation of the concerns and thoughts derived from the individual. Therefore, even though each PFS is dependent on the individual and may change given the circumstances, whether having new concerns or overcoming prior concerns, it is possible to measure PFS of the individual at that specific point in time and find the concerns the person has regarding his/her life domains. Based on psychological theorizing along with compelling empirical evidence from various cultural contexts, rendering certain PFS salient (through interventions or media messages) can increase children’s effort in school, reduce minority high-school students drop-out rate, improve health behaviors, and inspire college students to seek research careers (Schwartz, Luyckx, & Vignoles, 2011). These select examples of PFS analysis can be utilized for powerful prosocial change and intervention.

According to possible future selves, the act of thinking for the future and evaluating the present state allows the individual to create a bridge between the present and future. In this paper, the discussed data belongs to the study "The Allure of Aphrodite" by Kennard, Willis, Robinson, and Knobloch-Westerwick (Kennard, Willis, Robinson, & Knobloch-Westerwick, 2015) and consists of women aged between 21-35 who's PFS were recorded after exposure to magazine portrayals of women in gender-congruent and gender-incongruent roles. The study found that after the data collection session, the PFS of participants remained noticeable. In the context of changing to a gender-incongruent role compared to continuing a homemaker lifestyle, the participants exhibited concern over family relationships, health, and career. The exposure of homemaker roles, on the other hand, caused the participants to have concerns regarding motherhood and career roles. The difference in the concerns in these two situations might be caused by the disparity between what an individual is thinking about that given time, additionally each individuals' own train of thought will impact their own PFS. While someone with an already high paying job might not have concerns about career goals or finances, a person that was recently laid off from work might have various concerns regarding family, finances, and careers. However, this is not verified, and the cause of the concerns in association with the individual's current situation is not the focus of this paper. Instead, the capability and performance of machine learning frameworks over PFS in classification of concerns and life domains will be analyzed.

The possible concerns over the future are an indication of the disparity between the current situation of the person and the future including the possible risks that might come along with the change in time. A negative future where the person becomes fatally ill or drop out of school can also exist for a possible future an individual is envisioning. In that situation, the person would have a range of different concerns that might be realistic or not depending on the individual's train of thought and might give an indication of the variety of problems the person might be facing in the present. If these concerns are found to be mentally or physically harmful to the individual and noticed early on, it is possible for the person to receive the necessary help to obtain a better future and therefore current analysis methods of PFS need to be able to scale for high volumes of data in order to help as many individuals as possible and measure their concerns over various life domains.

Currently, there are two different methods to measure PFS; via analyzing open-ended question answers asking for person's thoughts regarding their future on specific goals and plans, and quantitative measurement of comparison across different goals on a scale of numbers given on a survey (Oyserman, & Markus, 2018). As open-ended questions allow individuals to answer in their own terms, the answers are directly related to the individuals and can supply more information regarding the person's thought process compared to the quantitative measurements. The open-ended answers are analyzed by researchers and are labeled with the topics that were discussed and the level of concern among each of them. However, as the data is analyzed by researchers reading the participants' data, the labeling processes takes time. Therefore, an automated way of analyzing PFS is required to mass-analyze data.

As a methodological advancement, this paper proposes a program to capture PFS using natural language processing (NLP). PFS measures derived from automated analysis—using long short-term memory networks, convolutional neural networks, and decision trees—are validated based on human-coded data. The created program will allow automated analysis to measure individuals' perceptions and goals in a particular life domain and assess the concerns within each part of their PFS. Uses of this newly created program include researchers labeling individuals' concerns over a selective exposure data collection session for analysis and mental health services sorting concerns of the individuals to help people better.

II. Artificial Intelligence Theory

Artificial intelligence became well-known after the Turing Test, in which a machine's intelligence is determined by the indistinguishability from a human's intelligence (Turing, 1950). Natural Language Processing, a subfield of artificial intelligence and linguistics, focuses on analyzing and interpreting human languages using machines, most commonly computers and started to emerge in 1957 by N. Chomsky with the book "Syntactic Structures" where the first computer identifiable grammar was designed (Chomsky, 1957). In machine learning, the goal is to train the machine learning model in order to classify, or analyze, data similar to the training dataset. There are various implementations of these models, this paper will specifically focus on neural networks and decision trees.

Neural networks consist of matrices where they originally start with an initial weight function and bias that determines the impact of each of the inputs that are entered into the neural network. Inspired by the biological neurons, neural networks have activation functions that simulate a decision taking place within the neurons and give a specific output based on the computation within the matrices (Hinton, Osindero, & Teh, 2006). In this paper, the activation function SoftMax, first designed by Ludwig Boltzmann (Boltzmann, 1868), will be utilized where the outputs are transformed into probabilistic distributions that sum to 1. In the beginning of training, the initial outputs of the neural networks will not give accurate results without training. In order to increase accuracy and obtain similar results to the training dataset at the training stage, the neural network is trained using optimizers, loss functions, and back-propagation (Sutskever, 2013).

Firstly, an optimizer function is chosen which contains the algorithm to maximize or minimize the loss function, used for calculating the amount of change the neural network needs to have to achieve the same result as the training sets outputs. Secondly, the loss is calculated between training sets and neural nets' outputs to find the effort the neural nets need to take to achieve the same result. Thirdly, the gradients of the neural nets are calculated for the weights of the neural nets that will allow the network to achieve similar results to the training data using the loss function. To apply the changes, the matrices backpropagate the gradients from the last layer to the first layer of the neural network and apply the weight changes based on the gradients on each layer. It is possible for the trained networks to overfit the training dataset, resulting in low accuracies throughout the testing dataset (Caruana, Lawrence, & Giles, 2000). Dropout technique is used to prevent overfitting by having a certain percentage of the locations in the neural network refrain from updating its weights based on the back propagation during training (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014)**.** The neural networks discussed in this paper utilized dropout on each of their layers.

In addition to neural networks, this paper also employs decision trees to analyze PFS. Decision trees are represented by a tree like structure, where every node is a question asked to the input data (Breiman, Friedman, Olshen, & Stone, 1983). Depending on the node condition and the input values, the tree is traversed left or right till the lowest node, a leaf node. The leaf nodes contain the conclusion reached by the decision tree. Compared to neural networks, the decision trees fit the node conditions according to information gain using the training dataset. To compare the accuracies of custom created LSTM and CNN networks with out-of-the-box decision trees, located under scikit-learn python package (Pedregosa et al., 2011), the decision tree classifiers were utilized. Ensemble learning (Dietterich, 2000) was created using the trained models where the most agreed prediction from all models was used as a result.

A subfield of NLP that focuses on understanding text-data and obtaining information is called information extraction. The current state-of-the-art information extraction methods include variations and optimizations of Convolutional Neural Networks (CNNs) and Long-Short Term Memories (LSTMs), an alteration of Recurrent Neural Networks (Lipton, 2015)**.** In this paper, both CNNs and LSTMs are utilized to show the performance of the neural networks in analyzing PFS.

CNNs utilize filters across the matrices, initially used in computer vision, and is one of the most commonly used neural network designs for natural language processing (LeCun, Haffner, Bottou, & Bengio, 1999). This method uses convolution in addition to using matrix operations. Convolution allows information to be extracted over a region that moves across the matrices of the network. The convolution is done by using a set of filters, where the matrix convolution filter is multiplied by the inner region of the neural network to transform the input matrix to a smaller feature-based convoluted matrix. This causes the input data to be shrunk from the size of filters to size of one, where it contains the most descriptive feature within the filter. From the achieve result, the matrix is pooled from to get the most prominent features within the input data.

LSTM exhibits a repeating model in its design using cycles (Hochreiter & Schmidhuber, 1997). A single block of LSTM contains several different types of gates to process the information. An input gate is used to process the input, an output gate to return the processed data via other inside gates, and a forget gate to allow the model to prioritize recent LSTM units over past LSTM units in the cycle model. In the analysis of text-data, LSTM performs with high accuracies as each of the words are encoded to work with the machine learning model and their positions are retained in the training of the neural network. This allows the model to read the sentences one word at a time in a sequence, creating a way to analyze the sentences as a whole with less focus on early words.

To analyze text-data, the data needs to be transformed to a usable form for matrix operations that represent the original data. One way is to use word embeddings, a method to map the words found in the text-data vectorized form. These vectors consist of multiple dimensions to represent a word, allowing words with far apart meanings to have very distinct vector forms compared to one another and have nearby vector representations for words with similar meanings (Zhang et al., 2016). After the text-data is changed into a matrix with vector representation of the words, the data can be used for training the neural networks.

III. Research Significance

An individual's possible future selves show a projection of that individual's current life situation and point of view. This makes it possible to extract the concerns the person has under each life domain in regard to future as compared to the present and past. If these concerns can be extracted with ease, the analysis can help motivate individuals towards their goals and keep them away from their fears. For example, the concerns regarding mental health can allow individuals to be redirected to resources that can help with their future selves. In relation to the concerns, their opinions on the importance of those specific life domains could be analyzed. The more important a life domain is to an individual, it is possible that they have concerns about that field.

Currently there is not an automated way of analyzing PFS, the current analysis methods take time as they are done by hand and therefore cannot be used for more than a handful of data. There is a need for an automated analysis of PFS to not only mass-analyze the data but also to help people's concerns in regard to their own PFS. This paper is proposing an automated way of analyzing PFS by using machine learning from computer science perspective. The proposed method will allow the mass-analysis of the data, find expressed concerns over a range of life domains, and find the relationship between the PFS and importance metrics of life goals. This program can be used by researchers to analyze created concerns of individuals over a selective exposure session in order to observe its effects to individuals' selves or by services working in the mental health sector to help promote healthy solutions to eliminate peoples' concerns by labeling the concerns.

Method

I. Data Explanation

In this research, the data was obtained from the study "The Allure of Aphrodite". There was a total of 214 participants, all women and aged between 21-35, who responded to open-ended questions about magazine portrayals of women in gender-congruent and gender-incongruent roles over the course of five days. The analyzed data was split into three: retrospective data which consists of importance metrics of life domains, text-data that contains the PFS description of the participants, and human-coded data by the researchers which contains the labels of perceived concerns from the participants' text-data. The importance metrics were on a scale of 0 to 100 with single point intervals, 0 indicated not important at all and 100 indicated very important, and the human-coded data consisted of either zeroes or ones, indicating whether a concern was expressed within the text data or not. The human-coded data was established in

Kennard et al. For the retrospective importance data, the participants responded to “Please place a mark on the line that represents how important the following items are to your happiness. This is about your personal views, there are no right or wrong answers.” For text data, the question “Over the past 7 days, how much have you thought about your current life situation and your future? What were your thoughts? How much have you thought about your goals in life and your relationships? What were your thoughts?” was answered.

II. Programming Environment

The programing language of choice was Python 3.7. PyTorch (Paszke et al., 2019) and Scikit-Learn packages were used for the machine learning frameworks. The performance of the machine learning methods was analyzed and exported to excel using the pandas package (McKinney, 2011). The part-of-speech tagger used in one part of the analysis is located under the NLTK python package (Bird, Loper, & Klein, 2009). To ensure repeatability, the data was randomized, and its randomization order along with the settings used for the parameters were saved. The source code can be found at https://github.com/BirkanGokbag/PFS.

III. Machine Learning Methods

Several different approaches were utilized to test machine learning theory for PFS analysis. Weighted average ensemble learning was utilized in predicting the human coded data and the importance metrics of the life domains. The method contained three different models and the predictions were based on each individual model’s performance, prioritizing models with higher individual performances. The models were CNNs, LSTMs, and decision trees. For CNNs and LSTMs, two different training methods and two different prediction formats were used while a single training and prediction approach was used for decision trees.

For decision trees, the Scikit-Learn library's decision tree class was used for the decision trees. The out-of-box decision trees were compared with the other neural networks, LSTMs and CNNs, where both of the neural networks were implemented on PyTorch. The two different used training methods were K-Fold Validation and iteration-based EPOCHS. In K-Fold validation approach the randomized training data was split into K equally sized sections where one section was chosen for validation and other (K-1) sections for training. The models were trained from the initial state for each of the folds, and the model with the highest accuracy among the folds was chosen for testing. For iteration-based training, the models are trained for a set number of EPOCHS, where after the model predicts the output the model back-propagates using linear algebra to reduce the error rate and then the data is tested on the validation set. The model weights that have the lowest loss, calculated by comparing the prediction to the actual set, is used for testing. All the weights, state of the training models, were saved to the computer to be used for analysis and can be loaded at any point in time. The detailed parameters table with each of the models' settings is located in Table 8 in the appendix.

The performance of training a single machine learning model to learn a single feature is compared to training a single model to learn all of the features for retrospective and human-coded datasets. The utilized evaluation methods were classification using probability and nominal prediction of the PFS. Probabilistic approach, using SoftMax function for prediction, utilized one LSTM or one CNN to analyze a feature, like a concern under a life domain or predicting an importance metric. The nominal prediction method, using feed forward network outputs, used one LSTM or one CNN network to predict all of the features in hand-coded data or the importance metrics of the participant. The nominal approach utilized stochastic gradient descent for optimizer with smooth L1 loss function, and probabilistic approach used stochastic gradient descent for optimizer with cross entropy loss function.

IV. Data Analysis

As the range of the importance metrics was between 0-100 and due to limited availability of the data, the range was lowered to 0-10, where each of the data points were mapped to the next multiple of unless it was a multiple of ten. Additionally, as the importance metrics data is subjective to the participant, it is possible that a life domain having a score of 7, 8, or 9 across three different participants have the same objective impact. Thus, if the machine learning models were close in their prediction to the original result in the performance analysis, they were counted as partially correct. The partial given score system is as follows:

- If the prediction exactly matches the correct value, 1 point.
- If the prediction is 1 off from the correct value, 0.5 points.
- If the prediction is 2 off from the correct value, 0.25 points.
- If the prediction is 3 off from the correct value, 0.13 points.
- Else, 0 points.

The accuracy of the machine learning models was compared to the human coded data analysis, which was established at Kennard et al and to the retrospective scores of the importance metrics. Due to small number of data points available for the study, Krippendorff's alpha reliability test and recall of results were found in addition to the obtained performance score of the machine learning models. Majority baseline, which assumes the majority is the answer for every feature, was utilized to assess the performance of the models from the data.

V. Study Dataset

PFS Dataset: Contains the original possible selves data from The Allure of Aphrodite with human coding of life domains and self-importance metrics of participants. The self-importance metrics were obtained prior to the study and used as a baseline, human coded data was obtained post-test by the researchers. The accuracy between determining concerns over life domains was compared to the importance metrics of the people.

VI. Additional Analysis Datasets

In addition to training the machine learning methods to analyze PFS data, the models can be trained to evaluate data that describes the person's happiness. While these additional evaluations independent of PFS are expected to result in similar or lower accuracies compared to the original dataset as the data is a representation of PFS, the representation of the current self might be possible to be extracted. The original dataset has been modified in two ways in order to test this theory, the text-data and retrospective-data obtained from the participants have been changed to include the happiness metrics that how happy the participants were. The details of the modified datasets are located under Table 9 in the appendix. The modifications are described below:

PFS Dataset with Happiness Averages: The original PFS Dataset's retrospective data was replaced by the happiness metrics of the participants which was asked in every session of the study.

PFS Dataset with Extended Text Data: The original PFS Dataset's PFS text data was combined with the open-ended happiness question which was asked in every session of the study.

In addition to the three datasets, part of speech (POS) tagger was used to filter out sections of the text-data to assess the importance of specific keywords in the portrayal of PFS compared to the performance on the original three datasets. The following POS tags were filtered using NLTK library**.**

- TO: The keyword “to”.
- POS: Possessive marker, “ ‘ “
- SYM: Symbols.
- EX: Existential “there” keyword.
- DT: Determiner keyword.

In total six datasets were used for machine learning, three different datasets with and without PFS filtering. This paper will focus on the original PFS dataset without the POS tagger, thus the analysis of the other five datasets can be found within the online code repository instead.

Results

I. Accuracies

The accuracy values for the PFS Dataset using LSTMs with different training methods are located in Table 1 for Retrospective Data and in Table 2 for human-coded data.

Table 1: LSTM Performance on Retrospective Data

	Nominal approach LSTM Trained with EPOCHS	Nominal approach LSTM Trained with KFOLDS	Probabilistic approach LSTM Trained with EPOCHS	Probabilistic approach LSTM Trained with KFOLDS
Exact Accuracy	20.8 %	6.3 %	23.8 %	12.6 %
Exact Accuracy + Partial Points	44.4 %	25.1 %	46.4 %	27.8 %

Table 2: LSTM Performance on Human-Coded Data

	Nominal approach LSTM Trained with EPOCHS	Nominal approach LSTM Trained with KFOLDS	Probabilistic approach LSTM Trained with EPOCHS	Probabilistic approach LSTM Trained with KFOLDS
Exact Accuracy	85.6 %	84.9 %	85.2 %	67.8 %

Across the LSTM models the highest accuracies were obtained by training the LSTMs using EPOCHS with the probabilistic method the neural network for the importance metrics retrospective data. However, under human-coded data the probabilistic approach had the lowest accuracy utilizing KFOLDS. The performance of CNNs under PFS Dataset with different approaches are located under Table 3 for Retrospective Data and under Table 4 for human-coded data.

Table 3: CNN Performance on Retrospective Data

	Nominal approach CNN Trained with EPOCHS	Nominal approach CNN Trained with KFOLDS	Probabilistic approach CNN Trained with EPOCHS	Probabilistic approach CNN Trained with KFOLDS
Exact Accuracy	14 %	9.7 %	24.3 %	15.1 %
Exact Accuracy + Partial Points	35.5 %	31.1 %	46.6 %	34.8 %

Table 4: CNN Performance on Human-Coded Data

	Nominal approach CNN Trained with EPOCHS	Nominal approach CNN Trained with KFOLDS	Probabilistic approach CNN Trained with EPOCHS	Probabilistic approach CNN Trained with KFOLDS
Exact Accuracy	86.5 %	85.5 %	86.4 %	69.2 %

Similar to LSTM, CNN models had the highest accuracies using probabilistic and iteration-based EPOCH training approach for the retrospective data and had comparably lower accuracies using probabilistic and KFOLD training approaches for the human-coded data. Comparison of the decision trees' performance with majority baseline is located below under Table 5 for retrospective and human-coded data.

Table 5: Decision Tree and Majority Baseline Performance on PFS Dataset

	Retrospective Data		Human-Coded Data	
	Decision Tree Classifier	Majority Baseline	Decision Tree Classifier	Majority Baseline
Exact Accuracy	18 %	27.8 %	80.6 %	85.8 %
Exact Accuracy + Partial Points	37.8 %	47.8 %	Not Applicable	Not Applicable

Decision tree classifier was able to achieve similar results to LSTM and CNN, where it had higher performance compared to their probabilistic and KFOLD training approaches, however it had lower performance compared to probabilistic and EPOCH training approaches for LSTMs and CNNs under retrospective data. Majority baseline, in comparison to decision trees and neural networks, had a higher retrospective data accuracy score. However, both majority baseline and decision trees had similar accuracies to the neural networks for human-coded data. The results of Ensemble learning for the probabilistic approach is located under Table 6 and for the nominal approach is located under Table 7.

Table 6: Ensemble Learning Performance on PFS Dataset Using Probabilistic Approach

	Retrospective Data		Human-Coded Data	
	Ensemble Learning Trained with EPOCHS	Ensemble Learning Trained with KFOLDS	Ensemble Learning Trained with EPOCHS	Ensemble Learning Trained with KFOLDS
Exact Accuracy	24.7 %	17.1 %	86.2 %	91.4 %
Exact Accuracy + Partial Points	46.6 %	36.8 %	Not Applicable	Not Applicable

Table 7: Ensemble Learning Performance on PFS Dataset Using Nominal Approach

	Retrospective Data		Human-Coded Data	
	Ensemble Learning Trained with EPOCHS	Ensemble Learning Trained with KFOLDS	Ensemble Learning Trained with EPOCHS	Ensemble Learning Trained with KFOLDS
Exact Accuracy	20.8 %	18 %	86.5 %	85.5 %
Exact Accuracy + Partial Points	44.4 %	37.8 %	Not Applicable	Not Applicable

The weighted ensemble learning was able to achieve similar accuracies with the individual machine learning models, LSTMs, CNNs, and decision trees, and was able to achieve slightly better accuracies in some cases.

II. Recall and Krippendorff's Alpha of the Machine Learning Models

The recall table for PFS Dataset using Decision Trees is located under Table 10 and Table 11 in the appendix to illustrate the data structure, the recall values for all of the datasets can be located in the code repository. While some of the features had very high recall values nearing 100 percent, some had lower values including zero percent. This is caused by the data used in this study as similar labeling of the features caused the machine learning algorithms to be correct most of the time, resulting in high recall values. On the other hand, combined with the scarcity of the data and the changing variety some of the features had very low recall. This caused Krippendorff's Alpha values to fluctuate between human-coded and retrospective data. Therefore, the reliability test scores are not reported in this paper but instead can be found under the online code repository, https://github.com/BirkanGokbag/PFS, for all of the datasets.

Discussion

The machine learning approach was found to be successful in predicting PFS with reasonable accuracy across all combinations of the training and evaluation methods, located in Table 12, the models' labeling of the concerns within the text-data was very similar to human coders across the trained features. In comparison with the original data, the created program was able to achieve a high accuracy in the human-coded data with above an 83% accuracy across all training methods. The high accuracy can be related to the way the text-data was human coded, only ranging between 0-1 to indicate whether a certain concern/emotion was present or not, and researchers only coding the explicit concerns located in the participants' text data.

The machine learning models had low accuracy across all of the training methods for the classification of the participants' importance metrics, located in Table 13. The cause of the low accuracy can be related to participants' varying opinions across life domains. Each individual has their own cognitive process and therefore it is difficult to be able to extract the level of importance each individual has to their own life domains as everyone thinks differently. Thus, the models had a very low accuracy across all training methods for labeling importance of life domains.

The recalls under human-coded data were the highest while the recalls under retrospective-data were the lowest, similar to Krippendorff's alpha reliability test scores located in the online repository. This is caused by the structure that the retrospective data was mapped to a value between 0 and 10 and was subjectively scored by the participants during the study, while the human coded data had a value of either 0 and 1 and was objectively labeled from the subjects' text data by the researchers. This difference between the data explains the changing accuracies between the models' accuracies on those two data types. When retrospective data was mapped to 0-10 from 0-100, some information was lost in the dataset and continued to have a higher range than human-coded data's 0-1 range. The data is shown to be skewed towards certain values under some of the features, lacking variety in the labels of the dataset.

In comparison between the employed methods, KFOLD training against EPOCH training and nominal approach against probabilistic approach, overall the highest performances were obtained using iteration-based EPOCH training with probabilistic evaluation approach. Both probabilistic and nominal approaches resulted in similar accuracies when utilized with iteration-based training, resulting in viable options for training. KFOLD training method is used to reduce the amount of overfit that can be caused from training the models, thus it could result in higher accuracies if applied on a larger dataset than the one used in this study. Ensemble learning model was created using the accuracy weighting distribution of the LSTMs, CNNs, and decision trees under similar categories. It was able to achieve slightly higher accuracies compared to the used machine learning models as the goal of ensemble learning was to find the most agreed result amongst the machine learning models. Thus, it was found to be the most optimal way to analyze PFS data.

Majority baseline had a high accuracy compared to the machine learning models as a baseline method, this shows the distribution of the features' labels under the PFS dataset for both retrospective and human-coded data. However, under human-coded data it had lower accuracies compared to the machine learning models as the models had difficulty learning the values within the range 0-10 due to 11 different labels under retrospective data while only learning two different labels under human-coded data. If the amount of data was increased for the retrospective data, enough to level out the distribution across the features, then majority baseline is expected to have a lower accuracy. Unlike the baseline method, the models are trained to have the text-data as inputs rather than the labels and therefore will be able to analyze the data at a higher accuracy in comparison.

After the datasets were altered using the happiness questions for both the text-data and retrospective data, including the POS filtering of the text-data, the accuracies were slightly different compared to the original dataset. However, no significant finding was discovered by changing the datasets or modifying them in any way. Similarly, after the PFS filtering was implemented to test the impact of the words the datasets did not gain or lose a substantial amount of accuracy. This can be explained from the set of the removed words under the POS filter, the removed words did not include any adjectives or nouns but the connecting keywords between words such as “there” or “to.”

Conclusion

The machine learning approach to mass-analyze high amounts of PFS data was found to be viable. Across different datasets, approaches, and methods the machine learning framework had high accuracies finding the concerns and emotions expressed within the PFS text data. When trained and compared with the human-coding of the text-data regarding concerns over life domains, the machine learning models had high accuracies in comparison to the baseline used in this study. For the retrospective data, the model was not able to achieve very high accuracies, but more data is required to verify whether the machine learning models could not determine the individuals’ importance metrics from their text-data, as it is highly possible that the participants did not talk about the life-domains they thought were highly important in addition to their own way of thinking. However, for human-coded data the model was able to get a very high accuracy of above 83% accuracy overall and was found to be viable in analyzing text-based PFS. In conclusion, the machine learning model was found to be successful in analyzing PFS across life-domains with high accuracy and can be used to mass-analyze the possible selves text data.

Future Work

As the machine learning approach was found to be viable for analyzing possible future selves, it could be used to analyze PFS across other life domains and dimensions for further research and the program could be employed by other researchers to study the behavior of self. The link between PFS and a person’s mental health could be further studied using machine learning to direct people to right resources depending on their concerns. If an abundance of PFS data is obtained, more complex neural networks could be used to increase the learning limit of the ones used in this research project and the program can be scaled to analyze thousands of PFS data.

References

Bird, S., Loper, E., & Klein, E. (2009). Natural Language Processing with Python. *O'Reilly Media Inc*.

Boltzmann, L. (1868). "Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten" [Studies on the balance of living force between moving material points]. *Wiener Berichte*, 58, 49–96

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. *CRC press*.

Caruana, R., Lawrence, S., & Giles, C. L. (2000). Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. *Advances in neural information processing systems*. 13, 402-408

Chomsky, N. (1957). Syntactic structures. *Mouton de Gruyter*.

Dietterich, T.G. (2000). Ensemble methods in machine learning. In J. Kittler & Roli, F., *Multiple classifier systems*, 1-15, doi: 10.1007/3-540-45014-9_1

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18, 1527-1554. doi: 10.1162/neco.2006.18.7.1527

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kennard, A. R., Willis, L. E., Robinson, M. J., & Knobloch-Westerwick, S. (2015). The allure of aphrodite: How gender-congruent media portrayals impact adult women's possible future selves. *Human Communication Research*, 42, 221-245. doi: 10.1111/hcre.12072

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object recognition with gradient-based learning. *Shape, Contour and Grouping in Computer Vision*, 1681, 319-345. https://doi.org/10.1007/3-540-46805-6_19

Lipton, Z. C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. Retrieved from https://arxiv.org/abs/1506.00019

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 4*, 954-969. doi: 10.1037/0003-066X.41.9.954

McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, Volume 14.

Oyserman, D., & Markus, H. (2018). Open and Close Ended Possible Selves Measures. Retrieved from https://dornsife.usc.edu/assets/sites/782/docs/Possible_Selves_Coding_Guide_--_July_2018.docx

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 8024–8035. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E.

(2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Schwartz, S. J., Luyckx, K., & Vignoles, V. L. (Eds.). (2011). *Handbook of identity theory and research*. New York, NY, US: Springer

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929-1958. doi: 10.5555/2627435.2670313

Sutskever, I. (2013). Training recurrent neural networks. [Doctoral Dissertation, University of Toronto]. doi: 10.5555/2604780

Turing, A. (1950). Computing Machinery and Intelligence. Mind, 49, 433-460. doi: 10.1093/mind/LIX.236.433

Zhang, Y., Rahman, M. M., Braylan, A., Dang, B., Chang, H., Kim, H., … Lease, M. (2016). Neural Information Retrieval: A Literature Review. Retrieved from https://arxiv.org/abs/1611.06792.

Appendix A. Tables

Table 8: Hyperparameters for CNN and LSTM Networks

Parameter	**LSTM**	**CNN**
Learning Rate for Optimizer, SGD	0.01	0.01
Neural Network Specific Param	Bidirectional	5 filters of sizes 2, 3, 5
Word Embedding Dimensions	50	50
Dropout	20 %	20 %
Value of K for KFOLD	20	20

Table 9: Data Descriptions

Descriptions	PFS Dataset	PFS Dataset with Happiness Averages	PFS Dataset with Extended Text Data
Retrospective Data	Consists of the importance metrics regarding the life domains, taken from the baseline, "Please place a mark on the line that represents how important the following items are to your happiness. This is about your personal views, there are no right or wrong answers."	Consists of the happiness metric questions for specific life domains that was asked each day of the data collection process, "For each of the following, please place a mark on the line that represents how happy you are TODAY with events and circumstances in that area of your life. This is about your personal views, there are no right or wrong answers." The happiness data was averaged.	Consists of the importance metrics regarding the life domains, taken from the baseline, "Please place a mark on the line that represents how important the following items are to your happiness. This is about your personal views, there are no right or wrong answers."
Text Data	Consists of the open-ended question "Over the past 7 days, how much have you thought about your current life situation and your future? What were your thoughts? How much have you thought about your goals in	Consists of the open-ended question "Over the past 7 days, how much have you thought about your current life situation and your future? What were your thoughts? How much have you thought about your goals in life and your relationships? What were	Consists of the merge of the open-ended questions "Over the past 7 days, how much have you thought about your current life situation and your future? What were your thoughts? How much have you thought about your goals in life and

	life and your relationships? What were your thoughts?" that was asked at the end of the data collection as posttest.	your thoughts?" that was asked at the end of the data collection as posttest.	your relationships? What were your thoughts?" and "How are you feeling TODAY about your current life situation and your future in comparison to other people? What are your thoughts on your goals in life?". The “TODAY” questions were asked in each day of the study.
Human Coded Data	Consists of the analyzed PFS in text form under several life domains by researchers. This was done after the data collection process had finished.	Consists of the analyzed PFS in text form under several life domains by researchers. This was done after the data collection process had finished.	Consists of the analyzed PFS in text form under several life domains by researchers. This was done after the data collection process had finished.

Table 10: Recall Table for Retrospective Data in PFS Dataset

Feature Description	Decision Trees	LSTM Average	CNN Average	Ensemble Learning Average
How happy you are today with...Your Health	14.3 %	24.4 %	22.6 %	22.6 %
How happy you are today with...Your Friends	31 %	19.6 %	17.3 %	31.5 %
How happy you are today with...Your Neighbors	21.4 %	7.7 %	5.4 %	10.7 %
How happy you are today with...Extend of which you help others	21.4 %	14.3 %	19.6 %	22 %
How happy you are today with...Things you do for fun	28.6 %	16.1 %	18.5 %	26.2 %
How happy you are today with...Your Romantic Life	14.3 %	10.7 %	14.3 %	16.1 %
How happy you are today with...Your Prospects of Having a Happy Marriage	9.5 %	7.1 %	12.5 %	11.9 %
How happy you are today with...Your Prospects of Having a Family with Children	16.7 %	19 %	17.9 %	20.2 %
How happy you are today with...Your Physical Attractiveness	14.3 %	14.9 %	20.2 %	17.3 %
How happy you are today with...Your Weight	16.7 %	16.1 %	11.3 %	18.5 %

How happy you are today with...Your Current career situation	16.7 %	17.9 %	11.9 %	20.8 %
How happy you are today with...Career prospects	4.8 %	7.7 %	8.9 %	8.9 %
How happy you are today with...Prestige of your current job/career status	11.9 %	13.1 %	14.3 %	11.9 %
How happy you are today with...Your Income	19 %	11.3 %	18.5 %	20.2 %
How happy you are today with...Your Finances	26.2 %	26.2 %	18.5 %	28 %
How happy you are today with...Your Achievement of personal goals	9.5 %	17.3 %	16.1 %	19.6 %
How happy you are today with...Your Life in general	28.6 %	17.3 %	18.5 %	29.8 %
How happy you are today with...Yourself in general	19 %	25 %	17.9 %	26.2 %

Table 11: Recall Table for Human Coded Data in PFS Dataset

Feature Description	Decision Trees	LSTM Average	CNN Average	Ensemble Learning Average
Romance_Past	92.9 %	95.2 %	94.6 %	95.2 %
Romance_Present	69 %	69 %	55.4 %	68.5 %
Romance_Future	57.1 %	78.6 %	64.3 %	73.2 %
Career_Past	100 %	100 %	100 %	100 %
Career_Present	52.4 %	40.5 %	60.7 %	59.5 %
Career_Future	71.4 %	58.9 %	58.9 %	76.2 %
School_Past	90.5 %	95.2 %	95.2 %	95.2 %
School_Present	69 %	66.1 %	83.3 %	79.8 %
School_Future	78.6 %	88.1 %	88.1 %	88.1 %
Children_Past	100 %	100 %	100 %	100 %
Children_Present	85.7 %	88.1 %	66.1 %	87.5 %
Children_Future	66.7 %	61.9 %	81 %	77.4 %
Appear_Past	95.2 %	100 %	100 %	100 %
Appear_Present	76.2 %	67.9 %	90.5 %	86.9 %
Appear_Future	97.6 %	97.6 %	97.6 %	97.6 %
Hopeful	73.8 %	88.1 %	66.1 %	84.5 %
Happy	61.9 %	76.2 %	76.2 %	76.2 %
Excited	85.7 %	69.6 %	92.9 %	91.1 %
Confident	95.2 %	97.6 %	97.6 %	97.6 %
Optimistic	73.8 %	62.5 %	83.3 %	81 %
Blessed	92.9 %	90.5 %	90.5 %	90.5 %
Thankful	81 %	90.5 %	90.5 %	90.5 %
Content	54.8 %	69 %	69 %	69 %
Angry	92.9 %	97.6 %	73.2 %	96.4 %
Sad	85.7 %	90.5 %	90.5 %	90.5 %

Anxious	88.1 %	90.5 %	90.5 %	90.5 %
Scared	95.2 %	97.6 %	97 %	97.6 %
Insecure	90.5 %	97.6 %	97.6 %	97.6 %
Uncertain	92.9 %	95.2 %	71.4 %	94.6 %
Frustrated	88.1 %	85.7 %	85.7 %	85.7 %
Stressed	88.1 %	92.9 %	92.3 %	92.9 %
Pessimistic	85.7 %	90.5 %	90.5 %	90.5 %
Settled	83.3 %	90.5 %	90.5 %	90.5 %
concern_romance	59.5 %	54.8 %	47 %	61.9 %
At Least 1 Concern	71.4 %	58.3 %	66.7 %	66.7 %
concern_career	64.3 %	35.7 %	54.2 %	60.1 %
concern_family	69 %	73.8 %	73.8 %	73.8 %
concern_appearance	85.7 %	90.5 %	90.5 %	90.5 %
concern_school	81 %	60.7 %	81 %	81 %

Table 12: Accuracy Comparisons of Machine Learning Models for Human Coded Data of PFS Dataset

Feature Description	LSTM Accuracy	CNN Accuracy	Decision Tree Accuracy	Ensemble Learning Accuracy
Romance_Past	95.2 %	94.6 %	95.2 %	95.2 %
Romance_Present	69.0 %	55.4 %	68.5 %	68.5 %
Romance_Future	78.6 %	64.3 %	73.2 %	73.2 %
Career_Past	100.0 %	100.0 %	100.0 %	100.0 %
Career_Present	40.5 %	60.7 %	59.5 %	59.5 %
Career_Future	58.9 %	58.9 %	76.2 %	76.2 %
School_Past	95.2 %	95.2 %	95.2 %	95.2 %
School_Present	66.1 %	83.3 %	79.8 %	79.8 %
School_Future	88.1 %	88.1 %	88.1 %	88.1 %
Children_Past	100.0 %	100.0 %	100.0 %	100.0 %
Children_Present	88.1 %	66.1 %	87.5 %	87.5 %
Children_Future	61.9 %	81.0 %	77.4 %	77.4 %
Appear_Past	100.0 %	100.0 %	100.0 %	100.0 %
Appear_Present	67.9 %	90.5 %	86.9 %	86.9 %
Appear_Future	97.6 %	97.6 %	97.6 %	97.6 %
Hopeful	88.1 %	66.1 %	84.5 %	84.5 %
Happy	76.2 %	76.2 %	76.2 %	76.2 %
Excited	69.6 %	92.9 %	91.1 %	91.1 %

Confident	97.6 %	97.6 %	97.6 %	97.6 %
Optimistic	62.5 %	83.3 %	81.0 %	81.0 %
Blessed	90.5 %	90.5 %	90.5 %	90.5 %
Thankful	90.5 %	90.5 %	90.5 %	90.5 %
Content	69.0 %	69.0 %	69.0 %	69.0 %
Angry	97.6 %	73.2 %	96.4 %	96.4 %
Sad	90.5 %	90.5 %	90.5 %	90.5 %
Anxious	90.5 %	90.5 %	90.5 %	90.5 %
Scared	97.6 %	97.0 %	97.6 %	97.6 %
Insecure	97.6 %	97.6 %	97.6 %	97.6 %
Uncertain	95.2 %	71.4 %	94.6 %	94.6 %
Frustrated	85.7 %	85.7 %	85.7 %	85.7 %
Stressed	92.9 %	92.3 %	92.9 %	92.9 %
Pessimistic	90.5 %	90.5 %	90.5 %	90.5 %
Settled	90.5 %	90.5 %	90.5 %	90.5 %
concern_romance	54.8 %	47.0 %	61.9 %	61.9 %
At Least 1 Concern	58.3 %	66.7 %	66.7 %	66.7 %
concern_career	35.7 %	54.2 %	60.1 %	60.1 %
concern_family	73.8 %	73.8 %	73.8 %	73.8 %
concern_appearance	90.5 %	90.5 %	90.5 %	90.5 %
concern_school	60.7 %	81.0 %	81.0 %	81.0 %

Table 13: Accuracy Comparisons of Machine Learning Models for Retrospective Data of PFS Dataset

Feature Description	LSTM Accuracy	CNN Accuracy	Decision Tree Accuracy	Ensemble Learning Accuracy
How happy you are today with...Your Health	24.4 %	22.6 %	22.6 %	22.6 %
How happy you are today with...Your Friends	19.6 %	17.3 %	31.5 %	31.5 %
How happy you are today with...Your Neighbors	7.7 %	5.4 %	10.7 %	10.7 %
How happy you are today with...Extend of which you help others	14.3 %	19.6 %	22.0 %	22.0 %
How happy you are today with...Things you do for fun	16.1 %	18.5 %	26.2 %	26.2 %
How happy you are today with...Your Romantic Life	10.7 %	14.3 %	16.1 %	16.1 %
How happy you are today with...Your Prospects of Having a Happy Marriage	7.1 %	12.5 %	11.9 %	11.9 %
How happy you are today with...Your Prospects of Having a Family with Children	19.0 %	17.9 %	20.2 %	20.2 %
How happy you are today	14.9 %	20.2 %	17.3 %	17.3 %

with...Your Physical Attractiveness				
How happy you are today with...Your Weight	16.1 %	11.3 %	18.5 %	18.5 %
How happy you are today with...Your Current career situation	17.9 %	11.9 %	20.8 %	20.8 %
How happy you are today with...Career prospects	7.7 %	8.9 %	8.9 %	8.9 %
How happy you are today with...Prestige of your current job/career status	13.1 %	14.3 %	11.9 %	11.9 %
How happy you are today with...Your Income	11.3 %	18.5 %	20.2 %	20.2 %
How happy you are today with...Your Finances	26.2 %	18.5 %	28.0 %	28.0 %
How happy you are today with...Your Achievement of personal goals	17.3 %	16.1 %	19.6 %	19.6 %
How happy you are today with...Your Life in general	17.3 %	18.5 %	29.8 %	29.8 %
How happy you are today with...Yourself in general	25.0 %	17.9 %	26.2 %	26.2 %