Synergy Conformal Prediction

Niharika Gauraha[a], Ola Spjuth[a]

[a]*Department of Pharmaceutical Biosciences*
Uppsala University
Uppsala, Sweden

Abstract

Conformal Prediction is a machine learning methodology that produces valid prediction regions under mild conditions. Ensembles of conformal predictors have been proposed to improve the informational efficiency of inductive conformal predictors by combining p-values, however, the validity of such methods has been an open problem. We introduce Synergy Conformal Prediction which is an ensemble method that combines monotonic conformity scores, and is capable of producing valid prediction intervals. We study the applicability in two scenarios; where data is partitioned in order to reduce the total model training time, and where an ensemble of different machine learning methods is used to improve the overall efficiency of predictions. We evaluate the method on 10 data sets and show that the synergy conformal predictor produces valid predictions and improves informational efficiency as compared to inductive conformal prediction and existing ensemble methods. The results indicate that synergy conformal prediction has advantageous properties compared to contemporary approaches, and we also envision that it will have an impact in Big Data and federated environments.

Keywords:
Conformal Prediction, Machine Learning, Synergy Conformal Prediction, Big Data, Federated Learning, Conformal Predictor Ensembles

Email addresses: niharika.gauraha@farmbio.uu.se (Niharika Gauraha), ola.spjuth@farmbio.uu.se (Ola Spjuth)

1. Introduction

Conformal prediction is an established method in the machine learning landscape that yields valid prediction regions for new objects (Vovk et al., 2005). Transductive conformal prediction (Vovk, 2013) and Inductive conformal prediction (Papadopoulos, 2008) are two approaches for the construction of prediction regions. The transductive approach requires re-training the model on each prediction, and inductive conformal prediction was developed to overcome this. The validity of Transductive Conformal Predictors (TCPs) and Inductive Conformal Predictors (ICPs) is proven in that they produce $1 - \epsilon$ expectation tolerance regions, where ϵ is the selected significance level (Vovk et al., 2005).

The inductive conformal prediction approach comes at the expense that part of the data needs to be set aside as a calibration set, and the model hence becomes less informational efficient. Conformal predictor ensembles, such as cross-conformal predictors (Vovk, 2015) and aggregated conformal predictors (Carlsson et al., 2014a), have been developed to improve the informational efficiency of ICPs. The idea is to train multiple ICPs to make better use of the examples for modeling and calibration, and aggregate the resulting p-values. However, the validity of these ensemble methods has been an open problem for researchers (Linusson et al., 2017). Though it has been shown to achieve empirical exact validity under certain conditions, its theoretical validity has not been proven yet and it remains a practical problem in many settings. Since the aforementioned conformal predictor ensembles try to aggregate conformal p-values, which usually does not yield standard uniform distribution, the validity is not guaranteed.

Our research tries to answer these two questions: (1) Can we partition a dataset and train multiple models and obtain valid prediction regions with good accuracy when combined? (2) Can we train multiple models on a dataset using different modeling methods, and obtain valid prediction regions with good accuracy when combined?

In this manuscript, we introduce *Synergy Conformal Prediction* (SCP) which is a method that aggregates monotone conformity scores instead of p-values. This work is inspired by the method of "Synergy of Monotonic Rules" proposed in Vapnik and Izmailov (2016), in which the authors have combined the estimated conditional probabilities to construct the optimal synergy rules for pattern recognition problems.

The organization of the paper is as follows. In section 2, we introduce the

background concepts and notations used throughout the paper. In Section 3, we present synergy conformal prediction and we prove its useful properties. In Section 4, we perform numerical analysis on a set of real data sets. In Section 5 we discuss the results and the applications of SCP. Finally, in Section 6 we conclude and provide directions for further research.

2. Background

This section gives a brief background about conformal prediction framework and fixes notations and assumptions used throughout the paper.

The object space is denoted by $\mathcal{X} \in \mathbb{R}^p$, where p is the number of features, and label space is denoted by $\mathcal{Y} \in \{1, \ldots, K\}$, where K is the number of labels. We assume that each example consists of an object and its label, and its space is given as $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$. In a typical classification problem, given a training dataset $Z = \{z_1, ..., z_\ell\}$ – where ℓ is the number of examples in the training set, and each example $z_i = (x_i, y_i)$ is labeled – we want to predict the label of a new object x_{new} whose label is unknown. Our experiments are limited to classification problems, and we also assume the exchangeability of examples throughout the paper.

The conformity measure is the score from a function that measures the closeness of an example in relation to the previous examples. In our experiments, we use the calibrated conditional probability from classifiers such as support vector machine and random forest as the comformity measure. To compute the corresponding p-values, we use the smoothed Mondrian approach (Vovk et al., 2005), where the taxonomy is defined by the labels.

Conformal prediction provides a layer on top of an existing machine learning method and uses available data to determine valid prediction regions for new objects (Vovk et al., 2005). Conformal prediction was primarily defined as an online transductive framework, in which the underlying model must be retrained each time an object is to be predicted and is hence computationally expensive. For further details on the transductive approach, we refer to Vovk et al. (2005). A more computational efficient inductive framework was then proposed as an alternative to the transductive approach. In the inductive approach, the training data has to be divided into a proper training set and a calibration set. Since ICP is a key building block for our method, we define it in a bit more detail in the following; for full length details we refer to Papadopoulos (2008).

Definition 1 (Inductive Conformal Predictor (ICP)). *Given a training set of ℓ examples, $Z = \{z_1, ..., z_\ell\}$, drawn from an exchangeable distribution P. The training data is first divided into a proper training set $\{Z_T\}$ and a calibration set $\{Z_C\}$, where (T, C) is a partition of $\{1, ..., \ell\}$. Let $\mathcal{A}$ be a conformity measure, and the conformity score $\mathcal{A}(Z_T, z)$ is used to measure how well the example z conforms to the proper training set Z_T. For example,*

$$\mathcal{A}(Z_T, (x, y)) = p(Y = y \mid f(x)), \tag{1}$$

where $f : X \to Y$, is a prediction rule of the model trained on the proper training set Z_T, and $p(Y = y \mid f(x))$ is a calibrated conditional probability. The predictive model trained on the proper training set Z_T, is then used to compute the conformity scores (i.e. class conditional probabilities) for the calibration set, $\alpha_j^y = p(Y = y \mid f(x_j))$, for $j \in C$ and $y \in \mathcal{Y}$. Let x_{new} (follows the same distribution P) be the example we want to predict, and let α_{new}^y be its conformity scores computed using the same function $\mathcal{A}(Z_T, .)$. The ICP p-values are then computed as

$$p_{new}^y = \frac{|j \in C_y; \alpha_{new}^y > \alpha_j^y| + \tau |j \in C_y; \alpha_{new}^y = \alpha_j^y|}{|C_y| + 1}, \tag{2}$$

where C_y denotes the class-wise partition of C (if $j \in C_y$, then $j \in C$ and $y_j = y$), $y \in \mathcal{Y}$, and $\tau \in [0, 1]$ is a random number.
The inductive conformal predictor corresponding to the tuple $(Z_C, \mathcal{A}(Z_T, .))$ is defined as a set predictor

$$\Gamma^\epsilon = \{y \mid p^y > \epsilon\}, \tag{3}$$

where $\epsilon \in (0, 1)$ is a chosen significance level, and $(1 - \epsilon)$ is known as confidence level.

As mentioned before, in ICP some examples are used for modeling only and some are used for calibration only, and this makes ICP less informational efficient than TCP. To improve the informational efficiency of ICP, ensembles of conformal predictors were proposed. Cross Conformal Predictor (CCP), and Bootstrap Conformal Predictor (BCP) were proposed in Vovk (2015) and it was generalized as Aggregated Conformal Predictor (ACP) in Carlsson et al. (2014b). In CCP, training data is divided into separate folds and each fold is used as calibration set and remaining data is used as a proper training set, and eventually p-values are averaged across all folds. In BCP,

training set is bootstrapped to obtain a proper training set and out-of-bag examples are used as a calibration set. The p-values are then aggregated across bootstrap replications. The ACP is a generalization of CCP and BCP, where a consistent resampling scheme is used for constructing the calibration set. We refer to Carlsson et al. (2014b) for details on ACP.

Other ensemble methods in the conformal prediction framework have been proposed, including Toccaceli and Gammerman (2018), Löfström et al. (2013) and Balasubramanian et al. (2015). In all the above ensemble methods, the main aim was to get more informational efficient conformal predictors by combining p-values. However, most of the resulting models are not guaranteed to be valid, as the combined p-values need not be uniformly distributed. For a detailed study on the validity of such ensemble methods, we refer to Linusson et al. (2017).

In our experiments, to assess the quality of a conformal predictor we consider validity and efficiency. Validity is empirically assessed in terms of calibration plots, the plot of the percentage of errors against $\epsilon \in (0, 1)$). For a valid prediction, the calibration plots are usually very close to the bisector of the first quadrant. We use observed fuzziness (Vovk et al., 2016) as our measure of efficiency, which is defined as the sum of all p-values for the incorrect class labels.

3. Synergy Conformal prediction

In all the ensemble methods discussed in the previous section, the key idea is to combine p-values computed from various ICPs. We propose a novel method of combining (monotone) conformity scores for the calibration set and for the test examples computed from various trained models. In the following, we define "Synergy Conformal Prediction".

Given ℓ examples, $Z = \{z_1, ..., z_\ell\}$, drawn from an exchangeable distribution P. Akin to the ICP method, here also the training data is first divided into the proper training set $\{Z_T\}$ and the calibration set $\{Z_C\}$, where (T, C) is a partition of $\{1, ..., \ell\}$. But then the proper training data is further divided into M non-empty disjoint subsets and each subset $Z_{T_m}, m = 1, \ldots, M$ is then used as an actual training set for modeling. $(T_1, ..., T_M)$ is a partition of T. The M predictive models trained on individual partitions are then used to compute the monotonic conformity scores (i.e. class conditional probabilities) for the calibration set denoted by α^y_{mj}, for $j \in C$, $m = 1, \ldots, M$, $y \in \mathcal{Y}$.

For example,

$$\alpha_{mj}^{y} = p_m(y = y \mid f_m(x_j)), \tag{4}$$

where $f_m(x)$ is the prediction rule defined by the predictive model trained on the m^{th} part of the training set. The aggregated conformity scores across models are then defined as

$$\alpha_j^y = \frac{1}{M} \sum_m \alpha_{mj}^y$$

Let x_{new} (follows the same distribution P) be the examples we want to predict, and let α_{new}^y be the aggregated conformity scores across models. The SCP p-values are then computed using eq. (1). The synergy conformal predictor corresponding to the tuple $(Z_C, \mathcal{A}(Z_{T_1}, .), ..., \mathcal{A}(Z_{T_M}, .))$ is defined as a set predictor as given in eq. (1).

Algorithm 1: Synergy Conformal Predictor

Input: training dataset:Z, object to predict:x_{new}, a conformity measure:A

Output: p-values

Step1: Split the training set into two smaller sets, $\{Z_T\}$ and the calibration set $\{Z_C\}$. (T, C) is a partition of $\{1, ..., |Z|\}$.

Step1: Split the set Z_T into M proper training sets, $\{Z_{T_m}, m = 1, \ldots, M\}$. $(T_1, ..., T_M)$ is a partition of T.

Step2: For each part Z_{T_m}, train and construct the rule to generate conformity scores.

Step3: Compute the aggregated conformity scores across M models for each example in the calibration set. α_j^y, for $j \in C$ and $y \in \mathcal{Y}$.

Step4: Compute the aggregated conformity scores across M predictive models to the object x_{new}, which results in α_{new}^y, for $y \in \mathcal{Y}$.

Step5: Compute p-values for each of the K classes using eq. 1.

return $(p_{new}^1, \ldots, p_{new}^y)$;

The SCP method differs from the ensemble methods discussed in the previous section by:

1. Ensemble methods combine conformal p-values obtained from different ICPs.

2. Within the inductive conformal prediction framework, SCP tries to improve the informational efficiency by combining monotone conformity scores (for example, calibrated probability estimates from support vector machines).

3.1. Properties of Synergy Conformal Predictors

In this section, we discuss the properties of SCP. First we prove that the SCP algorithm produces valid predictions.

Proposition 1. *The synergy conformal predictor is exactly valid.*

Proof. When we do not partition the proper training set into smaller subsets, in that case SCP is exactly ICP and hence exactly valid. In particular for this case, the whole partition Z_T is used for predictive modeling, and the pair calibration set Z_C and the conformity measure based on the prediction rule obtained from Z_T, forms an ICP, and an ICP is proven to be exactly valid (Vovk et al., 2005).

Alternatively, SCP can be viewed as a single ICP and hence exactly valid, where synergy of monotone conformity scores is considered as one function producing the (aggregated) conformity scores. To illustrate this, let us consider partition of the set Z_T into two subsets, Z_{T_1} and Z_{T_2}, and let their corresponding rules for computing conformity scores be $\mathcal{A}(Z_{T_1}, .)$ and $\mathcal{A}(Z_{T_2}, .)$ respectively. By the definition of ICP, each individual ICP based on the tuple $\{Z_C, \mathcal{A}(Z_{T_m}, z)\}$ for $m = 1, 2$, is valid. Let us define a new conformity score $\mathcal{A}(Z_T, .)$ which aggregated the conformity scores of an example z, $\mathcal{A}(Z_T, z) = \frac{1}{2}\sum_{m=1}^{2} \mathcal{A}(Z_{T_m}, z)$. The pair $\{Z_C, \mathcal{A}(Z_T, .)\}$ forms an ICP corresponding to the new conformity measure $\mathcal{A}(Z_T, .)$, hence exactly valid, and it can be generalized for any number of subsets. □

Next, we prove that the SCP is more efficient than each individual small ICP of a partitioned dataset.

Proposition 2. *The synergy conformal predictor corresponding to the tuple* $(Z_C, \mathcal{A}(Z_{T_1}, .), ..., \mathcal{A}(Z_{T_M}, .))$ *is more informational efficient that each individual small ICP corresponding to the tuple* $(Z_C, \mathcal{A}(Z_{T_m}, .))$ *for* $m = 1, \ldots, M$.

Proof. The choice of calibrated conditional probability of support vector machine as conformity measure, is a monotonically increasing function. The aggregation of monotone estimated conditional probabilities (conformity scores)

across various models are monotone, and has been proven to be more accurate than the individual scores (Vapnik and Izmailov, 2016). Moreover, the subsets used for training different predictive models are independent, thus averaging of M conditional probability values decreases the variance of resulting conditional probability by a factor of M. That follows the combined scores are more robust than the individual ones. Hence SCP is more informational efficient than each individual ICP. □

4. Experiments

We evaluate SCP on ten classification datasets from UCI machine learning repository (Lichman et al., 2013). Specific breakdowns for the corresponding training, calibration and test sets are given in Table 1. For the first two experiments we have the following splits for SCP, an overview of which is depicted in Figure 1. Each dataset is randomly partitioned into a test set (20%) and the remaining training set (80%) is then split into a proper training set and a calibration set in proportion $2:1$. The proper training data set is randomly split into three equal disjoint subsets. Each partition is considered as a actual training set to train the underlying machine learning algorithm and compute conditional probabilities (as conformity measure), and then conditional probabilities are aggregated across all the models to compute the p-values. The experiments are repeated 10 times (to get an idea of how much their results are affected by the random splits) and the average informational efficiency (observed fuzziness) across ten runs are reported.

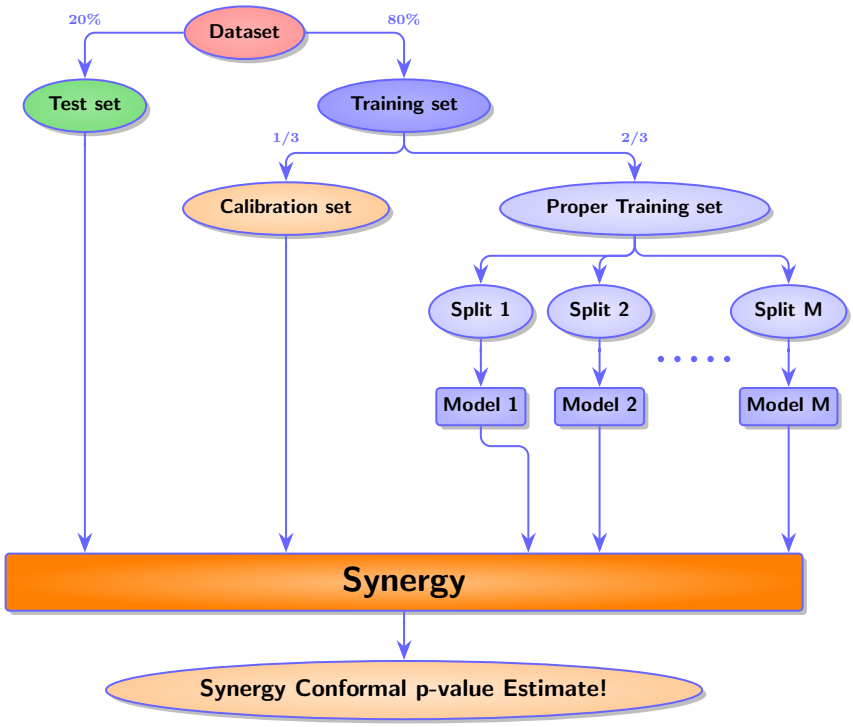


Figure 1: An overview of SCP

Table 1: Description of the datasets from UCI repository that are used in the evaluation.

Dataset	Training	Calibration	Test	Features
Spambase	2453	1227	921	57
Breast Cancer Wisconsin	303	152	114	30
Phishing Websites	5896	2948	2211	30
Covertype	9296	4648	3486	54
Adult	17365	8683	6513	14
Tic-tac-toe	510	256	192	9
Australian	368	184	138	14
MONK's-1	296	148	112	6
MONK's-2	320	160	121	6
Bank	2410	1206	905	16

4.1. Experiment 1: Synergy of Conformal Prediction using the Same Machine Learning Algorithm on Partitioned Data

The objective of this study is to compare SCP with each individual small ICPs (trained on individual part) and with the grand ICP (trained on the whole proper training set). We consider the form of SCP as given in Figure 1 with three equal partitions, $M = 3$. Linear Support Vector Machine (linear SVM) was used as the underlying machine learning method in this experiment. The average efficiency of SCP across 10 repetitions is reported in the third column of the Table 2. The first column is the average efficiency of the best of small ICPs trained on individual partitions (we choose the best ICP, which gives the minimum value for observed fuzziness), we call it the 'small (best) ICP'. The second column is the average efficiency of the ICP, where the proper training set as a whole is used for modeling.

Table 2: Comparison of efficiency between small (best) ICP, ICP and SCP, where linear SVM is used as the underlying machine learning algorithm.

Dataset	Small (best) ICP	ICP	SCP
Spambase	0.035	0.028	0.027
Breast Cancer Wisconsin	0.006	0.006	0.004
Phishing Websites	0.022	0.021	0.021
Covertype	0.434	0.371	0.376
Adult	0.144	0.144	0.144
Tic-tac-toe	0.400	0.387	0.397
Australian	0.083	0.067	0.069
MONK's-1	0.272	0.264	0.252
MONK's-2	0.487	0.460	0.485
Bank	0.152	0.139	0.139

4.2. Experiment 2: Synergy of Conformal Prediction using Different Machine Learning Algorithms on Partitioned Data

The objective of this study is to compare SCP with each individual small ICPs (trained on individual partitions), in order to show that using synergy of different machine learning algorithms in SCP actually improves the informational efficiency. We use the same setup as in Figure 1 with $M = 3$, and with three different machine learning algorithms: RBF SVM, linear SVM

and Random Forest (RF) one for each partition. The results are reported in the Table 3

Table 3: Comparison of efficiency between small (best) ICP and SCP with synergy of three different algorithms: RBF SVM, linear SVM and RF.

Dataset	Small (best) ICP	SCP
Spambase	0.033	0.018
Breast Cancer Wisconsin	0.009	0.005
Phishing Websites	0.022	0.010
Covertype	0.419	0.249
Adult	0.146	0.097
Tic-tac-toe	0.408	0.082
Australian	0.088	0.069
MONK's-1	0.275	0.077
MONK's-2	0.475	0.259
Bank	0.156	0.112

4.3. Experiment 3: Synergy of Conformal Prediction using Different Machine Learning Algorithms on Unpartitioned Data

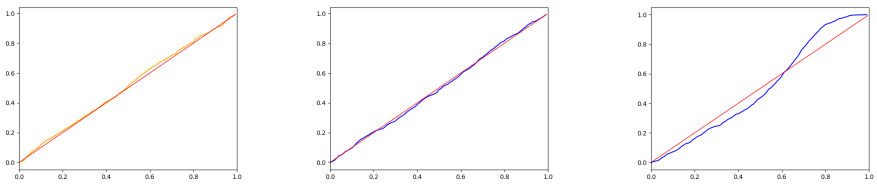
The objective of this study is to compare SCP with each individual big ICPs and CCP, in order to show that using different machine learning algorithm for training the same data (proper training set) in SCP also improves the informational efficiency. In this experiment, we do not partition the proper training set and instead we use three different machine learning algorithms: RBF SVM, linear SVM and RF on the same data for constructing SCP. The results are reported in the Table 4. The results of the individual (best) ICP and the SCP are given in first column and second column respectively. The third column is the average informational efficiency of the Cross Conformal Prediction (CCP) applied on the whole training set with three fold cross conformal prediction using linear SVM as an underlying machine learning algorithm.

Table 4: Comparison of efficiency between ICP, SCP and CCP. The first column shows results from the small best ICP while constructing SCP, The second column shows results for SCP when ensemble of different machine learning algorithms (RBF SVM, linear SVM and RF) are used on unpartitioned proper training set, and the third column shows results for the three fold CCP with linear SVM used for training.

Dataset	Best ICP	SCP	CCP
Spambase	0.028	0.014	0.028
Breast Cancer Wisconsin	0.008	0.006	0.007
Phishing Websites	0.023	0.006	0.021
Covertype	0.409	0.204	0.400
Adult	0.147	0.095	0.147
Tic-tac-toe	0.371	0.003	0.382
Australian	0.071	0.066	0.067
MONK's-1	0.261	0.026	0.259
MONK's-2	0.500	0.080	0.471
Bank	0.143	0.102	0.141

4.4. Experiment 4: Calibration of Conformal Predictors

In this section, we compare the calibration of ICP, SCP and CCP. We use the same setup as in Figure 1 with $M = 3$, and with random forest with 10 trees as an underlying ML algorithm. We also train three fold CCP using RF with 10 trees. We observe that the individual small ICPs as well as the SCP are always valid, however, CCP is not valid, as an example see Figure 2 for Spambase dataset where we have used random forest with 10 trees for training.



(a) ICP (b) SCP (c) CCP

Figure 2: Calibration plots of ICP, SCP and CCP for Spambase dataset using random forest with 10 trees for training all the underlying models

5. Discussions

The purpose of this study was to introduce SCP and explore its performance in different practical settings; for partitioned data and for valid ensemble predictions with confidence.

In Experiment 1, we showed that when using partitioned data, SCP is capable of combining the models and in all cases obtain improved efficiency than each individual small ICP (see Table 2). We also observed that SCP has comparable efficiency with an ICP trained on the entire dataset. These results open up for the possibility to instead of running long inductive training on a large dataset, to use the SCP method on partitioned data and benefit from reduced modeling time while still getting accurate and valid predictions. Another possibility is to develop new and efficient parallel implementations for predictions, such as in locality-aware Big Data framework.

In Experiment 2, we showed again that SCP has improved efficiency when comparing with ICPs trained on individual partitions, but here we also combined multiple modeling methods. This has implications in settings where models are combined without requiring the same modeling method. Another important implication is that SCP would be very useful in a federated setting, where data is located in different locations, and the data cannot be pooled due to privacy or regulatory reasons, or for practical reasons such as long data transfer times. The SCP method allows for efficiently combining such federated data sources, under the condition that a separate calibration set is available to the aggregation process.

In Experiment 3, it is shown that on unpartioned data SCP is in all cases better than an ICP trained on the entire data. We also observed that SCP is comparable with CCP. The fact that SCP produces valid predictions (in contrast to CCP) makes it an appealing alternative to both ICP and CCP.

In Experiment 4, we showed that individual small ICPs and SCP are well calibrated (close to the bisector of the first quadrant) even with the poor predictive models (random forest with 10 trees), however, CCP is very poorly calibrated, see Figure 2.

Finally, we point out the main drawback of SCP, it requires a shared calibration set across individual partitions (in the federated setting it can be a problem).

6. Conclusions and Future Directions

We presented Synergy Conformal Prediction (SCP), a new ensemble learning method that produces valid prediction intervals for new objects under mild conditions. We demonstrated that the method makes it possible to partition the data and aggregate the resulting predictions, improving the modeling time while retaining good accuracy. We also demonstrated that SCP is a viable approach to commonly used ICP and CCP approaches, making it widely applicable as a valid confidence predictor. Due to its low computational complexity and parallel scalability, we believe the method will be potentially useful for Big-data problems, and due to its isolated and non-sharing data sources and trained models it can have a potential application in federated settings. Future directions when working on partitioned data include (i) studying the effect of the number and size of data partitions, as well as the overlapping partitions (ii) trying with different monotone (non) conformity scores (using different underlying ML algorithms) with individual partitions.

Acknowledgements

We would like to acknowledge Akshay Chaturvedi for useful discussion. This project received financial support from the Swedish Foundation for Strategic Research (SSF) as part of the HASTE project under the call 'Big Data and Computational Science'. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project SNIC 2017/7-273.

References

Balasubramanian, V. N., Chakraborty, S., Panchanathan, S., 2015. Conformal predictions for information fusion. Annals of Mathematics and Artificial Intelligence 74 (1-2), 45–65.

Carlsson, L., Eklund, M., Norinder, U., 2014a. Aggregated conformal prediction. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, pp. 231–240.

Carlsson, L., Eklund, M., Norinder, U., 2014b. Aggregated conformal prediction. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H., Sioutas, S., Makris, C. (Eds.), Artificial Intelligence Applications and Innovations. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 231–240.

Lichman, M., et al., 2013. Uci machine learning repository.

Linusson, H., Norinder, U., Boström, H., Johansson, U., Löfström, T., 13–16 Jun 2017. On the calibration of aggregated conformal predictors. In: Gammerman, A., Vovk, V., Luo, Z., Papadopoulos, H. (Eds.), Proceedings of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications. Vol. 60 of Proceedings of Machine Learning Research. PMLR, Stockholm, Sweden, pp. 154–173.
URL http://proceedings.mlr.press/v60/linusson17a.html

Löfström, T., Johansson, U., Boström, H., 2013. Effective utilization of data in inductive conformal prediction. In: International Joint Conference on Neural Networks, Dallas, TX, USA, August 4-9, 2013. IEEE.

Papadopoulos, H., 2008. Inductive conformal prediction: Theory and application to neural networks. In: Tools in artificial intelligence. InTech.

Toccaceli, P., Gammerman, A., 2018. Combination of inductive mondrian conformal predictors. Machine Learning, 1–22.

Vapnik, V., Izmailov, R., 2016. Synergy of monotonic rules. The Journal of Machine Learning Research 17 (1), 4722–4754.

Vovk, V., 2013. Transductive conformal predictors. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, pp. 348–360.

Vovk, V., 2015. Cross-conformal predictors. Annals of Mathematics and Artificial Intelligence 74 (1-2), 9–28.

Vovk, V., Fedorova, V., Nouretdinov, I., Gammerman, A., 2016. Criteria of efficiency for conformal prediction. In: Symposium on Conformal and Probabilistic Prediction with Applications. Springer, pp. 23–39.

Vovk, V., Gammerman, A., Shafer, G., 2005. Algorithmic learning in a random world. Springer Science & Business Media.