Towards Automated Semi-Supervised Learning

AAAI 2019

Abstract

- Automated Machine Learning (AutoML) aims to build an appropriate machine learning model for any unseen dataset automatically, i.e., without human intervention.
- Great efforts have been devoted on AutoML while they typically focus on supervised learning.
- In many applications, however, semi-supervised learning (SSL) are widespread and current AutoML systems could not well address SSL problems.
- This paper proposes to present an automated learning system for SSL (AUTO-SSL).

Introduction

- In traditional machine learning, given a dataset, a fine-tuned learning model is built by human.
- Existing approaches based on manually fine-tuned learning models consume a large amount of human resources and efforts.
- To overcome this issue, the development of automated machine learning (AutoML), which attempts to build an appropriate machine learning model for unseen dataset in an automatic manner (without human intervention), has received increasing attention recently.
- AutoML is magnificent yet challenging, since absolute AutoML is infeasible.

Introduction

- Previous work on AutoML typically focuses on supervised learning problems, and addresses the difficulties including feature engineering, model selection and hyperparameter optimization.
- A couple of systematical schemes achieve promising performance:
  - AUTO-WEKA: combine the machine learning framework WEKA with a bayesian optimization method to select a good instantiation of WEKA for a new dataset.
  - AUTO- SKLEARN: improve AUTO-WEKA and use meta-learning to warmstart the bayesian optimization procedure, and include an automated ensemble construction step.
  - Cloud AutoML: a suite of machine learning products that can automatically train high quality models by leveraging google's state-of-the-art transfer learning techniques and neural architecture search techniques.

Introduction

- Except for supervised learning scenario, semi-supervised learning (SSL) is widespread in reality. However, the efforts on Automated SSL remain limited.
- The existing AutoML techniques could not directly be applied for the automated SSL problem, since SSL introduces some new challenges:
  - many meta-features used in meta learning are no longer available and suitable.
  - the use of auxiliary unlabeled instances may sometimes even be outperformed by direct supervised learning with only limited labeled examples.

Contribution

- This paper proposes to present an automated learning system (AUTO-SSL) for SSL:
  - use meta-learning to quickly suggest some instantiations of the SSL techniques that are likely to perform quite well. (meta-features with unlabeled data distribution)
  - propose a large margin separation method to fine-tune the hyperparameters and meanwhile alleviate the performance deterioration issue in SSL.
- Extensive empirical results on 40 datasets over 200 cases demonstrate that the proposal achieves highly competitive or better performance compared to the state-of-the-art AutoML system AUTO-SKLEARN and classical SSL techniques.

Related Work

- AutoML (supervised learning)
  - automated feature engineering, automated model selection and automated hyperparameter optimization.
- Safe SSL
  - Safeness is one important aspect to AutoSSL, since it is not desirable to have an automated yet performance-degenerated SSL system.

Towards Automated SSL

- An AutoML system consist of four procedures:
  - Given a collection of datasets, an AutoML system performs meta-learning which extracts meta-features of datasets and then uses a supervised learning model to select a learning algorithm which is likely to perform well for unseen dataset.
  - Then, the AutoML system performs hyperparameter optimization, to derive a good candidate hyperparameter for the selected algorithm.
  - Later, model evaluation is conducted to finalize the ultimate model.
- In the prediction phrase, given a new dataset, the AutoML system first represents the dataset via meta-features, and then predicts an appropriate algorithm with a good hyperparameter, which finally finalizes the learning model.
- Two major difficulties:
  - how to design appropriate SSL meta-features to facilitate a better meta-learning.
  - how to choose a good quality parameter and alleviate the performance de-generation issue in SSL.

Preliminaries and Problem Definition

- **SSL dataset:** $D_l^T = (x_i^T, y_i^T)|_{i=1}^{n_l}$ and $D_u^T = (x_i^T)|_{i=n_l+1}^{n_u+n_l}$.

Definition 1 (AutoSSL) *Let $\mathcal{S} = \{S^1, \cdots, S^N\}$ be a set of SSL algorithms, and the hyper-parameters of each algorithm S^j have a domain Θ^j. Let A be a baseline supervised learning algorithm, and the hyper-parameters of algorithm A have domain Λ. Suppose that M^{auto} is the output model of the automated SSL system on data set $\mathcal{D}$ and A_Λ the model of supervised learning algorithm trained on labeled data set $\mathcal{L}$. The goal of the automated SSL system is that $Per(M^{auto})$ is always significantly better than $Per(A_\Lambda)$, and rarely worse than $Per(A_\Lambda)$, where $Per(M)$ denotes the performance of model M on testing data.*

Meta-Learning with Enhanced Meta-Features

- Meta-learning aims to reason about the performance of learning algorithms across different datasets. Specifically, in AutoML, we collect the performance data and a set of meta-features for a large number of datasets, where meta-features are characteristics of the dataset that help determine which algorithm to use for a new dataset and can be computed in an efficient manner.
- Meta-feature is central to meta-learning. However, there is a lack of a principle way to design appropriate meta-features for meta-learning.
- This work proposes to characterize the distribution of unlabeled data or data distribution assumption, which is known as an important factor for SSL, by unsupervised clustering algorithms. Specifically, different SSL techniques prefer to different data distributions.

Meta-Learning with Enhanced Meta-Features

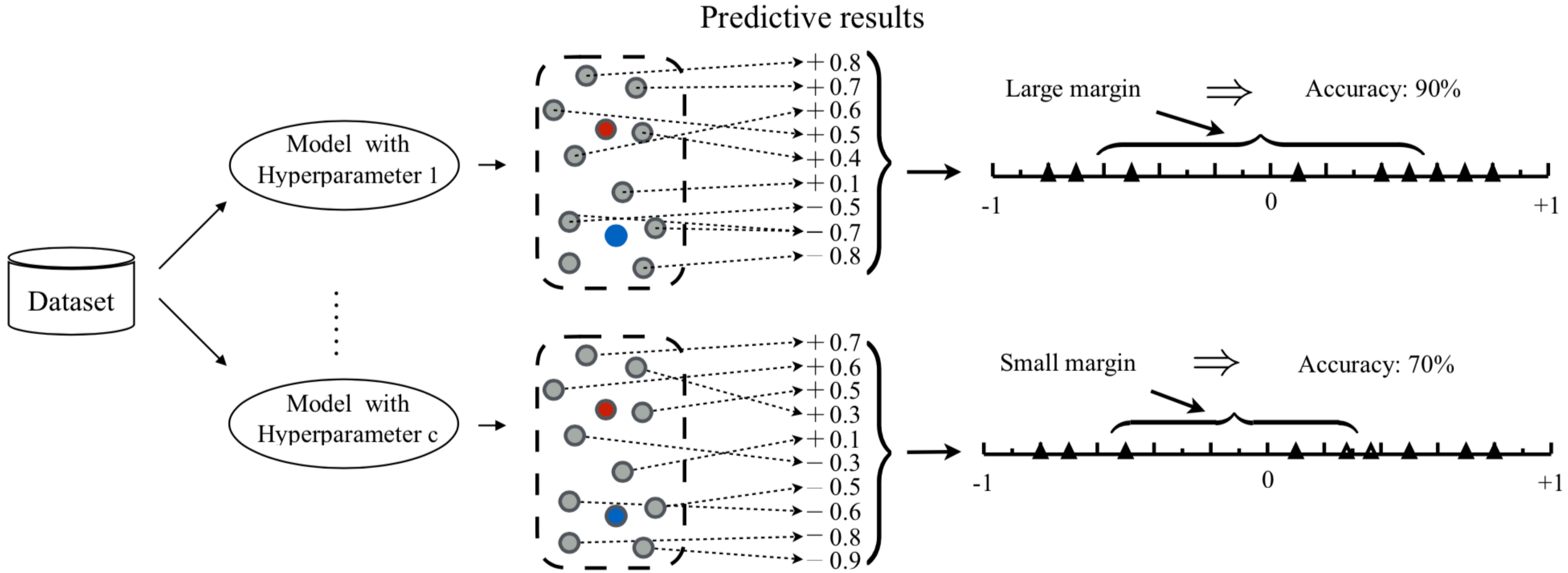
Traditional Meta-Features	
Simple Meta-Features:	**Statistic:**
number of instances	kurtosis min
log number of instances	kurtosis max
number of features	kurtosis mean
log number of features	kurtosis std
dataset dimensionality	skewness min
log dataset dimensionality	skewness max
inverse dataset dimensionality	skewness mean
log inverse dataset dimensionality	skewness std
class probability min	**PCA Statistic:**
class probability max	pca 95%
class probability mean	pca skewness first pc
class probability std	pca kurtosis first pc
Meta-Features with *Unsupervised Clustering*	
Algorithms:	**Meta-Features:**
K-Means	Intra-cluster cohesion
SpectralClustering	Inter-cluster separation
AgglomerativeClustering	Davies-Bouldin Index
	Dunn Validity Index

Large Margin Hyperparameter Selection

- Meta-learning is good at suggesting some instantiations of the SSL techniques which are likely to perform well in a quick manner, whereas it could not provide fine-grained performance.
- In contrast, hyperparameter optimization is good at fine-tuning performance over time, but it is much slower.
- The above two steps are complementary and help each other.
- Two problems:
  - it is not available to fine tune hyperparameters in SSL since the labeled examples are often too few to afford a reliable model selection.
  - the performance may be even degenerated compared to direct supervised learning with only labeled examples.

Large Margin Hyperparameter Selection

- This paper proposes to present a large margin separation method for hyperparameter optimization in AutoSSL.
- The basic idea is that once a certain hyperparameter owns a high quality, its predictive results on the unlabeled data may have a large margin separation and vice versa.



Large Margin Hyperparameter Selection

- Large margin separation is to select the high-quality hyperparameter of SSL model, such that the margin of the predictive results is maximized:

$$\theta^\star \in \arg\max_{\theta_k \in \Theta} \left| \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} f^{\theta_k}(\boldsymbol{x_i}) - \frac{1}{|\mathcal{N}|} \sum_{j \in \mathcal{N}} f^{\theta_k}(\boldsymbol{x_j}) \right|$$

- AUTO-SSL selects model that minimizes the empirical loss on the dataset:

$$M^\star \in \arg\min_{M \in S_\theta^* \cup A_\lambda, \lambda \in \Lambda} \frac{1}{K} \sum_{i=1}^{K} \mathcal{L}(M, D_{train}^i, D_{valid}^i)$$

Experiments and Results

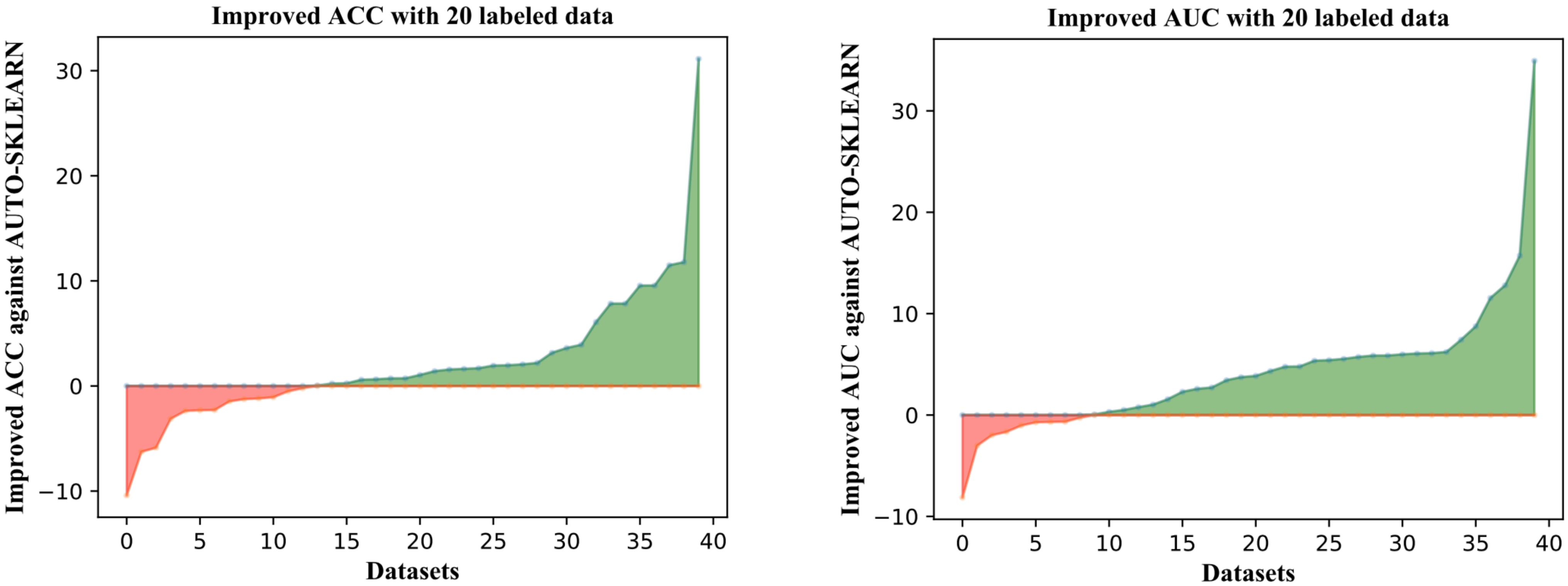


Figure 3: Improved performance of AUTO-SSL against AUTO-SKLEARN

Experiments and Results

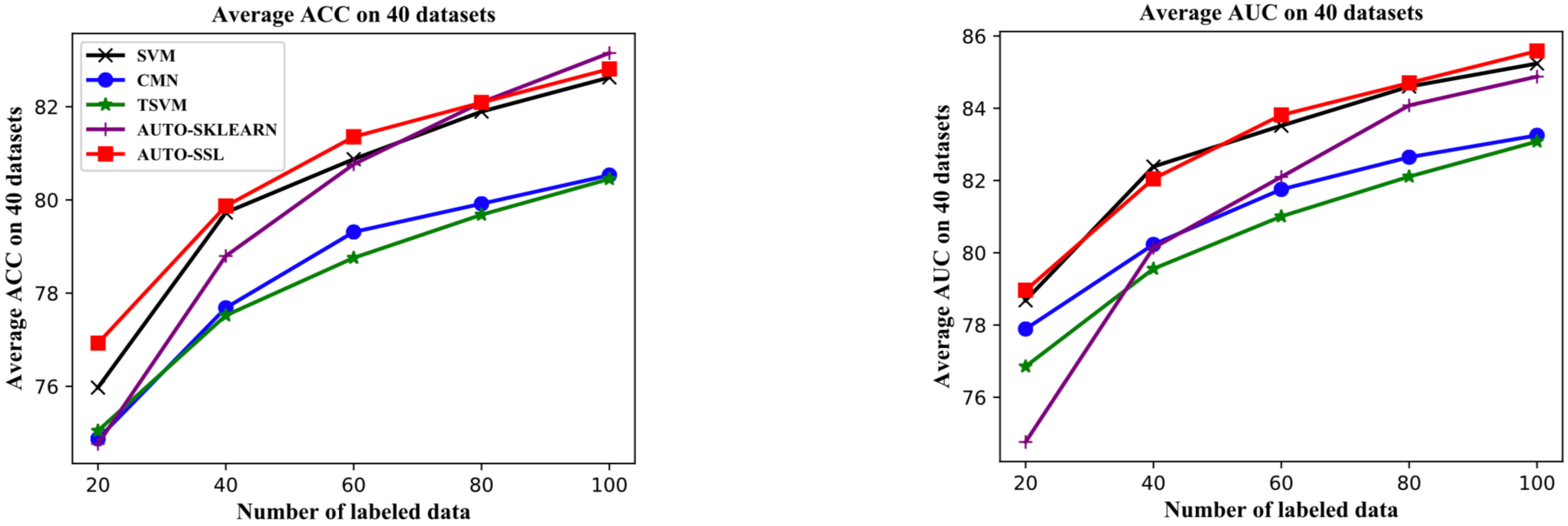


Figure 5: Average performance with different numbers of labeled insatnces on 40 datasets

Experiments and Results

	Method	Number of labeled instances 20	40	60	80	100
ACC	TSVM	12/15/13	13/11/16	13/10/17	12/11/17	13/11/16
	CMN	10/16/14	13/9/18	13/9/18	11/10/19	12/9/19
	ASSL	**11/25/4**	**13/22/5**	**13/24/3**	**10/29/1**	**11/26/3**
AUC	TSVM	10/14/16	11/13/16	9/19/12	12/13/15	12/16/12
	CMN	9/14/17	9/11/20	9/11/20	8/12/20	10/10/20
	ASSL	**8/24/8**	**8/27/5**	**9/28/3**	**10/26/4**	**9/28/3**

Conclusions

- This paper presents an automated SSL system (AUTO-SSL):
  - they first consider meta-learning that transforms automated SSL as a supervised learning and then exact appropriate features for data sets by not only traditional meta-features but also unsupervised learning.
  - To alleviate performance deterioration, they design a large margin principle to avoid low-quality hyperparameters, and save considerable computation overhead compared to direct cross-validation.
- Extensive empirical results show that the proposal outperforms classical SSL techniques and state-of-the-art AutoML system AUTO-SKLEARN, in addition clearly improves the reliability of SSL.