FoSIL at CheckThat! 2022: Using Human Behaviour-Based Optimization for Text Classification

Andy Ludwig[1], Jenny Felser[1], Jian Xi[1], Dirk Labudde[1] and Michael Spranger[1]

[1]University of Applied Sciences Mittweida, Technikumplatz 17, 09664 Mittweida, Germany

Abstract

Nowadays, a huge amount of information and news articles are available every day. The events of recent years have shown that Fake News can severely shake trust in politics and science. Unfortunately, a decision can only be made about the truthfulness of a fraction of all news and posts. In this respect, the CLEF2022-CheckThat! shared task 3a adresses this problem. In this paper, we propose a new classification approach using a novel metaheuristic feature selection algorithm that mimics human behavior. The results show that the performance of a baseline classifier can achieve higher performance by combining with this algorithm with only a fraction of the features.

Keywords

fake news detection, text classification, feature selection, human behavior-based optimization

1. Introduction

In times of constant availability of vast amounts of information, people have to judge the truth of news in a short time. This assessment is often neglected due to the fast-moving nature of the news. There are various reasons why authors, whether intentionally or unintentionally, contribute to the generation of untrustworthy content. In particular, sources that deliberately disseminate false information pose a danger to consumers of news.

Effective methods for detecting fake news are essential in the fight against the targeted spread of fake news. Continuous research and development of approaches to detect misinformation are highly important. This is one mission of the CLEF2022 - CheckThat! Lab [1, 2]. In general, the Lab's goal is to verify the veracity of claims. Task 3a takes up the challenge of assessing the truth content of news articles [3].

This paper presents a novel approach for text classification. The concept is based on human behaviour-based optimization (HBBO) [4]. This meta heuristic optimization approach uses some fundamental interactions and behaviours of humans. The potential of this adaptation was used for the fake news detection in task 3a.

The paper is structured as follows: section 2 presents related works, section 3 describes the human behaviour-based optimization, section 4 summarizes the adaption of the optimization approach for text classification, section 5 presents the given data and the conducted experiments, in section 6 the results are discussed and finally in section 7 a conclusion is given.

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

ludwig1@hs-mittweida.de (A. Ludwig); jfelser@hs-mittweida.de (J. Felser); xi@hs-mittweida.de (J. Xi); labudde@hs-mittweida.de (D. Labudde); spranger@hs-mittweida.de (M. Spranger)

2. Related Work

The assessment of the truthfulness of news or claims is a special case of text classification whose importance has increased greatly in recent times. Especially during the Covid-19 pandemic situation a larger group of people was challenged to judge news correctly. This situation makes the invention of automatic text classification systems necessary [5]. To pool a lot of competence and to push the development forward, a task for different participants can be done like the annual task CheckThat! Lab since 2018. The results of last year are summarized in [6]. These tasks need a high quality dataset [7], which is also reflected in the available labels [8].

To improve the results for fake news detection and the related text classification task, a wide range of possible approaches can be explored. In this paper, the research focus was on a new technique to select the best features for training. To solve these optimization tasks, nature inspired meta-heuristic techniques can be applied. An overview of already known approaches is given in [9]. For example, the ant colony optimization is a possible algorithm for feature selection in text classification tasks [10].

The nature-inspired algorithm used in this paper is derived from human behaviour. The basis of this approach was presented in [4]. This algorithm was used already in [11] in combination with self-organizing maps in context of cryptanalytic resistance. The authors compare this approach with other meta-heuristic solutions like ant colony optimization. Next to the approach of [4], other algorithms based on human behaviour were described. Firstly, in [12] the goal is to solve optimization tasks by simulating the phases of knowledge gaining and sharing in younger and older years of human life. A different aspect of optimization algorithms under usage of human behaviour is presented in [13]. This approach focuses on the adaption of behaviours and manners of other humans for example in family structures.

3. Human Behaviour-Based Optimization (HBBO)

In this section, we briefly describe Ahmadi's novel swarm intelligence-based optimization approach [4] considering human behavior, which forms the basis for the feature selection approach used here. A central aspect in all phases of the algorithm is an individual's pursuit of self-optimization at different stages of his or her life. At the same time, different individuals have different levels of experience in their field, and some of them become experts in one of them (e.g., art, music, or science, etc.).

The optimization of individual performance is carried out iteratively in four main steps: *Initialization*, *Education*, *Consultation*, and *Field Changing*. In the *Initialization* step the population is built. Each individual in the population is assigned an area of interest in which improvement should be achieved. Depending on the underlying optimization problem, an individual is represented as a vector of characteristic variables (features). $I = \begin{pmatrix} x_1 & x_2 & \dots & x_N \end{pmatrix}^T$. An expert is determined for each field. The expert provides the best function value depending on his individual feature set. The formal definition of an expert is shown in equation (1), where I denotes a person, E denotes an expert, and F denotes the actual field of expertise.

$$I^{(F)} \rightarrow E^{(F)} = \text{argmin}\ \ f^{\text{error}}(F) \tag{1}$$

The first step in the improvement process, is *Education*. Individuals in each area learn from the respective expert. The learning process comprises the improvement of the individual characteristic values by those of the expert and aims at the reduction of their own error function value.

A similar procedure is used in *Consultation*. In this stage, individuals can learn from any other individual, not only from an expert. For this purpose, some variables are merged between two randomly selected individuals. The consultation is called effective and the merged set of variables is kept if the updated variables lead to better function values. Otherwise, the update is reversed.

The last step is *Field Changing*. Whether an individual changes his field of interest is calculated using a rank probability method and the roulette wheel selection.

After Initialization, the three remaining steps are repeated until a stop criterion is met, e.g. if the average function values do not change (or change too little) or the number of iterations reaches a maximum.

This algorithm can be adapted to find an optimal feature set for text classification. The next section explains the adaptation in detail.

4. Adaptation for Text Classification

In order to adapt the HBBO approach for feature selection, the optimization objective, i.e., an objective function, must be determined. Here, the F_1-score is chosen on a test set. For this purpose, the preprocessed dataset must be divided into training and test data. To obtain reliable results and avoid overfitting, the data set is split in terms of k-fold cross-validation. Each fold results in an optimal feature set in the form of a document term matrix (DTM) that is able to classify the respective test data with the greatest possible success. Finally, the results of all the folds are merged into a DTM that contains only the most successful features to classify the entire dataset. For assessing the performance after each optimization step, a support vector machine (SVM) is used.

4.1. Initialization

During Initialization, each input document is considered as an single individual and encoded in the form of a bit vector. This vector represents an individual's knowledge. Each individual is assigned the class of the respective document, which also represents the field of interest. All vectorized documents together form the entire population. In contrast to the original algorithm where each individual is optimized, here subsets of individuals are optimized together. This leads to a smaller set of documents achieving good or even better classification results than the original set of all documents. A group of individuals contains all fields, which leads to a simultaneous optimization of all classes within a group.

For this purpose, the individuals are grouped into subsets, with each subset equally populated with individuals of each class (stratified approach). Each group of individuals is optimized separately. The size of the subsets is a hyperparameter.

The remaining three steps are applied iteratively to each group of each fold. The number of iterations needs to be specified as a hyperparameter.

4.2. Education

In the Education step, a group of individuals must first be determined for each field that achieves the lowest error with respect to the objective function on the test set according to equation (2). Each of these groups is considered to be the expert group for that particular field.

$$I^{(F)} \rightarrow E^{(F)} = \operatorname{argmax} F_1 \tag{2}$$

Subsequently, a subset S_E of features of the expert group, considering all classes, is merged with the features of other individuals, belonging to the remaining groups. This procedure corresponds to the non-experts approaching the expert group by updating their feature vector. The number of terms transferred in this step is another hyperparameter, as well as the number of adapted individuals, i.e., documents. Equation (3) shows the formal definition of this step, where $I_i^{(F)}$ denotes the i^{th} individual I in the specific field F and $S_E \subseteq E^{(F)}$.

$$I_i^{(F)} = \begin{cases} I_i^{(F)} \cup S_E, & \text{if} f^{\text{error}}(I_i^{(F)} \cup S_E) < f^{\text{error}}(I_i^{(F)}) \\ I_i^{(F)}, & \text{otherwise} \end{cases} \tag{3}$$

The feature update is considered successful and the merged feature set is retained only if the individual achieves an improvement with respect to the objective function, i.e., better classification results. Otherwise, the previous feature set remains unchanged.

4.3. Consultation

In Consultation, the basic procedure is very similar to that in the Education step. The main difference is in the group of individuals from which the features for merging are taken. Regardless of expert status, features are merged between two randomly selected groups of individuals. This leads to greater heterogeneity of terms across all groups. Equation (4) shows the formal definition of Consultation, where S_j is a subset of all features of individual I_j and $S_j \subseteq I_j^{(F)}$.

$$I_i^{(F)} = \begin{cases} I_i^{(F)} \cup S_j, & \text{if} f^{\text{error}}(I_i^{(F)} \cup S_j) < f^{\text{error}}(I_i^{(F)}) \\ I_i^{(F)}, & \text{otherwise} \end{cases} \tag{4}$$

As in Education, the updated feature set is retained only if the F_1-score improves; otherwise, the previous terms remain unchanged. Again, the number of terms exchanged and the number of individuals paired can be controlled using hyperparameters.

4.4. Field changing

The last step - Field Changing - does not manipulate the individuals' terms, but changes the field associated with them. The number of randomly selected individuals changing the field is another hyperparameter.

In a multilevel classification, an individual can simply change their area of interest to that of the most successful expert group. For this purpose, all expert groups are ordered by their F_1-score, as shown in equation (5), where $E^{(F_x)}$ is the expert group in the respective area.

$$R = \left\{ E^{(F_1)}, ..., E^{(F_n)} \mid f^{\text{error}} \left(E^{(F_1)} < E^{(F_2)} < ... < E^{(F_n)} \right) \right\} \tag{5}$$

Afterwards, the expert group with the highest rank in R determines the field of the individual willing to switch, as shown in equation (6).

$$F_x^* = \left\{ F_x | E^{(F_x)} = \text{argmin}_{E^{(F_x)} \in R} f^{error} \left(E^{(F_x)} \right) \right\} \tag{6}$$

In the special case of a binary classification, the process simply reduces to a field change, i.e., an inversion of the class label. Again, the new field is retained only if it leads to an improvement in the objective function value.

5. Experiments and Results

In this work, the adaptation of HBBO, as discussed in section 4, was applied to the problem of detecting fake news in the context of the CLEF2022-CheckThat! shared task 3a. The documents provided were news articles in English, which were to be grouped into the classes true, false, partially false and other with regard to potentially containing fake news. For training the model, the provided training and development sets were combined so that the total training corpus comprised $1,264$ documents, each of which was assigned to exactly one of the four mentioned classes. The resulting class distribution of the training corpus is shown in Figure 1.

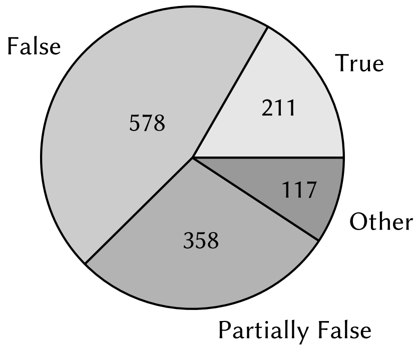

Figure 1: Distribution of classes and documents in the corpus of training data.

Before the HBBO algorithm can be applied, the input data must be cleaned. To this end, a wide range of cleaning steps were performed:

- Combination of article and headline into one pseudo document,
- Replacement of newlines and tabs with white space,
- Removal of emojis and links,
- Removal of special characters, punctuation and numbers,
- Removal of stop words,
- Lemmatization.

For simplicity, a bit vector was chosen to represent each single document. Subsequently, the cleaned dataset was divided into samples in terms of 5-fold cross-validation before HBBO could be applied to each sample, resulting in an optimized feature set for each fold.

In order to observe the specific behavior of the new algorithm, the problem is considered as several binary classification tasks. In this respect, for each class, a model is trained using the training data without applying any kind of balancing technique. Table 1 show the values chosen for the experiments.

Table 1
Summary of hyperparameter.

Phase of algorithm	Hyperparameter	Value
Initialization	Number of different data sets for cross validation	5
	Number of documents (individuals) per subset during optimization	24
	Number of iterations for all phases	125
Education	Number of terms for exchange	5
	Percent of documents for adaptation of features	100
Consultation	Number of terms for exchange	3
	Percent of documents for adaption of features	100
Field changing	Number of changed labels	1

Figure 2 shows the history of the F_1-score for all 125 iterations for the category *true*. Shown in green color is the best group of individuals in each iteration, while black dots show the mean of F_1-score of all groups and the red horizontal line symbolizes the baseline F_1-score trained without HBBO feature selection. The results of the remaining classes are shown in Figure 3, Figure 4, and Figure 5 accordingly.

In order to provide a quantitative overview, the respective results are summarized in Table 2. In addition, the mean value is calculated for each class. The system yields a $macro\ F1 = 0.602$ over all classes.

Table 2
Results in terms of F_1-Score for each fold of a 5-fold cross-validation.

Class	F_1-**Score**					
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
True	0.508	0.464	0.553	0.558	0.574	0.531
False	0.765	0.762	0.777	0.787	0.764	0.771
Partially False	0.633	0.601	0.633	0.602	0.645	0.623
Other	0.476	0.45	0.468	0.416	0.607	0.483

As mentioned earlier, the results of all folds must be merged to combine them into one classifier. This can be achieved simply by concatenating the best group of individuals in each fold. The final feature matrix contains 120 pseudo documents (5 groups of 24 features each).

In order to uniquely assign each document of the test data to a class, each separately optimized binary model was used for classification in the order of their performance. Table 3 summarizes the results of this final classification procedure. Here, $macro\ F_1 = 0.251$ with an accuracy of 0.462. With this result we have reached the 18^{th} place of the shared task 3a.

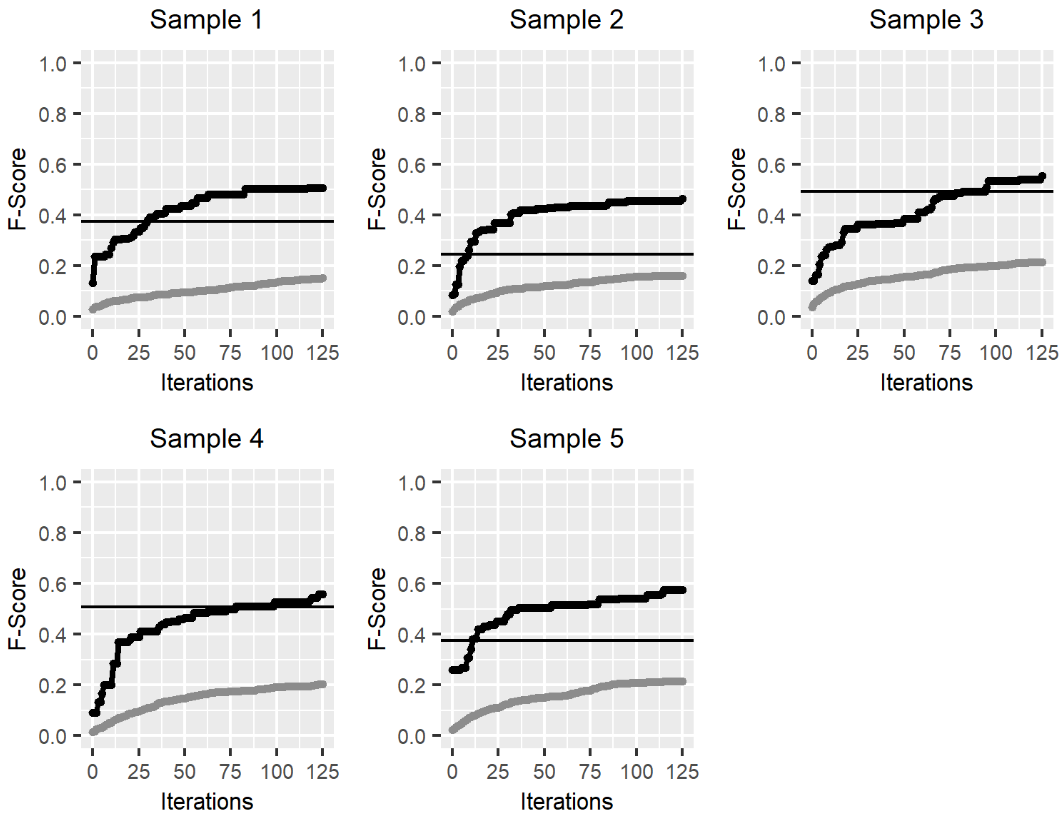

Figure 2: HBBO results after 125 iterations for first category: true; black horizontal line: own baseline without HBBO, black: F_1-score of the best group of individuals, grey: mean F_1-score of all groups of individuals.

Table 3
Evaluation results using hold-out data.

Class	Precision	Recall	F_1-Score
True	0.391	0.086	0.141
False	0.573	0.806	0.670
Partially False	0.161	0.179	0.169
Other	0.016	0.032	0.022

6. Discussion

The results shown are surprisingly low. Nevertheless, HBBO as a feature selection algorithm has a high potential for classification tasks. However, there are still some problems and open research questions. First of all, the drop of *macro* F_1 from training (0.602) and test (0.251) data,

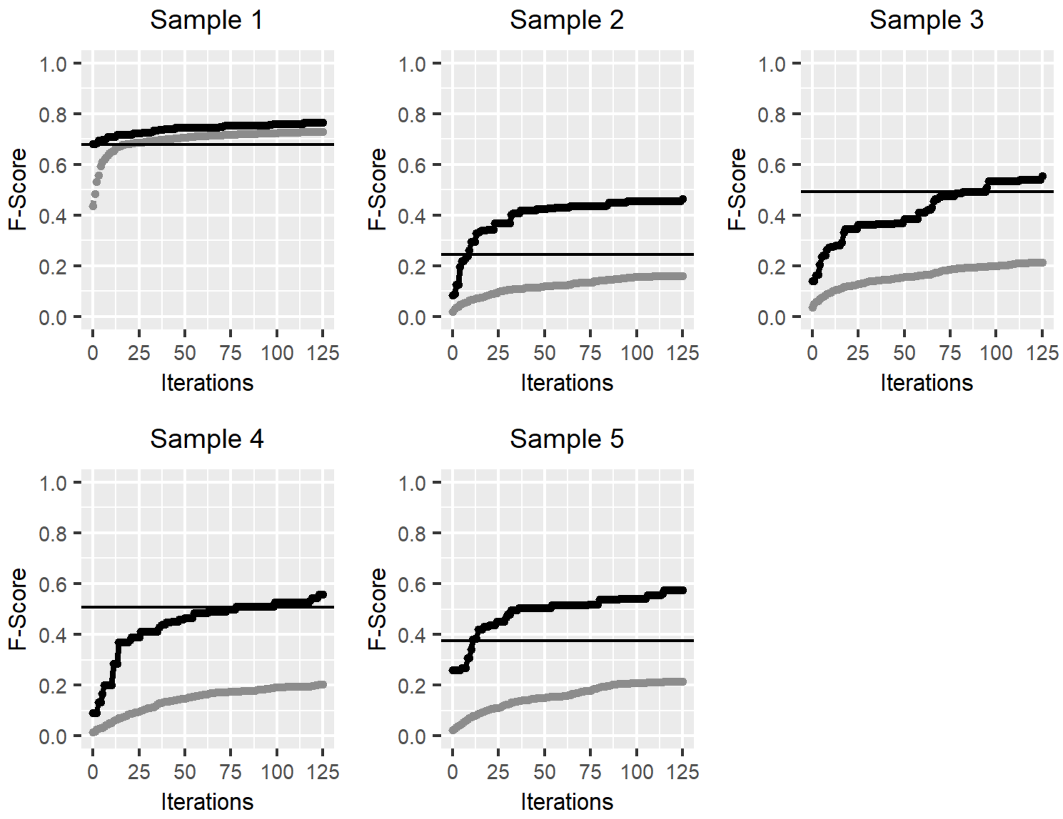

Figure 3: HBBO results after 125 iterations for second category: false; black horizontal line: own baseline without HBBO, black: F_1-score of the best group of individuals, grey: mean F_1-score of all groups of individuals.

which might indicate a slight overfitting. The optimization of each sample takes into account the respective test set. This process may lead to good classification results only within that particular test set. Further studies need to be performed to test this hypothesis.

Further performance gains could be achieved by structured experimentation with the hyperparameters or by using more advanced features, such as TF-IDF or BM25. Inseparable from the chaining of binary classifiers is the question of their best order. Among other strategies, the shift to true multi-label classification could also be beneficial. For this, further adaptations of the HBBO algorithm have to be made.

In addition, the strict requirement for improvement at each optimization step can lead to miss optimal feature combinations. This could be remedied by allowing temporary deteriorations.

In general, the task of automatically carrying out a fine-grained classification of a document with regard to its truthfulness is difficult to imagine. Inseparable from the spread of fake news is the task of making it appear as real as possible. Thus, there can be no statistically detectable linguistic features that clearly indicate the truthfulness of a document, as can be

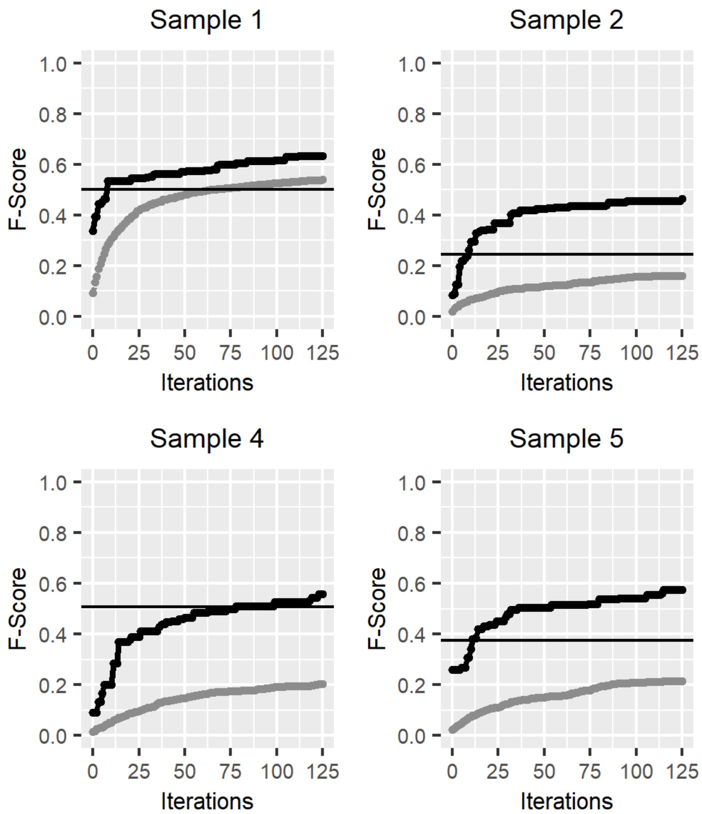

Figure 4: HBBO results after 125 iterations for third category: partially false; black horizontal line: own baseline without HBBO, black: F_1-score of the best group of individuals, grey: mean F_1-score of all groups of individuals.

easily demonstrated analytically. Ultimately, this can only be determined by a fact check. This insight is also underlined by the overall results achieved by all participants in this shared task.

7. Conclusion

In this paper a novel feature selection algorithm for text classification using an human behavior-based optimization approach is presented in order to solve task of fine-grained fake news detection. The algorithm shows an improved performance compared to classification using a single SVM. In addition, the enormous reduction in training input after optimization is remarkable. In each sample, only 24 optimized pseudo-documents were able to outperform the baseline calculated considering all documents.

Nevertheless, further experiments must be carried out to find the best values for the hyperparameters. Further improvements might achieved by using a more advanced term representation.

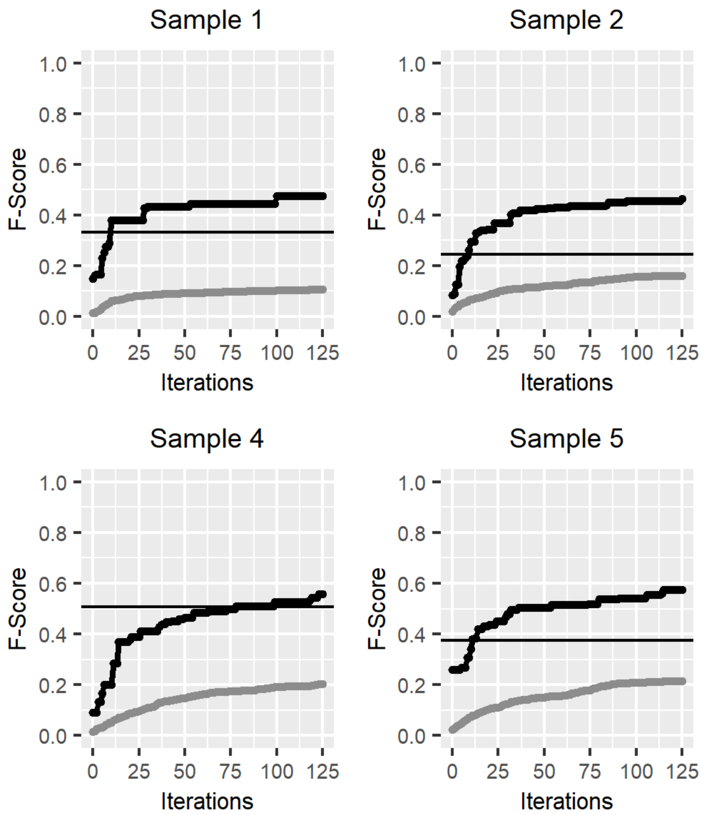

Figure 5: HBBO results after 125 iterations for fourth category: other; black horizontal line: own baseline without HBBO, black: F_1-score of the best group of individuals, grey: mean F_1-score of all groups of individuals.

References

[1] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, The clef-2022 checkthat! lab on fighting the covid-19 infodemic and fake news detection, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2022, pp. 416–428.

[2] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, M. Wiegand, M. Siegel, J. Köhler, Overview of the CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection, in: Proceedings of the 13th International Conference of the CLEF Association: Information Access Evaluation meets Multilinguality, Multimodality, and Visualization, CLEF '2022,

Bologna, Italy, 2022.
[3] J. Köhler, G. K. Shahi, J. M. Struß, M. Wiegand, M. Siegel, T. Mandl, Overview of the CLEF-2022 CheckThat! lab task 3 on fake news detection, in: Working Notes of CLEF 2022—Conference and Labs of the Evaluation Forum, CLEF '2022, Bologna, Italy, 2022.
[4] S.-A. Ahmadi, Human behavior-based optimization: A novel metaheuristic approach to solve complex optimization problems, Neural Comput. Appl. 28 (2017) 233–244. URL: https://doi.org/10.1007/s00521-016-2334-4. doi:10.1007/s00521-016-2334-4.
[5] G. K. Shahi, D. Nandini, FakeCovid – a multilingual cross-domain fact check news dataset for covid-19, in: Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media, 2020. URL: http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf.
[6] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the clef-2021 checkthat! lab task 3 on fake news detection, Working Notes of CLEF (2021).
[7] G. K. Shahi, Amused: An annotation framework of multi-modal social media data, arXiv preprint arXiv:2010.00502 (2020).
[8] G. K. Shahi, A. Dirkson, T. A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, Online Social Networks and Media 22 (2021) 100104.
[9] M. Sharma, P. Kaur, A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem, Archives of Computational Methods in Engineering 28 (2020). doi:10.1007/s11831-020-09412-6.
[10] J. Xi, M. Spranger, D. Labudde, Music event detection leveraging feature selection based on ant colony optimization 13 (2020) 36–47.
[11] R. Soto, B. Crawford, F. G. Molina, R. Olivares, Human behaviour based optimization supported with self-organizing maps for solving the s-box design problem, IEEE Access 9 (2021) 84605–84618. doi:10.1109/ACCESS.2021.3087139.
[12] A. Wagdy, A. Hadi, A. Khater, Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm, International Journal of Machine Learning and Cybernetics 11 (2020). doi:10.1007/s13042-019-01053-x.
[13] M. Kumar, A. J. Kulkarni, S. C. Satapathy, Socio evolution & learning optimization algorithm: A socio-inspired optimization methodology, Future Generation Computer Systems 81 (2018) 252–272. doi:https://doi.org/10.1016/j.future.2017.10.052.