南京大学
人工智能学院
SCHOOL OF ARTIFICIAL INTELLIGENCE, NANJING UNIVERSITY

LAMDA
Learning And Mining from DatA

在线集成：非稳态在线学习的理论框架

赵 鹏

机器学习与数据挖掘研究所

南京大学人工智能学院

zhaop@lamda.nju.edu.cn

Outline

- Background
- Problem Setup
- Online Ensemble
- Conclusion

Outline

- Background

- Problem Setup

- Online Ensemble

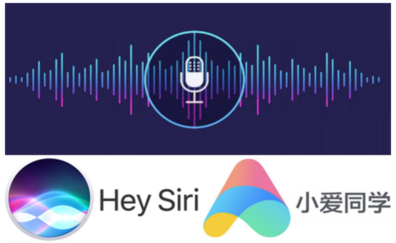
- Conclusion

Machine Learning

- Machine Learning has achieved great success in recent years.

image recognition

search engine



voice assistant

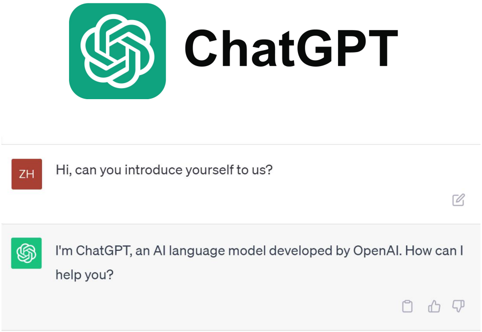
recommendation

nature
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO

AlphaGo Games

automatic driving

medical diagnosis



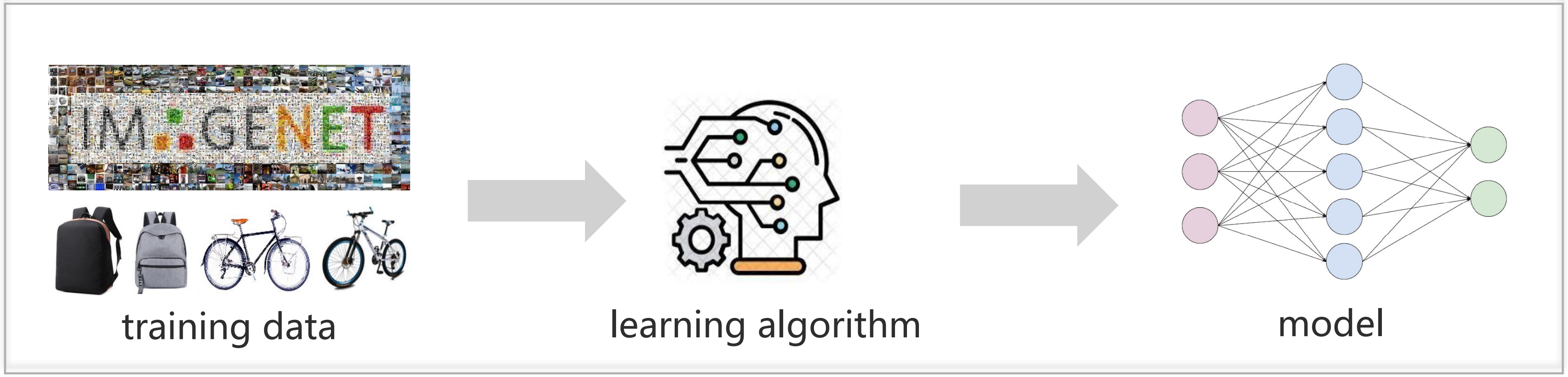
ChatGPT

ZH Hi, can you introduce yourself to us?

I'm ChatGPT, an AI language model developed by OpenAI. How can I help you?

large language model

Machine Learning



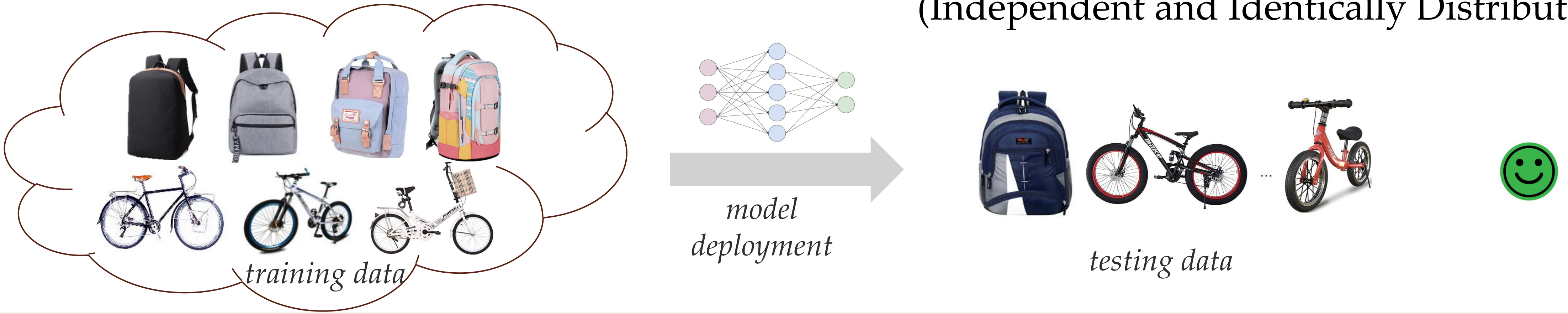
training data

learning algorithm

model

- The theoretical foundation for ML to work well: **I.I.D. assumption**

(Independent and Identically Distributed)

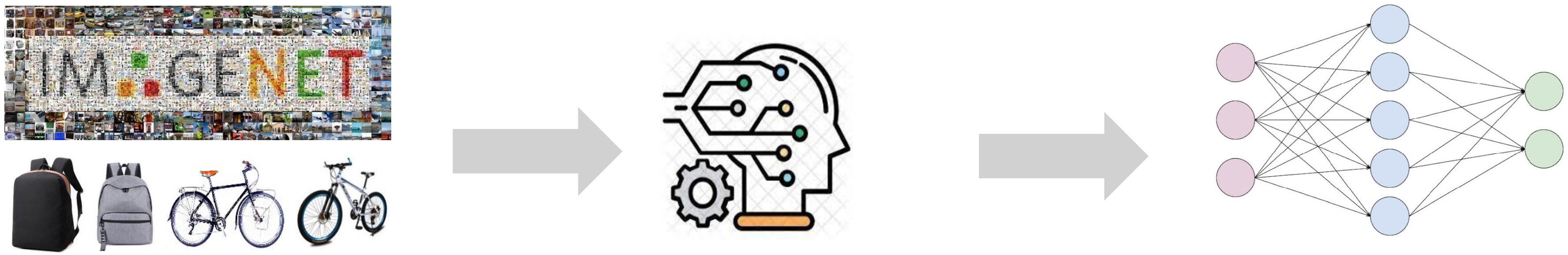


training data

model deployment

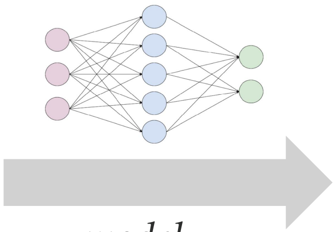
testing data

Machine Learning


training data
learning algorithm
model

- The theoretical foundation for ML to work well: **I.I.D. assumption**
(Independent and Identically Distributed)

training data

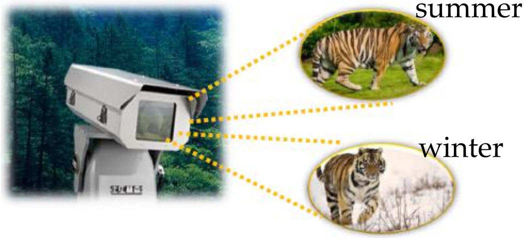

model deployment

testing data ***in practical scenario***

Open-environment Machine Learning

- **Distribution shift**: data are usually collected in open environments

species monitoring



urban computing

route planning

- In many applications, data are coming in an online fashion, like a "*stream*"

→

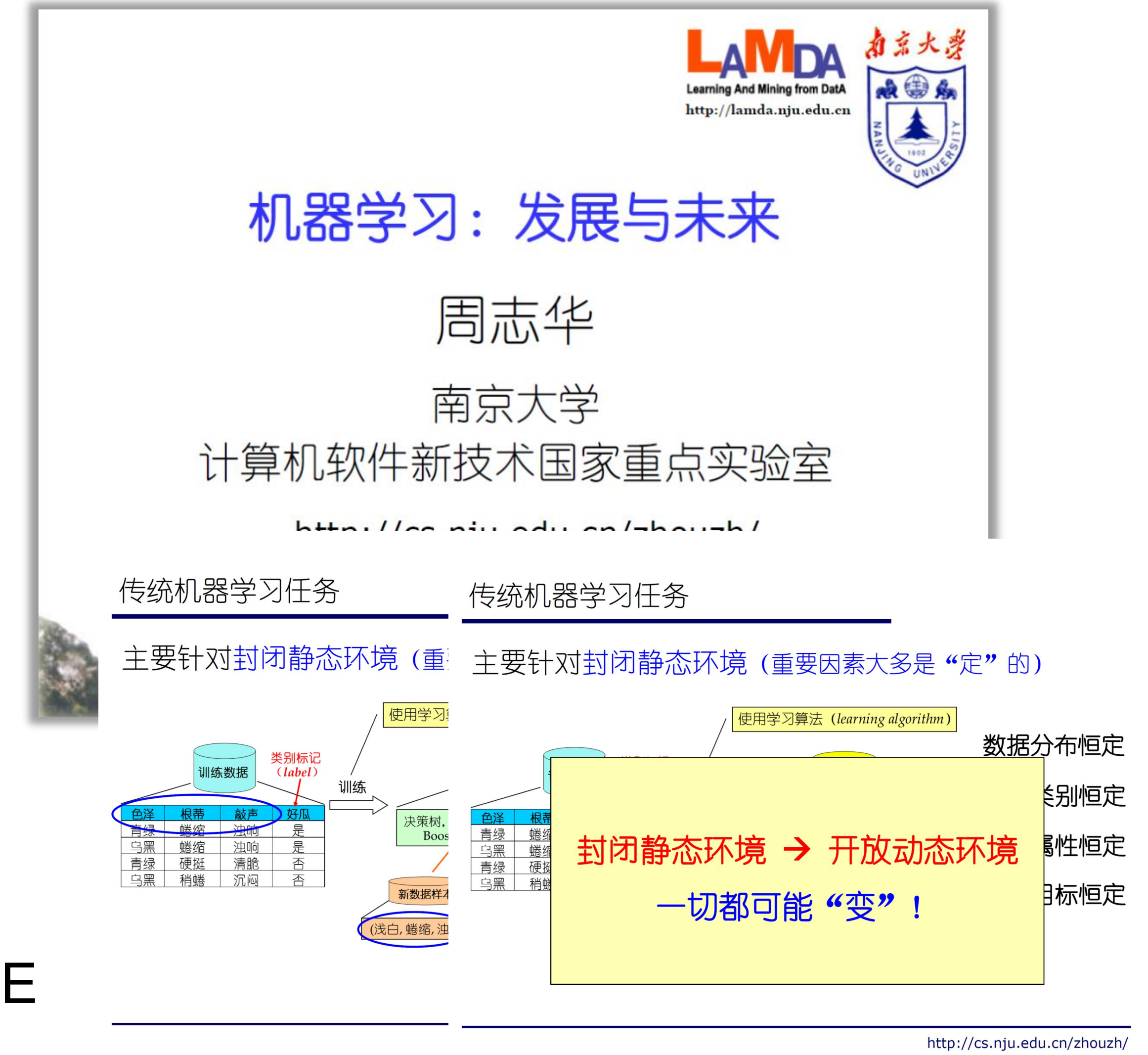
provably robust methods for non-stationary online learning

Community Discussions

Learning And Mining from DatA

“机器学习：发展与未来”

2016年中国计算机大会 特邀报告



Learning And Mining from DatA
http://lamda.nju.edu.cn

机器学习：发展与未来

周志华

南京大学
计算机软件新技术国家重点实验室

http://cs.nju.edu.cn/zhouzh/

传统机器学习任务

主要针对封闭静态环境（重

色泽	根蒂	敲声	好瓜
青绿	蜷缩	浊响	是
乌黑	蜷缩	浊响	是
青绿	硬挺	清脆	否
乌黑	稍蜷	沉闷	否

传统机器学习任务

主要针对封闭静态环境（重要因素大多是“定”的）

数据分布恒定

类别恒定

属性恒定

目标恒定

封闭静态环境 → 开放动态环境

一切都可能“变”！

http://cs.nju.edu.cn/zhouzh/

Zhi-Hua Zhou

Nanjing University

IJCAI President

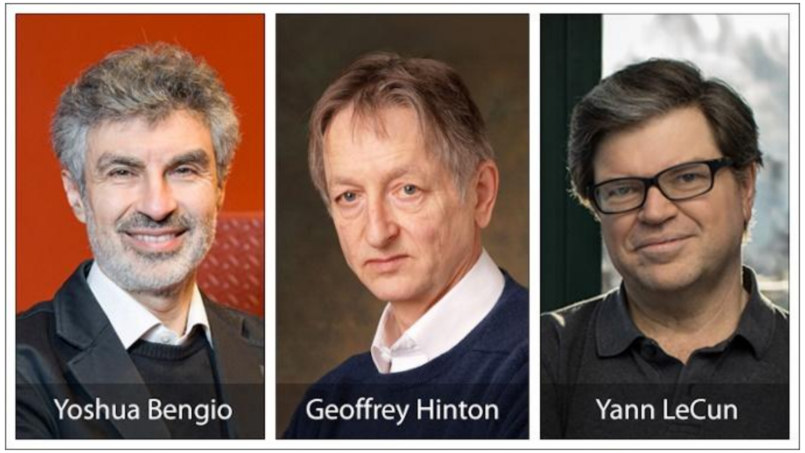
Fellow of AAAI/ACM/IEEE

Community Discussions

“Deep Learning for AI”

Communication of ACM

July, 2021. Vol 64. No 7.



2018 Turing Award Recipients

turing lecture

DOI:10.1145/3448250

How can neural networks learn the rich internal representations required for difficult tasks such as recognizing objects or understanding language?

BY YOSHUA BENGIO, YANN LECUN, AND GEOFFREY HINTON

Deep Learning for AI

TURING LECTURE

Yoshua Bengio, Yann LeCun, and Geoffrey Hinton are recipients of the 2018 ACM A.M. Turing Award for breakthroughs that have made deep neural networks a critical component of computing.

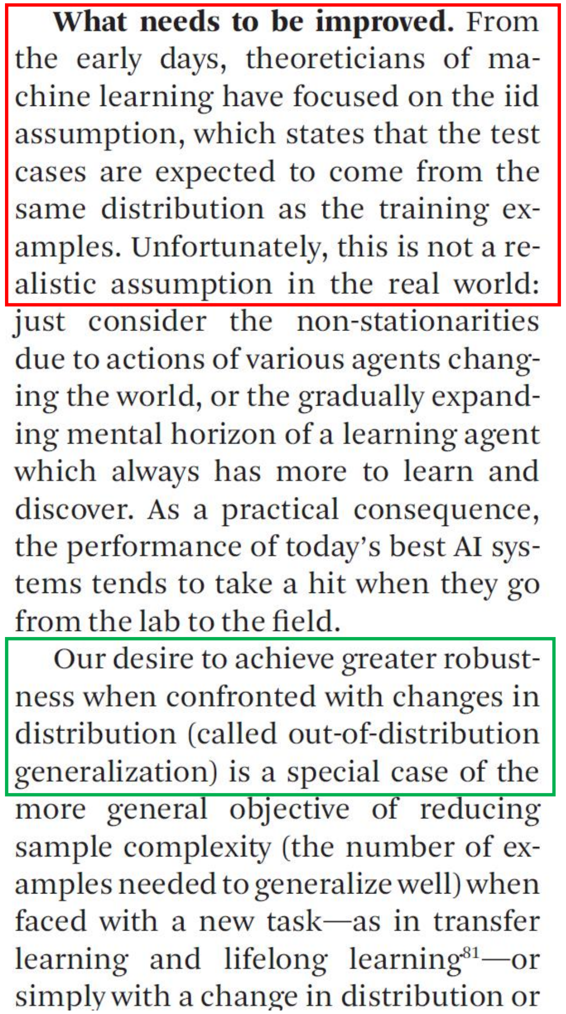
RESEARCH ON ARTIFICIAL neural networks was motivated by the observation that human intelligence emerges from highly parallel networks of relatively simple, non-linear neurons that learn by adjusting the strengths of their connections. This observation leads to a central computational question: How is it possible for networks of this general kind to learn the complicated internal representations that are required for difficult tasks such as recognizing objects or understanding language? Deep learning seeks to answer this question by using many layers of activity vectors as representations and learning the connection strengths that give rise to these vectors by following the stochastic gradient of an objective function that measures how well the network is performing. It is very surprising that such a conceptually simple approach has proved to be so effective when applied to large training sets using huge amounts of computation and it appears that a key ingredient is depth: shallow networks simply do not work as well.

We reviewed the basic concepts and some of the breakthrough achievements of deep learning several years ago.[63] Here we briefly describe the origins of deep learning, describe a few of the more recent advances, and discuss some of the future challenges. These challenges include learning with little or no external supervision, coping with test examples that come from a different distribution than the training examples, and using the deep learning approach for tasks that humans solve by using a deliberate sequence of steps which we attend to consciously—tasks that Kahneman[56] calls *system 2* tasks as opposed to *system 1* tasks like object recognition or immediate natural language understanding, which generally feel effortless.

From Hand-Coded Symbolic Expressions to Learned Distributed Representations

There are two quite different paradigms for AI. Put simply, the logic-inspired paradigm views sequential reasoning as the essence of intelligence and aims to implement reasoning in computers using hand-designed rules of inference that operate on hand-designed symbolic expressions that formalize knowledge. The brain-inspired paradigm views learning representations from data as the essence of intelligence and aims to implement learning by hand-designing or evolving rules for modifying the connec-

58 COMMUNICATIONS OF THE ACM | JULY 2021 | VOL. 64 | NO. 7



What needs to be improved. From the early days, theoreticians of machine learning have focused on the iid assumption, which states that the test cases are expected to come from the same distribution as the training examples. Unfortunately, this is not a realistic assumption in the real world: just consider the non-stationarities due to actions of various agents changing the world, or the gradually expanding mental horizon of a learning agent which always has more to learn and discover. As a practical consequence, the performance of today’s best AI systems tends to take a hit when they go from the lab to the field.

Our desire to achieve greater robustness when confronted with changes in distribution (called out-of-distribution generalization) is a special case of the more general objective of reducing sample complexity (the number of examples needed to generalize well) when faced with a new task—as in transfer learning and lifelong learning[81]—or simply with a change in distribution or

Outline

- Background
- Problem Setup
- Online Ensemble
- Conclusion

Online Learning

- View online learning as a game between *learner* and *environment*.

Online Convex Optimization

At each round $t = 1, 2 \cdots, T$

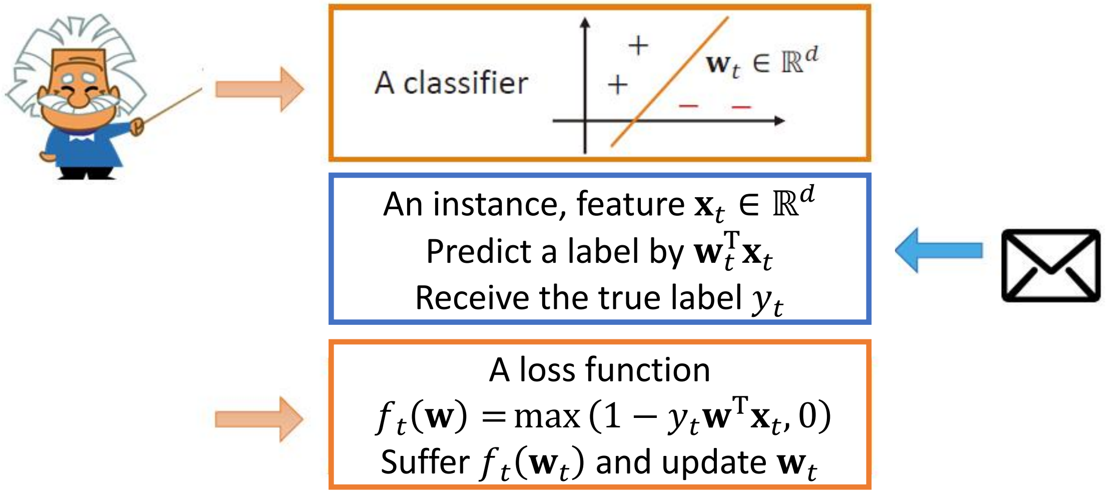
1. learner first provides a model $\mathbf{w}_t \in \mathcal{W}$;
2. and simutaneously the environment picks a convex $\boxed{\text{online function } f_t : \mathcal{W} \mapsto [0,1]}$;
3. the learner then suffers loss $f_t(\mathbf{w}_t)$ and observes some information of f_t.

Example: online function $f_t : \mathcal{W} \mapsto \mathbb{R}$ is composition of

(i) loss $\ell : \hat{\mathcal{Y}} \times \mathcal{Y} \mapsto \mathbb{R}$, and

(ii) data item: $(\mathbf{x}_t, y_t) \in \mathcal{X} \times \mathcal{Y}$.

$$\Longrightarrow \quad f_t(\mathbf{w}) = \ell(\mathbf{w}^\top \mathbf{x}_t, y_t)$$



An instance, feature $\mathbf{x}_t \in \mathbb{R}^d$

Predict a label by $\mathbf{w}_t^{\mathrm{T}}\mathbf{x}_t$

Receive the true label y_t

A loss function

$f_t(\mathbf{w}) = \max(1 - y_t \mathbf{w}^{\mathrm{T}}\mathbf{x}_t, 0)$

Suffer $f_t(\mathbf{w}_t)$ and update $\mathbf{w}_t$

Spam Filtering

Regular vs Spam ?

Online Learning

- View online learning as a game between *learner* and *environment*.

Online Convex Optimization

At each round $t = 1, 2 \cdots, T$

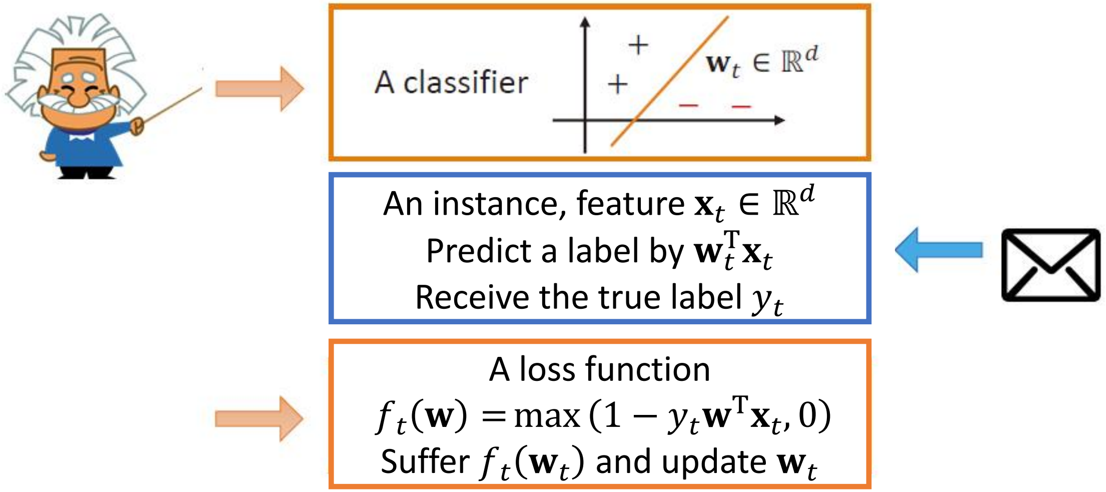
1. learner first provides a model $\mathbf{w}_t \in \mathcal{W}$;
2. and simutaneously the environment picks a convex online function $f_t : \mathcal{W} \mapsto [0, 1]$;
3. the learner then suffers loss $f_t(\mathbf{w}_t)$ and observes some information of f_t.

full information

horse racing

partial information

multi-armed bandits



A classifier

An instance, feature $\mathbf{x}_t \in \mathbb{R}^d$

Predict a label by $\mathbf{w}_t^\mathrm{T}\mathbf{x}_t$

Receive the true label y_t

A loss function

$f_t(\mathbf{w}) = \max(1 - y_t\mathbf{w}^\mathrm{T}\mathbf{x}_t, 0)$

Suffer $f_t(\mathbf{w}_t)$ and update $\mathbf{w}_t$

Spam Filtering

Regular vs Spam ?

Performance Measure

Regret: online prediction as good as the best offline model

$$\mathrm{Regret}_T \triangleq \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \boxed{\min_{\mathbf{w}\in\mathcal{W}} \sum_{t=1}^{T} f_t(\mathbf{w})}$$

cumulative loss of the best offline model

optimal model changes in non-stationary environments

Dynamic Regret

$$\text{D-Regret}(\mathbf{u}_1, \cdots, \mathbf{u}_T) \triangleq \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)$$

allow changing comparators

The comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T$ essentially depict the underlying (unknown) distributions of all rounds.

- stationary environments: $\mathbf{u}_t = \mathbf{w}_* \in \arg\min_{\mathbf{w}\in\mathcal{W}} \sum_{t=1}^{T} f_t(\mathbf{w})$
- piecewise-stationary environments: $\mathbf{u}_t = \mathbf{w}_*^{\mathcal{I}_k}$ for a stationary interval $t \in \mathcal{I}_k$

Outline

- Background
- Problem Setup
- Online Ensemble
- Conclusion

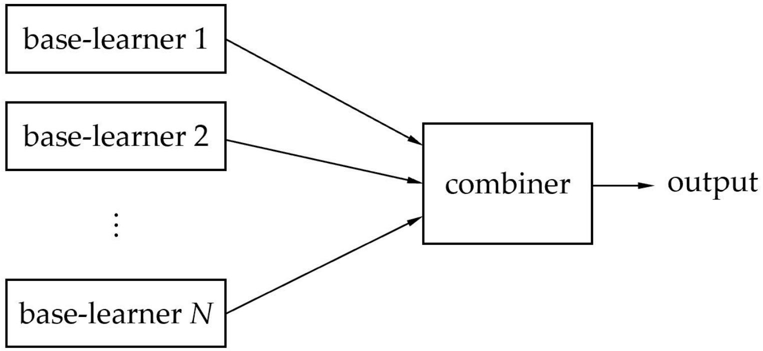
Fundamental Challenge

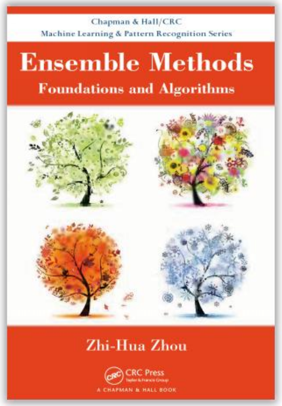
$$\text{D-Regret}(\mathbf{u}_1, \cdots, \mathbf{u}_T) = \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)$$

Key difficulty: the ***uncertainty*** due to unknown environmental changes.

Basic idea: **Ensemble Methods**

- ***Protocol***: combine multiple base learners to achieve robustness
- ***Advantage***: achieve more robust results under uncertain or even changing environments





Chapman & Hall/CRC
Machine Learning & Pattern Recognition Series
Ensemble Methods
Foundations and Algorithms
Zhi-Hua Zhou
CRC Press

Zhi-Hua Zhou. Ensemble Methods: Foundations and Algorithms. Chapman & Hall/CRC, Jun. 2012.

Online Ensemble (在线集成)

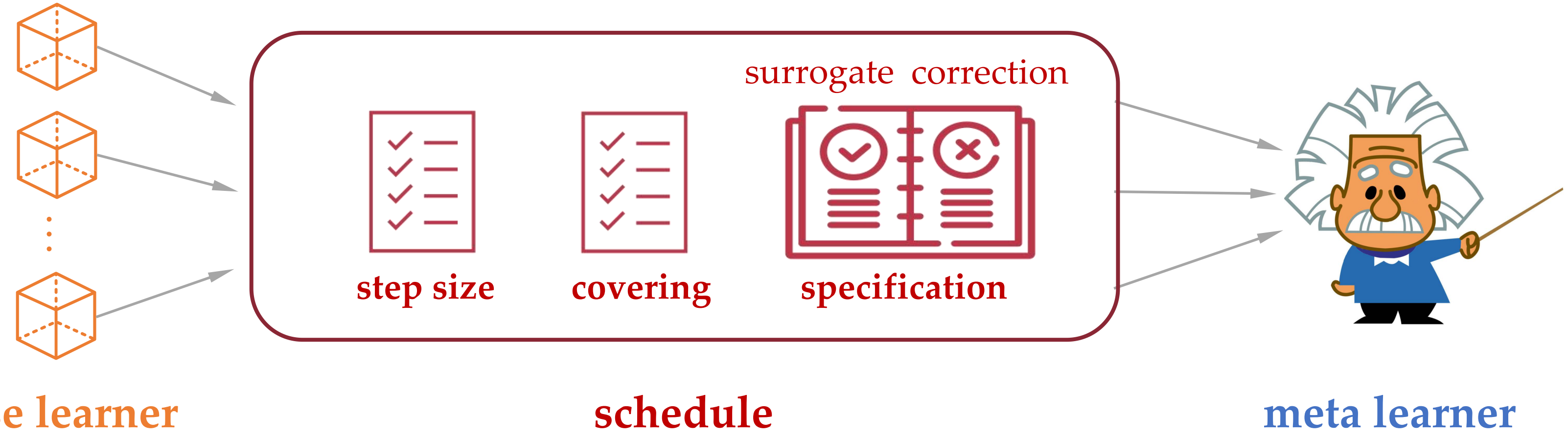
Learning And Mining from DatA

Basic Components

(1) **base learner**: an online learner to cope with a certain amount of non-stationarity

(2) **schedule**: a set of parameters for initiating base learners that encourage diversity

(3) **meta learner**: an expert-tracking learner that can combine base learners' decisions



Deploying Online Ensemble

We will showcase that properly deploying online ensemble can effectively resolve several important online learning problems.

- Dynamic Regret of Bandit Convex Optimization

- Problem-dependent Dynamic Regret

Deploying Online Ensemble

We will showcase that properly deploying online ensemble can effectively resolve several important online learning problems.

- Dynamic Regret of Bandit Convex Optimization

- Problem-dependent Dynamic Regret

Bandit Convex Optimization (BCO)

- BCO with one-point feedback

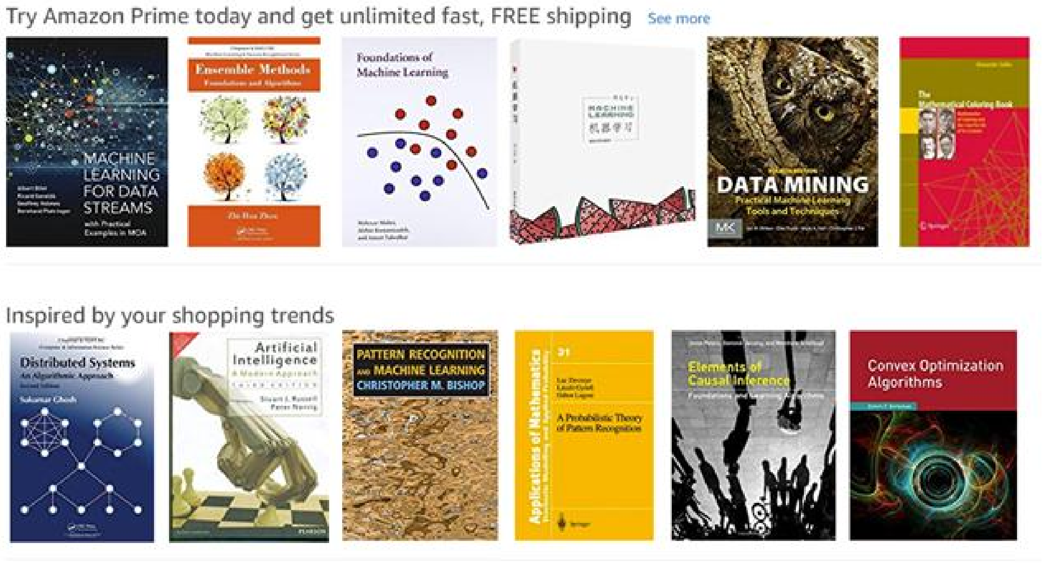
  the learner sends a single point $\mathbf{w}_t \in \mathcal{W}$, and then receives the ***function value*** $f_t(\mathbf{w}_t)$ only

  [Flaxman et al., SODA 2005; Bubeck et al., STOC 2017]

- BCO with two-point feedback

  the learner sends two points $\mathbf{w}_t^1, \mathbf{w}_t^2 \in \mathcal{W}$, and then receives their ***function values***, namely, $f_t(\mathbf{w}_t^1)$ and $f_t(\mathbf{w}_t^2)$, only

  [Agarwal et al., COLT 2010; Shamir, JMLR 2017]



Try Amazon Prime today and get unlimited fast, FREE shipping See more

Inspired by your shopping trends

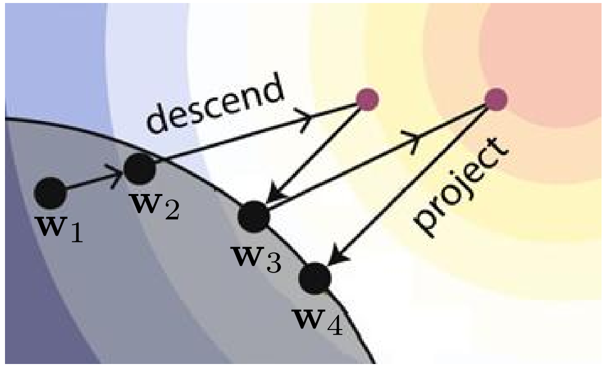
MACHINE LEARNING FOR DATA STREAMS
Ensemble Methods
Foundations of Machine Learning
DATA MINING
Distributed Systems
Artificial Intelligence
PATTERN RECOGNITION AND MACHINE LEARNING
CHRISTOPHER M. BISHOP
A Probabilistic Theory of Pattern Recognition
Elements of Causal Inference
Convex Optimization Algorithms

online recommendation

A Gentle Start

Online Gradient Descent (OGD)

```
for $t = 1$ to $T$ do
    Play model $\mathbf{w}_t$ and suffer loss $f_t(\mathbf{w}_t)$
    Update the model
        $\mathbf{w}_{t+1} = \Pi_{\mathcal{W}}[\mathbf{w}_t - \underline{\eta \nabla f_t(\mathbf{w}_t)}]$
end for
```



https://www.nature.com/articles/s41534-017-0043-1

Challenge: with only bandit feedback, the learner *cannot evaluate the gradient*

FKM estimator [Flaxman et al., SODA'05]

construct $\mathbf{w}_t$ using the perturbation technique

$$\mathbf{w}_t \triangleq \widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t$$

$\mathbf{s}_t$ is random vector sampled from ball $\mathbb{B} = \{\mathbf{v} \mid \|\mathbf{v}\| \leq 1\}$

$$\Rightarrow \quad \mathbb{E}\left[\frac{d}{\delta} f_t(\mathbf{w}_t) \cdot \mathbf{s}_t\right] = \nabla \widehat{f}_t(\widetilde{\mathbf{w}}_t)$$

[proved by Stokes equation]

with $\widehat{f}_t(\mathbf{w}) \triangleq \mathbb{E}_{\mathbf{v} \in \mathbb{B}}[f_t(\mathbf{w} + \delta \mathbf{v})]$ being smoothed function.

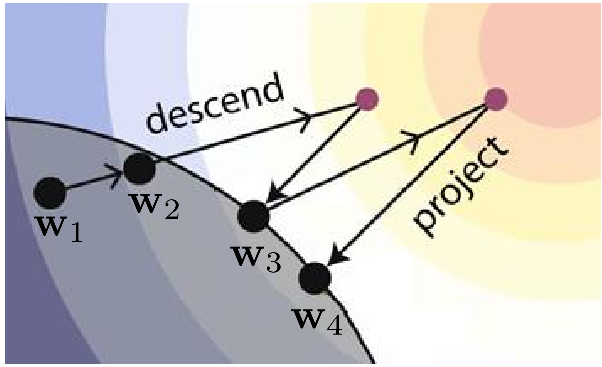
$\Rightarrow$ define $\mathbf{g}_t \triangleq \frac{d}{\delta} f_t(\widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t) \cdot \mathbf{s}_t$ as gradient estimator

A Gentle Start

Online Gradient Descent (OGD)

```
for $t = 1$ to $T$ do
    Play model $\mathbf{w}_t$ and suffer loss $f_t(\mathbf{w}_t)$
    Update the model
        $\mathbf{w}_{t+1} = \Pi_{\mathcal{W}}[\mathbf{w}_t - \underline{\eta \nabla f_t(\mathbf{w}_t)}]$
end for
```

Challenge: with only bandit feedback, the learner ***cannot evaluate the gradient***



https://www.nature.com/articles/s41534-017-0043-1

FKM estimator [Flaxman et al., SODA'05]

construct $\mathbf{w}_t$ using the perturbation technique

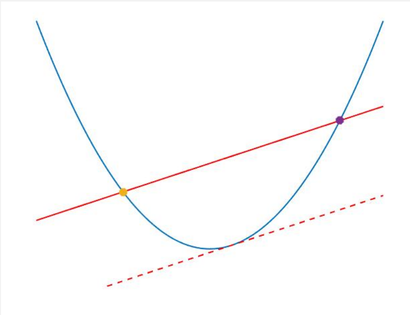
$$\mathbf{w}_t \triangleq \widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t$$

$\mathbf{s}_t$ is random vector sampled from ball $\mathbb{B} = \{\mathbf{v} \mid \|\mathbf{v}\| \leq 1\}$

Consider the 1-dim case ($d = 1$).

$$\mathbb{E}_{\mathbf{s} \in \mathbb{S}}\left[\frac{d}{\delta} f_t(\widetilde{\mathbf{w}} + \delta \mathbf{s}) \cdot \mathbf{s}\right]$$

$$= \frac{1}{2\delta} f_t(\widetilde{w} + \delta) - \frac{1}{2\delta} f_t(\widetilde{w} - \delta)$$



Base Algorithm: BGD

- Gradient estimator: $\mathbf{g}_t = \frac{d}{\delta} f_t(\widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t) \cdot \mathbf{s}_t$
- Perform Online Gradient Descent using this gradient estimator.

```
Bandit Gradient Descent (BGD)
for $t = 1$ to $T$ do
    Select a unit vector $\mathbf{s}_t$ uniformly at random
    Submit $\mathbf{w}_t = \widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t$
    Receive $f_t(\mathbf{w}_t)$ as the feedback
    Construct the gradient estimator by $\mathbf{g}_t = \frac{d}{\delta} f_t(\widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t) \cdot \mathbf{s}_t$
    $\widetilde{\mathbf{w}}_{t+1} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_t - \eta \mathbf{g}_t]$
end for
```

$\mathbb{E}[\mathbf{g}_t] = \nabla \widehat{f}_t(\widetilde{\mathbf{w}}_t)$

$\widehat{f}_t(\mathbf{w}) \triangleq \mathbb{E}_{\mathbf{v} \in \mathbb{B}}[f_t(\mathbf{w} + \delta \mathbf{v})]$

Base Algorithm: Dynamic Regret

Theorem 1. Under certain standard assumptions, for any perturbation parameter $\delta > 0$, step size $\eta > 0$, and shrinkage parameter $\alpha = \delta/r$, the expected dynamic regret of $\mathrm{BGD}(T, \delta, \alpha, \eta)$ for the one-point feedback model satisfies

$$\begin{aligned}\mathbb{E}\left[\text{D-Regret}(\mathbf{u}_1, \ldots, \mathbf{u}_T)\right] &\leq \frac{7R^2 + RP_T}{4\eta} + \frac{\eta d^2 C^2 T}{2\delta^2} + \left(3L + \frac{LR}{r}\right)\delta T \\ &= \mathcal{O}\left(\frac{1 + P_T}{\eta} + \frac{\eta T}{\delta^2} + \delta T\right),\end{aligned}$$

where $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|$ measures the non-stationarity level.

Optimal parameter setting is

- step size $\eta_* = \left(\frac{7R^2 + RP_T}{T}\right)^{\frac{3}{4}}$ $\Longrightarrow$ $\mathcal{O}\big(T^{3/4}(1 + P_T)^{1/4}\big)$
- perturbation parameter $\delta_* = \eta_*^{\frac{1}{3}}$

Base Algorithm: Dynamic Regret

Theorem 1. Under certain standard assumptions, for any perturbation parameter $\delta > 0$, step size $\eta > 0$, and shrinkage parameter $\alpha = \delta / r$, the expected dynamic regret of BGD$(T, \delta, \alpha, \eta)$ for the one-point feedback model satisfies

$$\begin{aligned}\mathbb{E}\left[\text{D-Regret}(\mathbf{u}_1, \ldots, \mathbf{u}_T)\right] &\leq \frac{7R^2 + RP_T}{4\eta} + \frac{\eta d^2 C^2 T}{2\delta^2} + \left(3L + \frac{LR}{r}\right)\delta T \\ &= \mathcal{O}\left(\frac{1 + P_T}{\eta} + \frac{\eta T}{\delta^2} + \delta T\right),\end{aligned}$$

where $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|$ measures the non-stationarity level.

Optimal parameter setting is

- step size $\eta_* = \left(\frac{7R^2 + RP_T}{T}\right)^{\frac{3}{4}}$
- perturbation parameter $\delta_* = \eta_*^{\frac{1}{3}}$

Comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T$ can be arbitrary, we cannot know non-stationarity P_T in advance, ***so how to tune the step size ?***

Online Ensemble for BCO

Deploying a proper online ensemble to deal with the issue of ***unknown non-stationarity***, so that we can ***optimally tune step size***.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i}$$

- **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items

η_1
η_2
η_3
$\cdots$
η_N

$$\mathcal{H} = \left\{ \eta_i = 2^{i-1}\frac{\sqrt{7}R}{dCT^{3/4}} \mid i = 1, \ldots, N \right\}$$

with $N = \lceil \log_2(1 + 2T/7) \rceil + 1 = \mathcal{O}(\log T)$.

- **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\text{BGD}(\eta_i): \ \widetilde{\mathbf{w}}_{t+1,i} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_{t,i} - \eta_i \mathbf{g}_t^{\eta_i}]$$

$$\mathbf{w}_{t+1,i} = \widetilde{\mathbf{w}}_{t+1,i} + \delta \boldsymbol{s}_t$$

- **Meta algorithm:** provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

increase weight on base-learners with better performance

$$\text{Hedge:} \ \ p_{t+1,i} \propto p_{t,i} \exp(-\varepsilon f_t(\mathbf{w}_{t,i}))$$

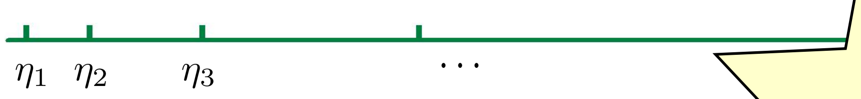
Online Ensemble for BCO

Deploying a proper online ensemble to deal with the issue of *unknown non-stationarity*, so that we can ***optimally tune step size***.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i}$$

- **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{ \eta_i = 2^{i-1}\frac{\sqrt{7}R}{dCT^{3/4}} \mid i = 1, \ldots \right\}$$

with $N = \lceil \log_2(1 + 2T/7) \rceil + 1 = \mathcal{O}(\log T)$.

- **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\text{BGD}(\eta_i):\ \widetilde{\mathbf{w}}_{t+1,i} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_{t,i} - \eta_i \mathbf{g}_t^{\eta_i}]$$

$$\mathbf{w}_{t+1,i} = \widetilde{\mathbf{w}}_{t+1,i} + \delta \boldsymbol{s}_t$$

orithm: provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

weight on base-learners with better performance

$$\text{Hedge:}\ \ p_{t+1,i} \propto p_{t,i}\exp(-\varepsilon f_t(\mathbf{w}_{t,i}))$$

bandit *feedback makes it hard to initiate* ***multiple*** *base learners*

Peng Zhao (Nanjing University)

Multiple base learners in BCO

Learning And Mining from DatA

- A closer look at dynamic regret analysis

$$\widetilde{\mathbf{w}}_{t+1} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_t - \eta \mathbf{g}_t], \ \mathbb{E}[\mathbf{g}_t] = \nabla \widehat{f}_t(\widetilde{\mathbf{w}}_t).$$

smoothed function $\widehat{f}_t(\mathbf{w}) = \mathbb{E}_{\mathbf{v}\in\mathbb{B}}[f_t(\mathbf{w}+\delta\mathbf{v})]$

rescaled comparator $\mathbf{v}_t = (1-\alpha)\mathbf{u}_t$

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)$$

$$= \underbrace{\sum_{t=1}^{T} \widehat{f}_t(\widetilde{\mathbf{w}}_t) - \sum_{t=1}^{T} \widehat{f}_t(\mathbf{v}_t)}_{\substack{\texttt{term (a)} \\ \textit{depends on } P_T}} + \underbrace{\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} \widehat{f}_t(\widetilde{\mathbf{w}}_t)}_{\substack{\texttt{term (b)} \\ \leq 2L\delta T}} + \underbrace{\sum_{t=1}^{T} \widehat{f}_t(\mathbf{v}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)}_{\substack{\texttt{term (c)} \\ \leq (L\delta + L\alpha R)T}}$$

crucial term, related to non-stationarity measure P_T

not involve the unknown non-stationarity measure P_T

(approximation error due to the perturbation operation)

Multiple base learners in BCO

- Key idea: ***surrogate optimization***

Proposition 1. For any $t \in [T]$, the following holds true:

$$\mathbb{E}[\widehat{f}_t(\widetilde{\mathbf{w}}_t) - \widehat{f}_t(\mathbf{v}_t)] \leq \mathbb{E}[\langle \mathbf{g}_t, \widetilde{\mathbf{w}}_t - \mathbf{v}_t \rangle],$$

where $\mathbf{g}_t = \frac{d}{\delta} f_t(\widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t) \cdot \mathbf{s}_t$ is the one-point gradient estimator.

- Construct the surrogate loss $\ell_t(\mathbf{w}) \triangleq \langle \mathbf{g}_t, \mathbf{w} \rangle$

  which is a linearized loss parametrized by the gradient estimator $\mathbf{g}_t$.

Feed this surrogate loss to online ensemble to maintain multiple base learners!

Surrogate Loss

- Construct the surrogate loss $\ell_t(\mathbf{w}) \triangleq \langle \mathbf{g}_t, \mathbf{w} \rangle$ and feed it to online ensemble.

Theorem 2. The constructed surrogate loss satifies the following properties:

(i) $\mathbb{E}\left[\widehat{f}_t(\widetilde{\mathbf{w}}_t) - \widehat{f}_t(\mathbf{v})\right] \le \mathbb{E}\left[\ell_t(\widetilde{\mathbf{w}}_t) - \ell_t(\mathbf{v})\right]$ holds for any $\mathbf{v} \in \mathcal{W}$.

(ii) $\nabla \ell_t(\mathbf{w}) = \mathbf{g}_t$ holds for any $\mathbf{w} \in \mathcal{W}$.

- Property (i) implies that it suffices to optimize ***dynamic regret of surrogate loss***.
- Property (ii) implies that it is feasible to ***deploy multiple base learners*** to perform BGD over the ***surrogate loss***.

All the gradients $\nabla \ell_t(\widetilde{\mathbf{w}}_t^1) = \nabla \ell_t(\widetilde{\mathbf{w}}_t^2) = \cdots = \nabla \ell_t(\widetilde{\mathbf{w}}_t^N) = \mathbf{g}_t$, so they can be obtained by querying the function value of f_t ***only once***.

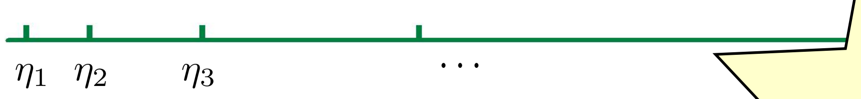
Online Ensemble for BCO

Deploying a proper online ensemble to deal with the issue of ***unknown non-stationarity***, so that we can ***optimally tune step size***.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i}$$

- **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{ \eta_i = 2^{i-1}\frac{\sqrt{7}R}{dCT^{3/4}} \mid i = 1, \ldots \right\}$$

with $N = \lceil \log_2(1 + 2T/7) \rceil + 1 = \mathcal{O}(\log T)$.

***bandit** feedback makes it hard to initiate **multiple** base learners*

- **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\text{BGD}(\eta_i)\text{:}\quad \widetilde{\mathbf{w}}_{t+1,i} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_{t,i} - \eta_i \mathbf{g}_t^{\eta_i}]$$

$$\mathbf{w}_{t+1,i} = \widetilde{\mathbf{w}}_{t+1,i} + \delta \boldsymbol{s}_t$$

orithm: provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

weight on base-learners with better performance

$$\text{Hedge:}\quad p_{t+1,i} \propto p_{t,i}\exp(-\varepsilon f_t(\mathbf{w}_{t,i}))$$

Online Ensemble for BCO

Deploying a proper online ensemble to deal with the issue of *unknown non-stationarity*, so that we can ***optimally tune step size***.

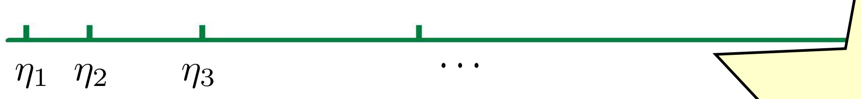
$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i} \mathbf{w}_{t+1,i}$$

$$\mathbf{g}_t = \frac{d}{\delta} f_t(\widetilde{\mathbf{w}}_t + \delta \mathbf{s}_t) \cdot \mathbf{s}_t$$

$$\ell_t(\mathbf{w}) = \langle \mathbf{g}_t, \mathbf{w} \rangle$$

- **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{ \eta_i = 2^{i-1} \frac{\sqrt{7}R}{dCT^{3/4}} \mid i = 1, \ldots \right.$$

with $N = \lceil \log_2(1 + 2T/7) \rceil + 1 = \mathcal{O}(\log T)$.

- **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\text{BGD}(\eta_i): \quad \widetilde{\mathbf{w}}_{t+1,i} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_{t,i} - \eta_i \mathbf{g}_t]$$

$$\mathbf{w}_{t+1,i} = \widetilde{\mathbf{w}}_{t+1,i} + \delta \boldsymbol{s}_t$$

- **...orithm:** provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

...weight on base-learners with better performance

$$\text{Hedge:} \quad p_{t+1,i} \propto p_{t,i} \exp(-\varepsilon \ell_t(\mathbf{w}_{t,i}))$$

surogate loss

makes online ensemble possible in bandit!

Dynamic Regret

Theorem 3. *Under certain standard assumptions, with a proper setting of the pool of candidate step sizes $\mathcal{H}$ and the learning rate ϵ for the meta-algorithm, our* PBGD *algorithm enjoys the following expected dynamic regret guarantees.*

- *For the one-point feedback model,* $\mathbb{E}[\text{D-Regret}_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)] \leq \mathcal{O}(T^{\frac{3}{4}}(1 + P_T)^{\frac{1}{2}})$.
- *For the two-point feedback model,* $\mathbb{E}[\text{D-Regret}_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)] \leq \mathcal{O}(T^{\frac{1}{2}}(1 + P_T)^{\frac{1}{2}})$.

We further establish the lower bound to demonstrate the hardness of the problem: an $\Omega(\sqrt{TP_T})$ regret is unavoidable for bandit feedback models.

⇨ *Our algorithm is minimax optimal for two-point BCO model; while it remains open how to close the gap in one-point BCO.*

Online Ensemble for BCO

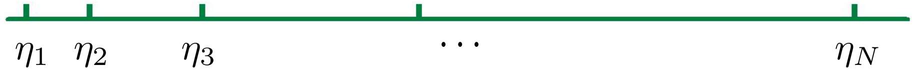
Deploying a proper online ensemble to deal with the issue of ***unknown non-stationarity***, so that we can ***optimally tune step size***.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i} \mathbf{w}_{t+1,i}$$

*Proper **surrogate loss** is essential for deploying online ensemble to bandit online problems.*

- **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{ \eta_i = 2^{i-1} \frac{\sqrt{7}R}{dCT^{3/4}} \mid i = 1, \ldots, N \right\}$$

with $N = \lceil \log_2(1 + 2T/7) \rceil + 1 = \mathcal{O}(\log T)$.

- **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\text{BGD}(\eta_i): \ \widetilde{\mathbf{w}}_{t+1,i} = \Pi_{(1-\alpha)\mathcal{W}}[\widetilde{\mathbf{w}}_{t,i} - \eta_i \mathbf{g}_t]$$

$$\mathbf{w}_{t+1,i} = \widetilde{\mathbf{w}}_{t+1,i} + \delta \boldsymbol{s}_t$$

- **Meta algorithm:** provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

increase weight on base-learners with better performance

$$\text{Hedge:} \ \ p_{t+1,i} \propto p_{t,i} \exp(-\varepsilon \ell_t(\mathbf{w}_{t,i}))$$

Deploying Online Ensemble

We will showcase that properly deploying online ensemble can effectively resolve several important online learning problem.

- Dynamic Regret of Bandit Convex Optimization

- Problem-dependent Dynamic Regret

Beyond the worst-case analysis

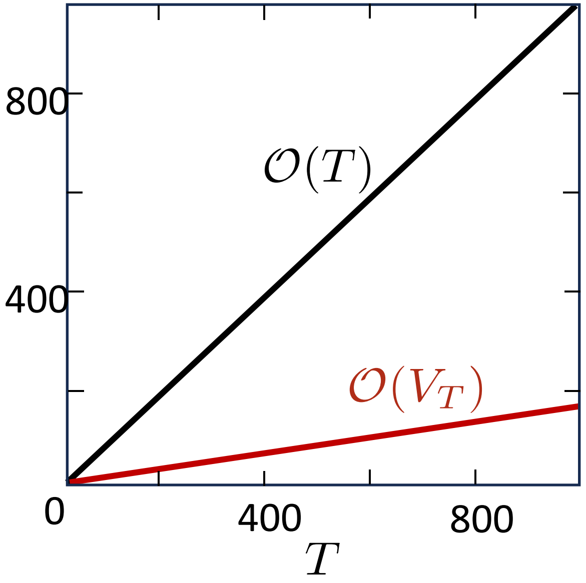
- Previously, we have achieved minimax results like $\mathcal{O}(\sqrt{T(1+P_T)})$.

- More ambitious: achieving ***problem-dependent*** guarantees
  - ▶ become tighter than worst-case results for benign problems
  - ▶ safeguard the same minimax rate in the worst case

gradient variation

$$V_T = \sum_{t=2}^{T} \sup_{\mathbf{w}\in\mathcal{W}} \|\nabla f_{t-1}(\mathbf{w}) - \nabla f_t(\mathbf{w})\|_2^2$$

It is also essential due to profound connections with many other areas such as online games, stochastic optimization, etc.



Exploiting historical information

- How to exploit the niceness of the environments?

focusing on the gradient feedback for simplicity

Optimistic Online Gradient Descent [Rakhlin and Sridharan, 2013]

$$\begin{aligned} \widehat{\mathbf{w}}_{t+1} &= \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_t - \eta \nabla f_t\left(\mathbf{w}_t\right)\right] \\ \mathbf{w}_{t+1} &= \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t+1} - \eta M_{t+1}\right]. \end{aligned}$$

where $\{M_1, M_2, \ldots, M_T\}$ is the ***hint sequence*** encoding prior knowledge of future.

- If the environment is benign, which means it is ”predictable”, and thus we can provide the $\{M_t\}_{t=1}^T$ sequence by exploiting historical information.
- A two-step update fashion, and it will degenerate as the standard OGD when there is no external hint (simply setting $M_t = \mathbf{0}$).

Base Algorithm Analysis

- Optimistic OGD can serve as the base learner for problem-dependent dynamic regret minimization.

$$\widehat{\mathbf{w}}_{t+1} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_t - \eta \nabla f_t(\mathbf{w}_t)\right]$$
$$\mathbf{w}_{t+1} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t+1} - \eta M_{t+1}\right].$$

Theorem 4. *Under certain standard assumptions, the dynamic regret of optimistic OGD over comparator sequence* $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{W}$ *is bounded as*

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \le GD + \frac{1}{2\eta}\underbrace{(D^2 + 2DP_T)}_{\text{non-stationarity}} + \eta \underbrace{\sum_{t=2}^{T} \|\nabla f_t(\mathbf{w}_t) - M_t\|^2}_{\text{adaptivity}} \underbrace{- \frac{1}{\eta}\sum_{t=2}^{T} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|^2}_{\text{negative term}}$$

crucial for gradient variation

$$= \mathcal{O}\left(\frac{1 + P_T}{\eta} + \eta A_T\right),$$

where $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|$ *measures non-stationarity and* $A_T = \sum_{t=2}^{T} \|\nabla f_t(\mathbf{w}_t) - M_t\|^2$ *reflects adaptivity.*

Online Ensemble for Adaptive Bounds

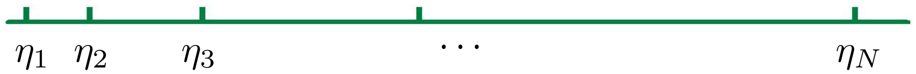
- An online ensemble to balance between *non-stationarity* and *adaptivity*.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i}$$

$$\boxed{\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\frac{1+P_T}{\eta} + \eta A_T\right)}$$

▸ **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{\eta_i = 2^{i-1}\frac{D}{2GT} \mid i = 1, \ldots, N\right\}$$

with $N = \lceil \log_2(GT/(8D^2L^2)) \rceil + 1 = \mathcal{O}(\log T)$.

▸ **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\widehat{\mathbf{w}}_{t+1,i} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t,i} - \eta_i \nabla f_t(\mathbf{w}_t)\right]$$
$$\mathbf{w}_{t+1,i} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t+1,i} - \eta_i M_{t+1}\right].$$

▸ **Meta algorithm:** provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

also include the "hint" in the performance evaluation

Hedge: $p_{t+1,i} \propto \exp\left(-\varepsilon(L_{t,i} + m_{t+1,i})\right), \ \forall i \in [N].$

$$L_{t,i} \triangleq \sum_{s=1}^{t} \ell_s(\mathbf{w}_{s,i}) = \sum_{s=1}^{t} \langle \nabla f_s(\mathbf{w}_s), \mathbf{w}_{s,i} \rangle, \ m_{t+1,i} \triangleq \langle M_{t+1}, \mathbf{w}_{t,i} \rangle.$$

Online Ensemble for Adaptive Bounds

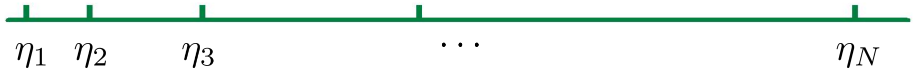
- An online ensemble to balance between ***non-stationarity*** and ***adaptivity***.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i}$$

$$\boxed{\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\frac{1+P_T}{\eta} + \eta A_T\right) = \mathcal{O}\left(\sqrt{A_T(1+P_T)}\right)}$$

▶ **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{\eta_i = 2^{i-1}\frac{D}{2GT} \mid i = 1, \ldots, N\right\}$$

with $N = \lceil \log_2(GT/(8D^2L^2)) \rceil + 1 = \mathcal{O}(\log T)$.

▶ **Base learners:** each updated using $\eta_i \in \mathcal{H}$

$$\widehat{\mathbf{w}}_{t+1,i} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t,i} - \eta_i \nabla f_t(\mathbf{w}_t)\right]$$
$$\mathbf{w}_{t+1,i} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t+1,i} - \eta_i M_{t+1}\right].$$

▶ **Meta algorithm:** provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

also include the "hint" in the performance evaluation

Hedge: $p_{t+1,i} \propto \exp\left(-\varepsilon(L_{t,i} + m_{t+1,i})\right), \ \forall i \in [N].$

$$L_{t,i} \triangleq \sum_{s=1}^{t} \ell_s(\mathbf{w}_{s,i}) = \sum_{s=1}^{t} \langle \nabla f_s(\mathbf{w}_s), \mathbf{w}_{s,i}\rangle, \ m_{t+1,i} \triangleq \langle M_{t+1}, \mathbf{w}_{t,i}\rangle.$$

Gradient-Variation Dynamic Regret

Learning And Mining from DatA

- From adaptive bound to ***gradient-variation*** regret bound

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{A_T(1+P_T)}\right)$$

non-stationarity $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|$

adaptivity $A_T = \sum_{t=2}^{T} \|\nabla f_t(\mathbf{w}_t) - M_t\|^2$

gradient variation $$V_T \triangleq \sum_{t=2}^{T} \sup_{\mathbf{w} \in \mathcal{W}} \|\nabla f_{t-1}(\mathbf{w}) - \nabla f_t(\mathbf{w})\|_2^2$$

problem-dependent

⇨ setting $M_{t+1} = \nabla f_t(\mathbf{w}_t)$ as the last-round gradient

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{(1+P_T) \cdot \sum_{t=2}^{T} \|\nabla f_t(\mathbf{w}_t) - \nabla f_{t-1}(\mathbf{w}_{t-1})\|^2}\right)$$

only "data-dependent"

need to analyze $\|\mathbf{w}_t - \mathbf{w}_{t-1}\|^2$ (*stability* of the dynamics)

Stability Analysis

- Stability of the meta-base online ensemble

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i} \quad \Longrightarrow \quad \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 \leq \underline{2D^2 \left\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\right\|_1^2} + \underline{2\sum_{i=1}^{N} p_{t,i}\left\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\right\|_2^2}$$

meta stability

weighted combine of base stability

- Decompose the overall dynamic regret into the meta-base two levels:

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) = \underbrace{\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{w}_{t,i})}_{\texttt{meta-regret}} + \underbrace{\sum_{t=1}^{T} f_t(\mathbf{w}_{t,i}) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)}_{\texttt{base-regret}}$$

- `meta-regret` $\leq \mathcal{O}\left(\varepsilon V_T + \varepsilon \sum_{t=2}^{T} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 + \frac{1+P_T}{\varepsilon} - \frac{1}{\varepsilon}\sum_{t=2}^{T}\left\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\right\|_1^2\right)$ *negative term for self-cancellation*
- `base-regret` $\leq \mathcal{O}\left(\eta_i V_T + \eta_i \sum_{t=2}^{T} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 + \frac{1}{\eta_i} - \frac{1}{\eta_i}\sum_{t=2}^{T}\left\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\right\|_2^2\right)$ *only for a particular base learner, not sufficient for cancellation*

Stability Analysis

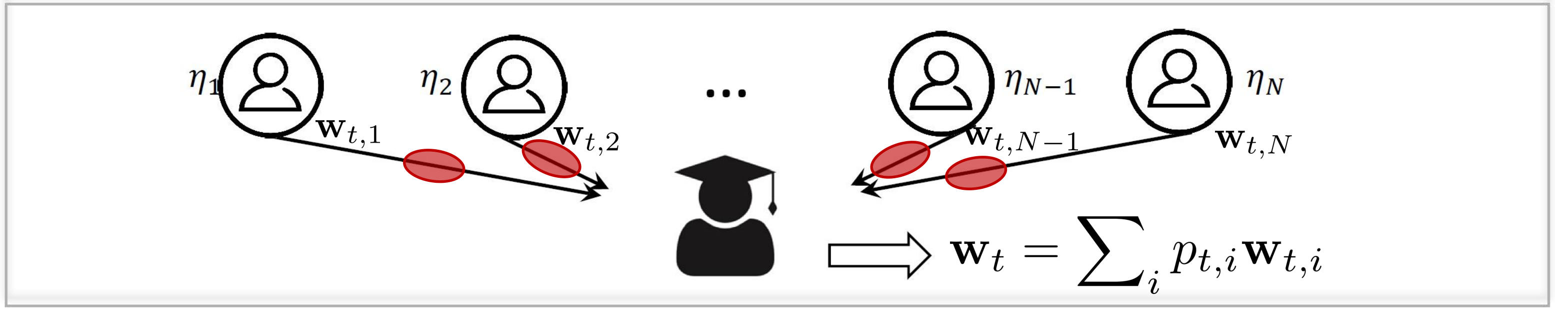
Learning And Mining from DatA

- Stability of the meta-base online ensemble

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i} \quad \Longrightarrow \quad \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 \leq \underbrace{2D^2 \left\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\right\|_1^2}_{\text{meta stability}} + \underbrace{2\sum_{i=1}^{N} p_{t,i} \left\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\right\|_2^2}_{\text{weighted combine of base stability}}$$

- **Stablization**: meta algorithm $p_{t+1,i} \propto \exp\left(-\varepsilon(L_{t,i} + m_{t+1,i})\right)$ with
  - surrogate loss $\boldsymbol{\ell}_t \in \mathbb{R}^N$ with $\ell_{t,i} = \langle \nabla f_t(\mathbf{w}_t), \mathbf{w}_{t,i}\rangle + \lambda \|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2$;
  - hint prediction $\boldsymbol{m}_{t+1} \in \mathbb{R}^N$ with $m_{t+1,i} = \langle M_{t+1}, \mathbf{w}_{t+1,i}\rangle + \lambda \|\mathbf{w}_{t+1,i} - \mathbf{w}_{t,i}\|_2^2$.

correction:penalizing instable base learners



Collaborative Online Ensemble

- Dynamic regret of the modified algorithm (***with corrections***):
  - $\texttt{meta-regret} \leq \mathcal{O}\left(\frac{1+P_T}{\varepsilon} + \varepsilon V_T + \varepsilon \sum_{t=2}^{T} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 - \frac{1}{\varepsilon}\sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 + \frac{1}{\eta_i}\sum_{t=2}^{T} \|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2 - \sum_{t=2}^{T}\sum_{i=1}^{N} \frac{1}{\eta_i} p_{t,i} \|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2\right)$

    these two terms are due to correction

  - $\texttt{base-regret} \leq \mathcal{O}\left(\eta_i V_T + \frac{1}{\eta_i} + \eta_i \sum_{t=2}^{T} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 - \frac{1}{\eta_i}\sum_{t=2}^{T} \|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2\right)$

Collaborative Online Ensemble

- Dynamic regret of the

$$\|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 \le 2D^2 \underbrace{\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2}_{\textit{meta stability}} + \underbrace{2\sum_{i=1}^N p_{t,i}\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2}_{\textit{weighted combine of base stability}}$$

ons):

- $$\texttt{meta-regret} \le \mathcal{O}\Bigg(\frac{1+P_T}{\varepsilon} + \varepsilon V_T + \varepsilon \cancel{\sum_{t=2}^T \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2} - \frac{1}{\varepsilon}\cancel{\sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2}$$
$$+ \frac{1}{\eta_i}\sum_{t=2}^T \|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2 - \sum_{t=2}^T \cancel{\sum_{i=1}^N \frac{1}{\eta_i} p_{t,i}\|\mathbf{w}_{t,i}} - \mathbf{w}_{t-1,i}\|_2^2\Bigg)$$

these two terms are due to correction

- $$\texttt{base-regret} \le \mathcal{O}\Bigg(\eta_i V_T + \frac{1}{\eta_i} + \eta_i \cancel{\sum_{t=2}^T \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2} - \frac{1}{\eta_i}\sum_{t=2}^T \|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2\Bigg)$$

with suitable parameter configurations

$$\text{D-Regret}_T \le \mathcal{O}\left(\sqrt{V_T(1+P_T)}\right)$$

Collaborations between meta and base learners:
simultaneously exploiting

- ✷ *negative terms in the regret analysis*
- ✷ *correction terms in the algorithm design*

Collaborative Online Ensemble

$$\|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 \le 2D^2 \underbrace{\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2}_{\textit{meta stability}} + \underbrace{2\sum_{i=1}^{N} p_{t,i}\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2}_{\textit{weighted combine of base stability}}$$

- Dynamic regret of the ***ns***):

  - $$\texttt{meta-regret} \le \mathcal{O}\left(\frac{1+P_T}{\varepsilon} + \varepsilon V_T + \cancel{\varepsilon\sum_{t=2}^{T}\|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2} - \cancel{\frac{1}{\varepsilon}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2} + \cancel{\frac{1}{\eta_i}\sum_{t=2}^{T}\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2} - \cancel{\sum_{t=2}^{T}\sum_{i=1}^{N}\frac{1}{\eta_i}p_{t,i}\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2}\right)$$

    these two terms are due to correction

  - $$\texttt{base-regret} \le \mathcal{O}\left(\eta_i V_T + \frac{1}{\eta_i} + \cancel{\eta_i\sum_{t=2}^{T}\|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2} - \cancel{\frac{1}{\eta_i}\sum_{t=2}^{T}\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2}\right)$$

with suitable parameter configurations

$$\text{D-Regret}_T \le \mathcal{O}\left(\sqrt{V_T(1+P_T)}\right)$$

Collaborations between meta and base learners: *simultaneously exploiting*

✷ *negative terms in the regret analysis*

✷ *correction terms in the algorithm design*

Online Ensemble for Gradient Variation

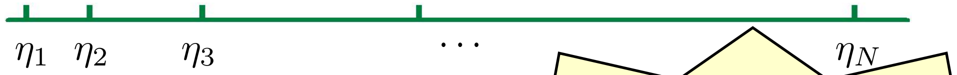
- An online ensemble to balance between non-stationarity and adaptivity.

$$\mathbf{w}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{w}_{t+1,i}$$

$$\boxed{\sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{V_T(1+P_T)}\right)}$$

▶ **Multiple candidates:** to cover uncertainty

diversity consideration: cover all the possible range using as fewer as possible discretization items



$$\mathcal{H} = \left\{ \eta_i = 2^{i-1}\frac{D}{2G}\right.$$

with $N = \lceil \log_2(GT/(8D^2$

correction terms *enable* ***collaborations*** *between meta and base levels*

▶ **Base-learners:** each updated using $\eta_i \in \mathcal{H}$

$$\widehat{\mathbf{w}}_{t+1,i} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t,i} - \eta_i \nabla f_t(\mathbf{w}_t)\right]$$
$$\mathbf{w}_{t+1,i} = \Pi_{\mathcal{W}}\left[\widehat{\mathbf{w}}_{t+1,i} - \eta_i \nabla f_t(\mathbf{w}_t)\right].$$

▶ **Meta-algorithm:** provide the weight $\boldsymbol{p}_{t+1} \in \Delta_N$

Hedge: $p_{t+1,i} \propto \exp\left(-\varepsilon(L_{t,i} + m_{t+1,i})\right),\ \forall i \in [N].$

- surrogate loss $\ell_{t,i} = \langle \nabla f_t(\mathbf{w}_t), \mathbf{w}_{t,i}\rangle + \lambda\|\mathbf{w}_{t,i} - \mathbf{w}_{t-1,i}\|_2^2$;
- hint prediction $m_{t+1,i} = \langle M_{t+1}, \mathbf{w}_{t+1,i}\rangle + \lambda\|\mathbf{w}_{t+1,i} - \mathbf{w}_{t,i}\|_2^2$.

Summary of Our Results

- Full-information online learning

  gradient information is available to the learner

  [Zhang et al., NeurIPS'18; Zhao et al., NeurIPS'20; Zhao et al., NeurIPS'22; Zhao et al., JMLR'23]

- Partial-information online learning

  gradient information cannot be observed, only function value is available

  [Zhao et al., JMLR'21; Luo et al., COLT'22; Yan et al., JMLR'23]

- Decision-dependent online learning

  current decision will affect the future (incl. gradient & function value)

  [Zhao et al., ICML'22; Zhao et al., AISTAST'23; Li et al., NeurIPS'23]

Outline

- Background
- Problem Setup
- Online Ensemble
- Conclusion

Conclusion

- **Online Ensemble**: an effective theoretical framework (base learners; meta learners; schedule) to handle ***uncertainty*** in online environments

- **Non-stationary online learning**: online ensemble for dynamic regret
  - bandit convex optimization: ***surrogate loss*** is essential to exploit limited feedback
  - problem-dependent guarantee: incorporating ***hint prediction***, enable ***collaboration*** between meta and base layers (via negative terms and corrections)
  - other results: online MDPs, game theory, online weakly supervised learning, etc.

- Beyond non-stationarity: universal online learning (agnostic to curvatures)

- Many todo: efficiency/real-time response? non-convexity? continuous learning? …

Thanks!

References

- Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Adaptivity and Non-stationarity: Problem-dependent Dynamic Regret for Online Convex Optimization. Under review. Arxiv preprint: 2112.14368, 2021. **(major reference)**

- Peng Zhao, Guanghui Wang, Lijun Zhang, and Zhi-Hua Zhou. Bandit Convex Optimization in Non-stationary Environments. Journal of Machine Learning Research (**JMLR**), 22(125):1–45, 2021.

- Peng Zhao, Yu-Hu Yan, Yu-Xiang Wang, Zhi-Hua Zhou. Non-stationary Online Learning with Memory and Non-stochastic Control. Journal of Machine Learning Research (**JMLR**), 24(206):1–70, 2023.

Thanks!

References

- Peng Zhao, Long-Fei Li, and Zhi-Hua Zhou. Dynamic Regret of Online Markov Decision Processes. In: Proceedings of the 39th International Conference on Machine Learning (**ICML'22**), Baltimore, Maryland, 2022. Page: 26865-26894.

- Peng Zhao, Yan-Feng Xie, Lijun Zhang, and Zhi-Hua Zhou. Efficient Methods for Non-stationary Online Learning. In: Advances in Neural Information Processing Systems 35 (**NeurIPS'22**), New Orleans, Louisiana, 2022. Page: 11573-11585.

- Haipeng Luo, Mengxiao Zhang, Peng Zhao, and Zhi-Hua Zhou. (alphabetical order). Corralling a Larger Band of Bandits: A Case Study on Switching Regret for Linear Bandits. In: Proceedings of the 35th Annual Conference on Learning Theory (**COLT'22**), London, UK, 2022. Page: 3635-3684.

Thanks!

敬请各位专家批评指正！