Transfer Learning Algorithms for Chemical Sensor Arrays

Prof. Lei Zhang

School of Microelectronics and Communication Engineering,
Chongqing University, Chongqing, China
https://www.leizhang.tk

Paper Resources and Open-source Codes (Python vs. Matlab)

Google scholar: https://scholar.google.com/citations?user=Nt9es7kAAAAJ&hl=zh-CN

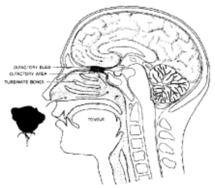
ISOCS Winter School
Bormio, Italy, 17/01/2024

Contents

- Part I: Background and Preliminary
- Part II: Transfer learning: Concept, Theory and Algorithms
- Part III: Transfer Learning Algorithms for E-nose
- Summary vs. Future

Contents

- Part I: Background and Preliminary
- Part II: Transfer learning: Concept, Theory and Algorithms
- Part III: Transfer Learning Algorithms for E-nose
- Summary vs. Future



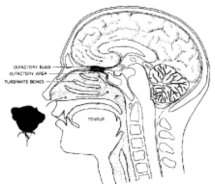
1982

Electronic Nose

- Electronic olfaction system equipped with a model nose was firstly proposed to mimic biological olfaction mechanism as early as 1982.
- One key characteristic of model nose is the odorant detectors (primary neurons) respond to a wide range of chemicals.

K. Persaud and G. Dodd, “Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose,” Nature, 1982.

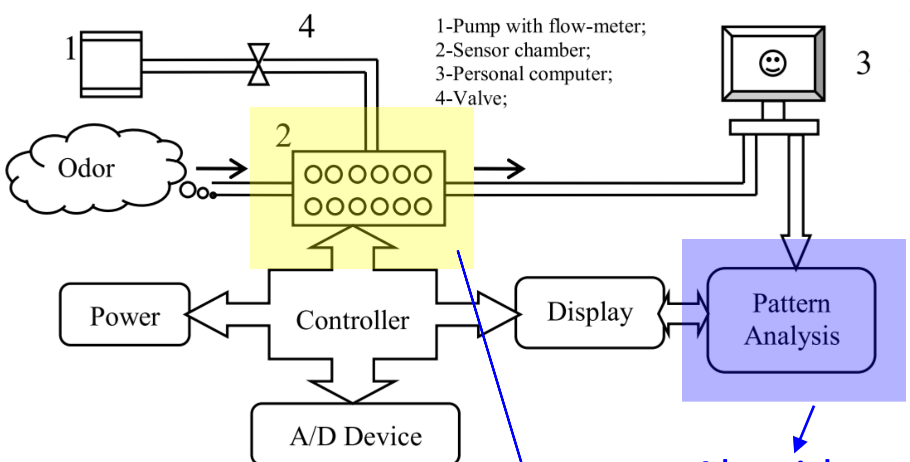
J.W. Gardner and P.N. Bartlett, “A brief history of electronic noses,” Sens. Actu. B Chem., 1994.

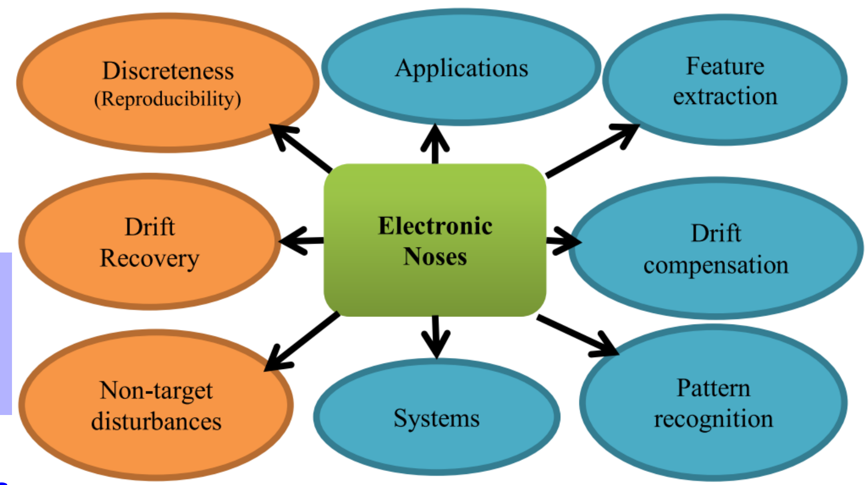


nature International journal of science 1982

Electronic Nose

- A basic e-nose system includes some components.
- Great progress is made in e-nose.





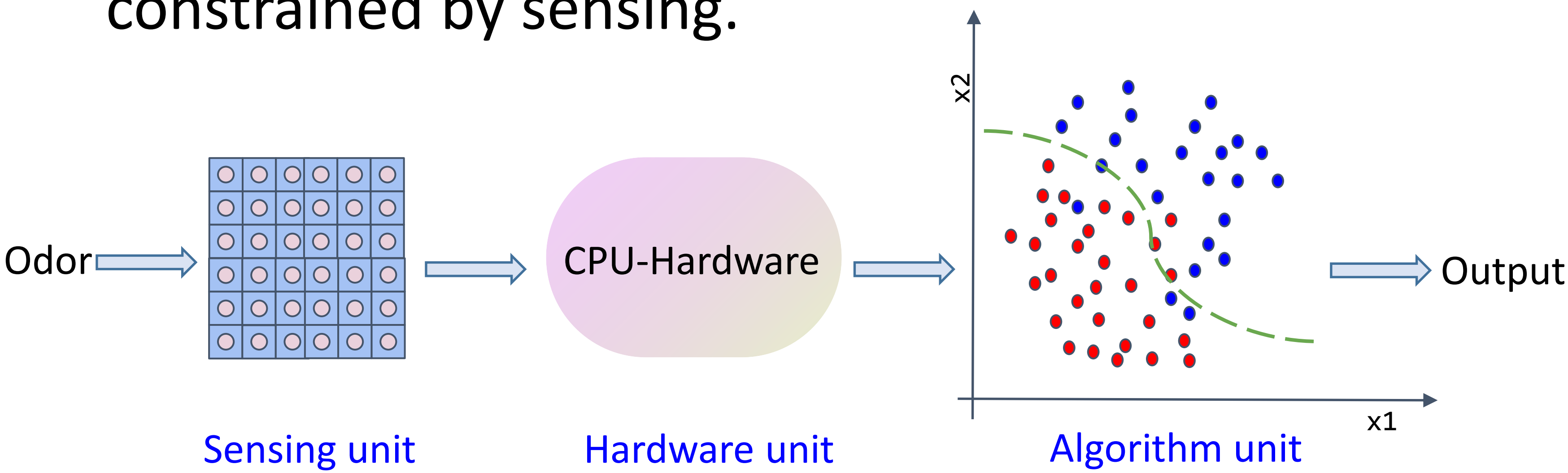
Algorithms

Metal oxide semiconductor (MOS) chemical sensor array

L. Zhang and F.C. Tian, "Performance Study of Multilayer Perceptrons in Low-Cost Electronic Nose," IEEE Trans. Instrumentation and Measurement, 2014.
L. Zhang and D. Zhang, "Efficient Solutions for Discreteness, Drift, and Disturbance (3D) in Electronic Olfaction," IEEE Trans. Systems, Man, and Cybernetics: Systems, 2018.

Challenge on MOS Sensors

- Sensing unit plays a crucial role in e-nose.
- Algorithm benefits from sensing, but is also constrained by sensing.



Drift of MOS sensors is always a tricky problem and a disaster of e-nose.

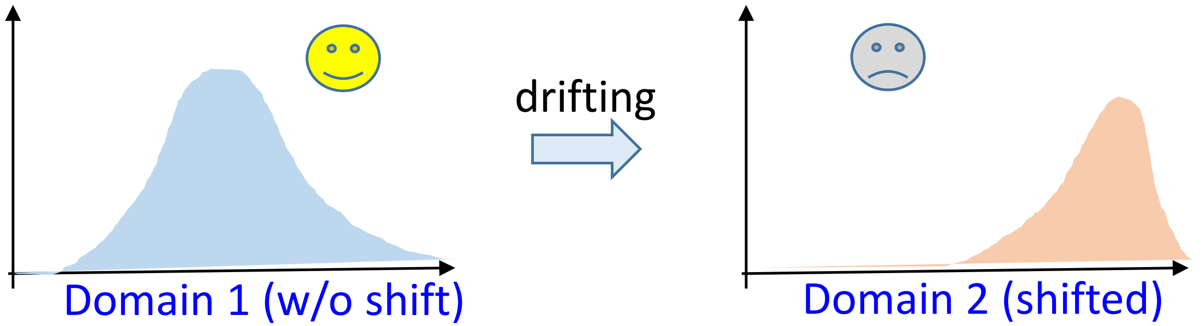
Lei Zhang, Fengchun Tian, and David Zhang, "Electronic Nose: Algorithmic Challenge," Springer, 2018. pages: 1-339.

Challenge on MOS Sensors

- What is sensor drift?

This is caused by unknown dynamic process such as poisoning, aging or environmental variations (temperature, humidity, etc.).

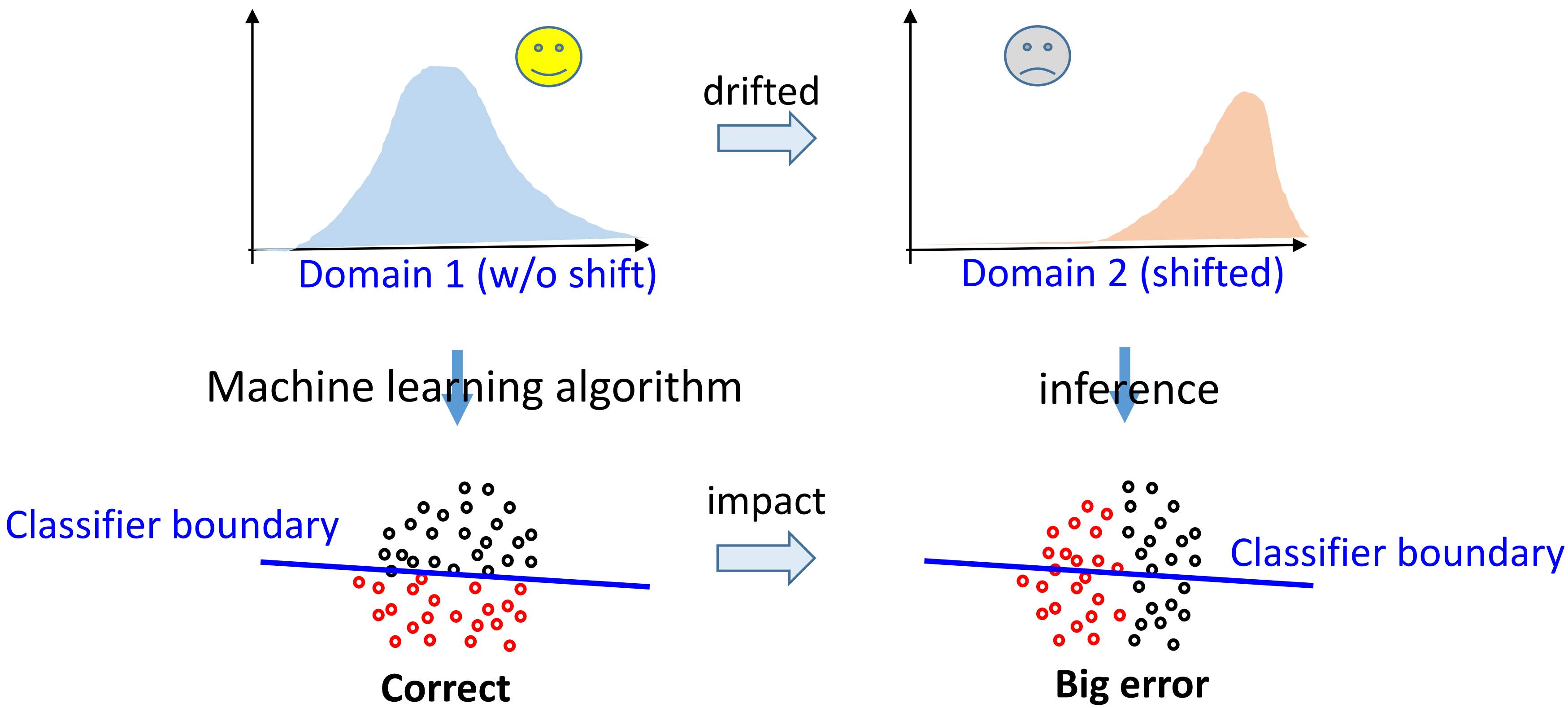
Drift causes outliers in data samples and data distribution shift (covariate shift).



M. Holmberg, et al., “Drift counteraction in odour recognition applications: lifelong calibration method,” Sens. Actu. B: Chem, 1997.

Challenge on MOS Sensors

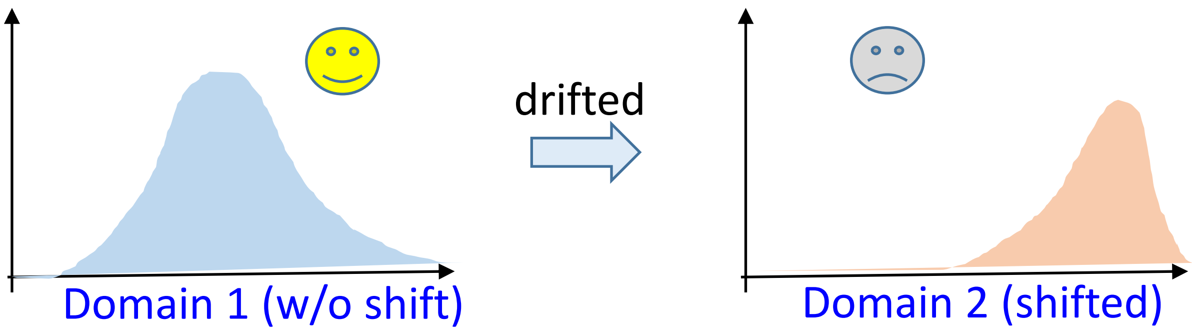
- What is the impact of drift?



Impact: drift brings classification algorithm to big errors.

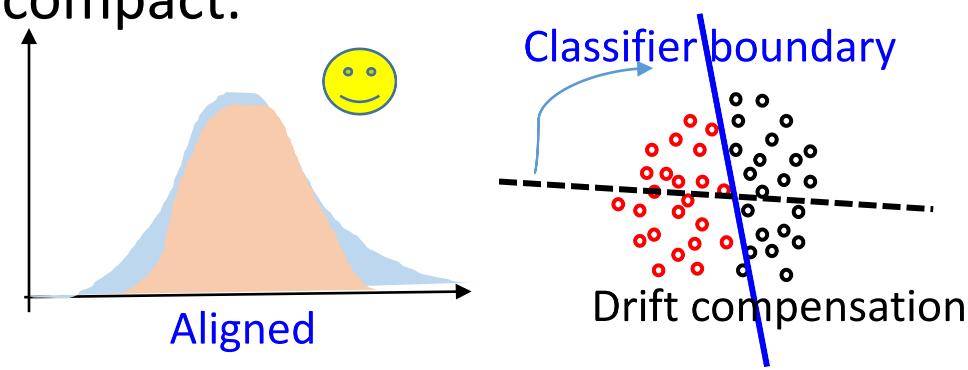
Challenge on MOS Sensors

- How to deal with drift?



Align feature distribution between domain 1 and domain 2, then compensate drift compact.

Transfer learning



L. Zhang and D. Zhang, "Domain adaptation extreme learning machines for drift compensation in e-nose systems," IEEE Trans. Instru. Measu., 2015.

Contents

Difference from Calibration Transfer

- Calibration transfer in E-nose mainly refers to instrumental calibration between systems (master vs. slave).

✓ Standardization is the classical technique for calibration transfer, including standardization on **model coefficients** f, **sensor response** x, and **prediction output** y.

✓ Standardization on sensor responses such as direct standardization (DS), piece-wise direct standardization (PDS), OSC, CC, and GLSW are most popular.

- But, Transfer learning is a new learning methodology towards out-of-distribution generalization and adaptation.

R. Laref, et al., “Support vector machine regression for calibration transfer between electronic noses dedicated to air pollution monitoring,” Sensors, 2018.

A Revisit of Machine Learning

- Machine learning is a modeling technique with statistics for parameters estimation of unknown func.
- To be simple, given a dataset (X, y) with **label y**, a statistical learning model is to find a mapping func. (hypothesis) f(.) between X and y, such that

$$y=f(x)$$

- A learning problem to be solved is how to find f(.)?
- Many learning techniques from shallow to deep.
- Gradient descent based optimization techniques.

A Revisit of Machine Learning

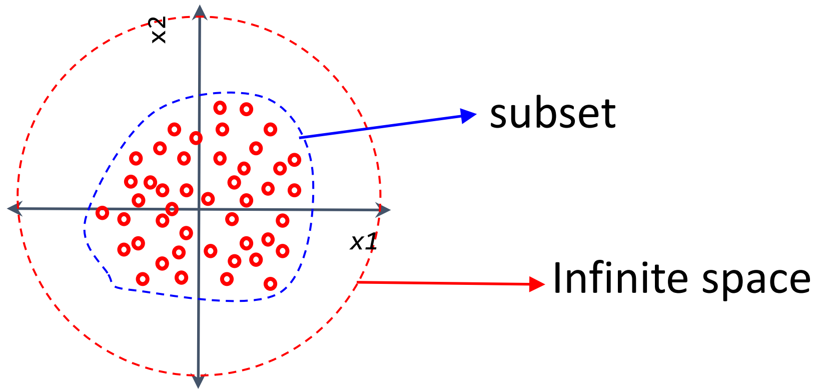
- To find an optimal mapping (solution) f(.), machine learning is transformed to an *optimization* technique.
- A general optimization (*minimization*) problem of learning is,

$$\min R[\mathrm{Pr}, \theta, l(x, y, \theta)] = \mathbf{E}_{(x,y)\sim\mathrm{Pr}}[l(x, y, \theta)]$$

- R[.] is the *expected* risk defined by the loss function with input (X,y) sampled from a probabilistic distribution *Pr* and parameter θ of f(.)
- Pr should be an independent identical distribution (i.i.d.) of test samples.

A Revisit of Machine Learning

- However, due to the infinite space of the data distribution, we can only have a subset of the data (i.e., training data).



- So, the expected risk minimization is transformed into an *empirical* risk optimization problem,

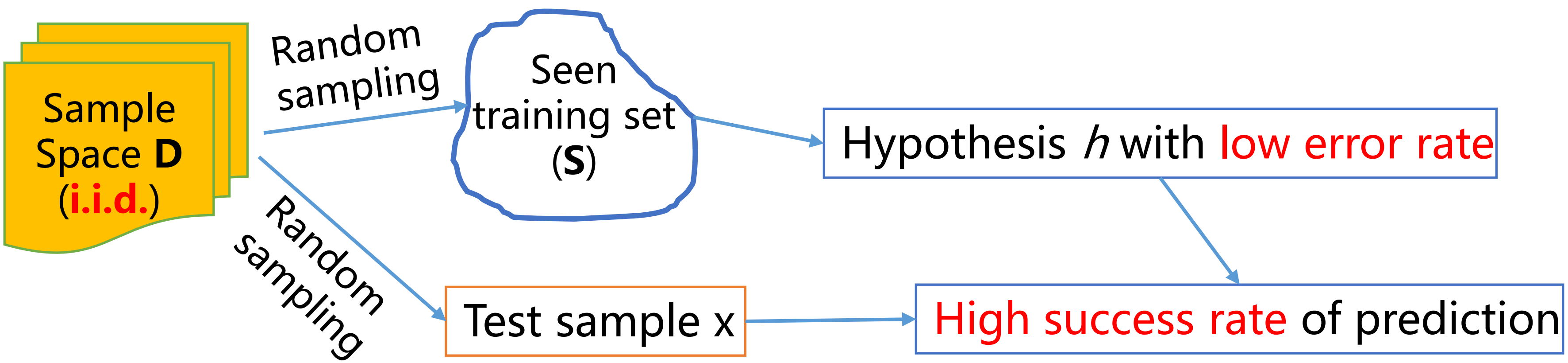
$$R[\Pr, \theta, l(x, y, \theta)] = \mathbf{E}_{(x,y)\sim\Pr}\left[l(x, y, \theta)\right]$$

$$R_{\text{emp}}[Z, \theta, l(x, y, \theta)] = \frac{1}{m}\sum_{i=1}^{m} l(x_i, y_i, \theta)$$

Where m is the size (number) of the finite training subset sampled from the distribution Pr.

A Revisit of Machine Learning

- Okay, now we can have a view of a general machine learning framework (**Probably Approximate Correct, PAC**).



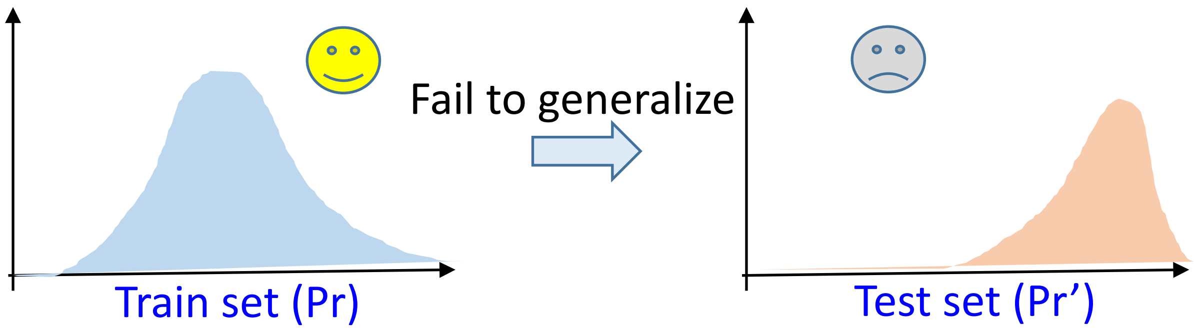
- A prior assumption is the i.i.d. condition.

The training set and test sample should be sampled from an independent identical distribution (i.i.d.)

Unfortunately, i.i.d. is not realistic. What should we do?

Dilemma of Non i.i.d.

- Generalization is the final objective of ML task.
- The optimized parameter θ of the mapping function $f\,(.)$ on a training subset sampled from Pr often fails to *generalize* a test subset sampled from a non *i.i.d.* distribution Pr'.

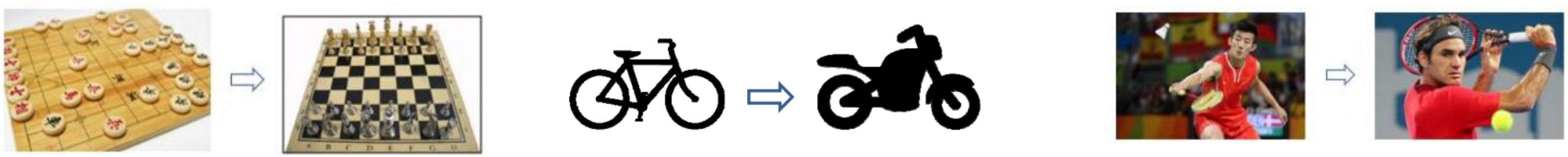


- This is analogy to sensor drift (distribution shift), which **opens the research on TL for drift in e-nose**.

A Preliminary of Transfer Learning

Toy Examples:

Semantic related but distribution different tasks



Behavior learning skills (domain common knowledge)

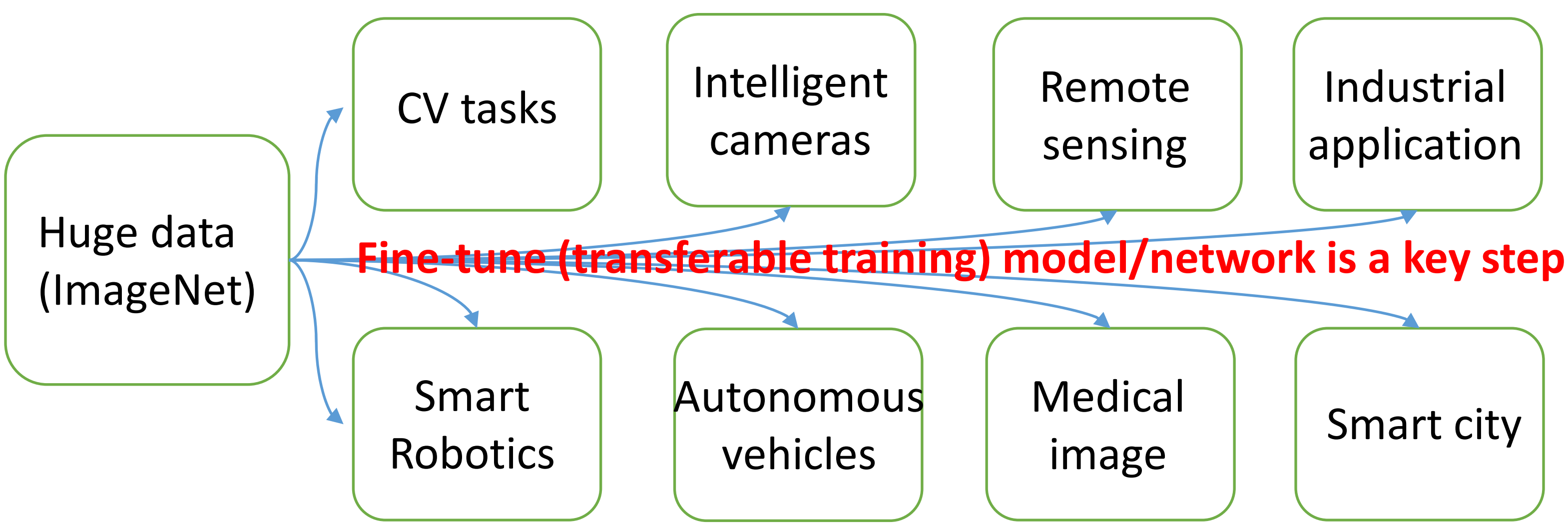
Computer Vision
(image classification)

Natural Language Processing
(translation)

Text Recognition

Fine-tune Paradigm

- Transfer learning has been a widely used technique in a wide spread of applications.
- In deep learning era, you may hear from about the “fine-tune” technique for down-stream tasks.



A Preliminary of Transfer Learning

Problem definition:

Given a target task (D_T) without labels (or few labels), *how to learn a reliable predictor/classifier for D_T*,

Not feasible? (TL emerges)

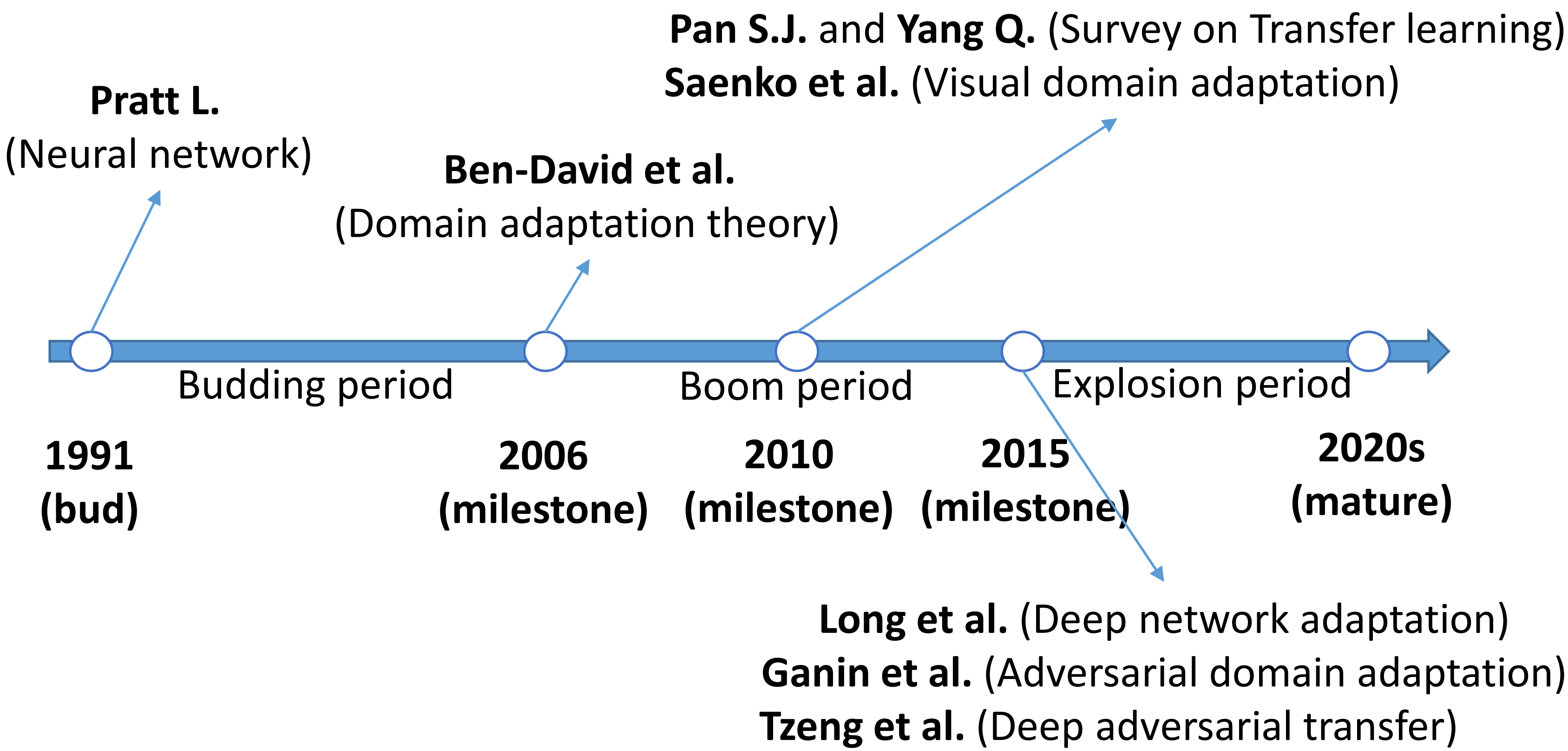
- A sufficiently labeled, semantic related but distribution different source task (D_S) is leveraged as auxiliary training data.
- Two key points:

1) Overcomes the label deficiency problem;

2) But introduces non i.i.d. problem between D_T and D_S

A Preliminary of Transfer Learning

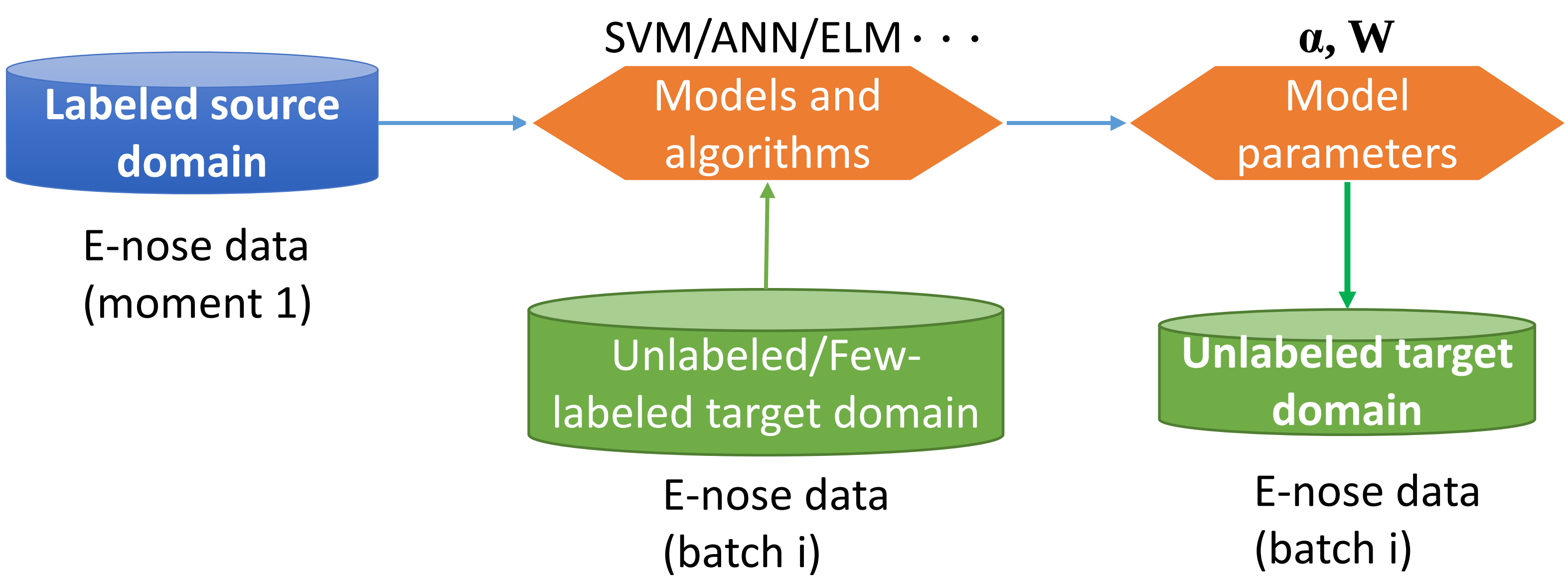
History of Transfer Learning (1990s-2020s):



Definition of TL

What is transfer learning?

Transfer learning or domain adaptation aims to leverage a sufficiently labeled, distribution different but semantic related source domain for training and recognizing target domain samples.



Domain Adaptation Theory

Why are transfer learning models or algorithms effective and reliable?

Namely, *how to guarantee the models or algorithms to have low generalization error on target data?*

Ben-David Shai et al. induced a generalization bound of domain adaptation, widely used as a theoretical guidance for a series of models and algorithms.

Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2006). Analysis of representations for domain adaptation. In: *Advances in neural information processing systems*

Domain Adaptation Theory

Shai Ben-David's generalization bound theorem:

- To be simple, the expected target error $\epsilon_T(h)$ is bounded as (proof based on triangular inequality is removed)

$$\epsilon_T(h) \leq \hat{\epsilon}_S(h) + \sqrt{\frac{4}{m}\left(d \log \frac{2em}{d} + \log \frac{4}{\delta}\right)} + d_{\mathcal{H}}(\tilde{\mathcal{D}}_S, \tilde{\mathcal{D}}_T) + \lambda$$

- $\mathcal{H}$ is the set of hypothesis.
- The upper bound of $\epsilon_T(h)$ consists of four terms.

Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2006). Analysis of representations for domain adaptation. In: *Advances in neural information processing systems*

Distribution Difference Measure

Distribution alignment is the key part of transfer learning.

How to measure distribution difference between two distributions P and Q?

- MMD (Maximum Mean Discrepancy) (Gretton et al. NIPS'06, NIPS'09, JMLR'12)
- HSIC (Hilbert Schmidt Independence Criterion) (Gretton et al. ALT'05; Yan et al. TCYB'17, Wang et al. ICCV'17, CRTL)
- Bregman divergence (Si et al. TKDE'10, TSL)
- Moment statistics (Herath et al. CVPR'17, ILS; Sun et al. arXiv'17, CORAL; Peng et al. ICCV'19)

Theory---->Algorithm

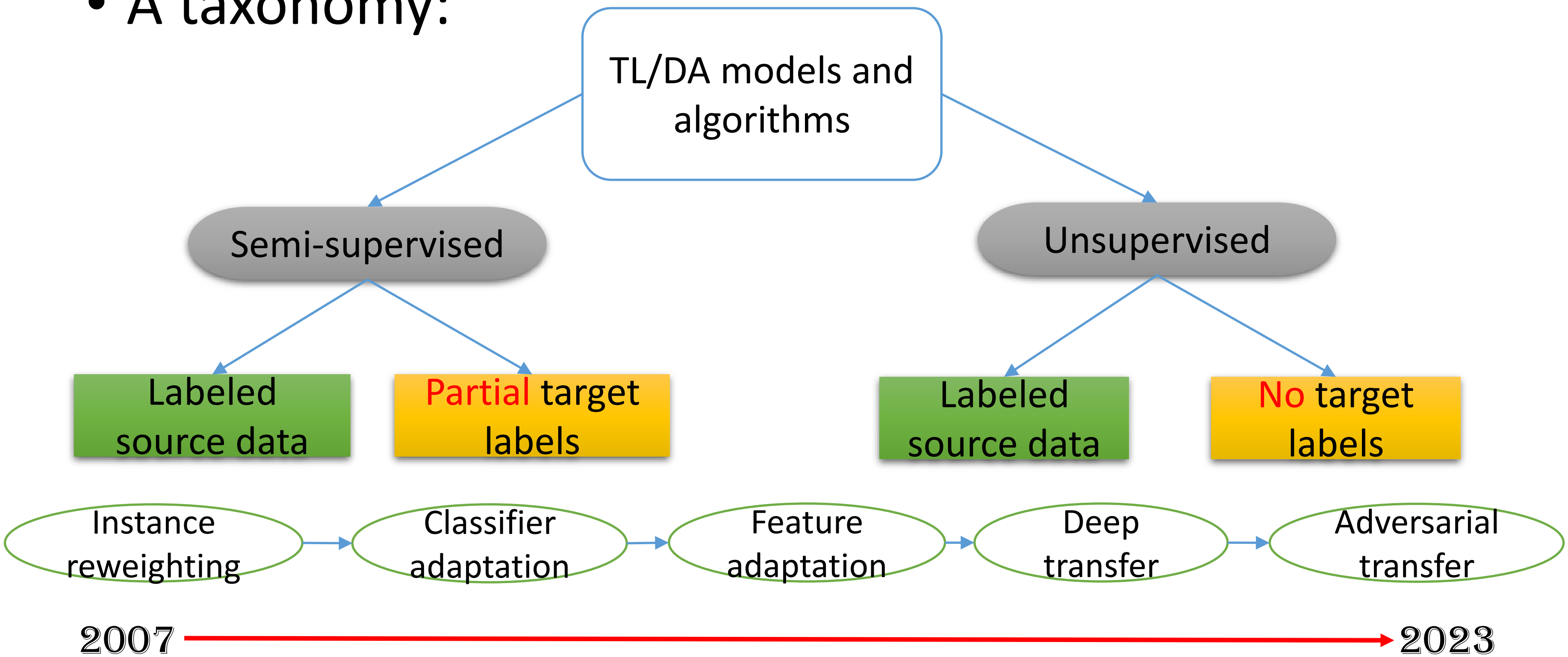
- Induced by the generalization bound theory, a number of models and algorithms are emerged, by focusing on three components during design.

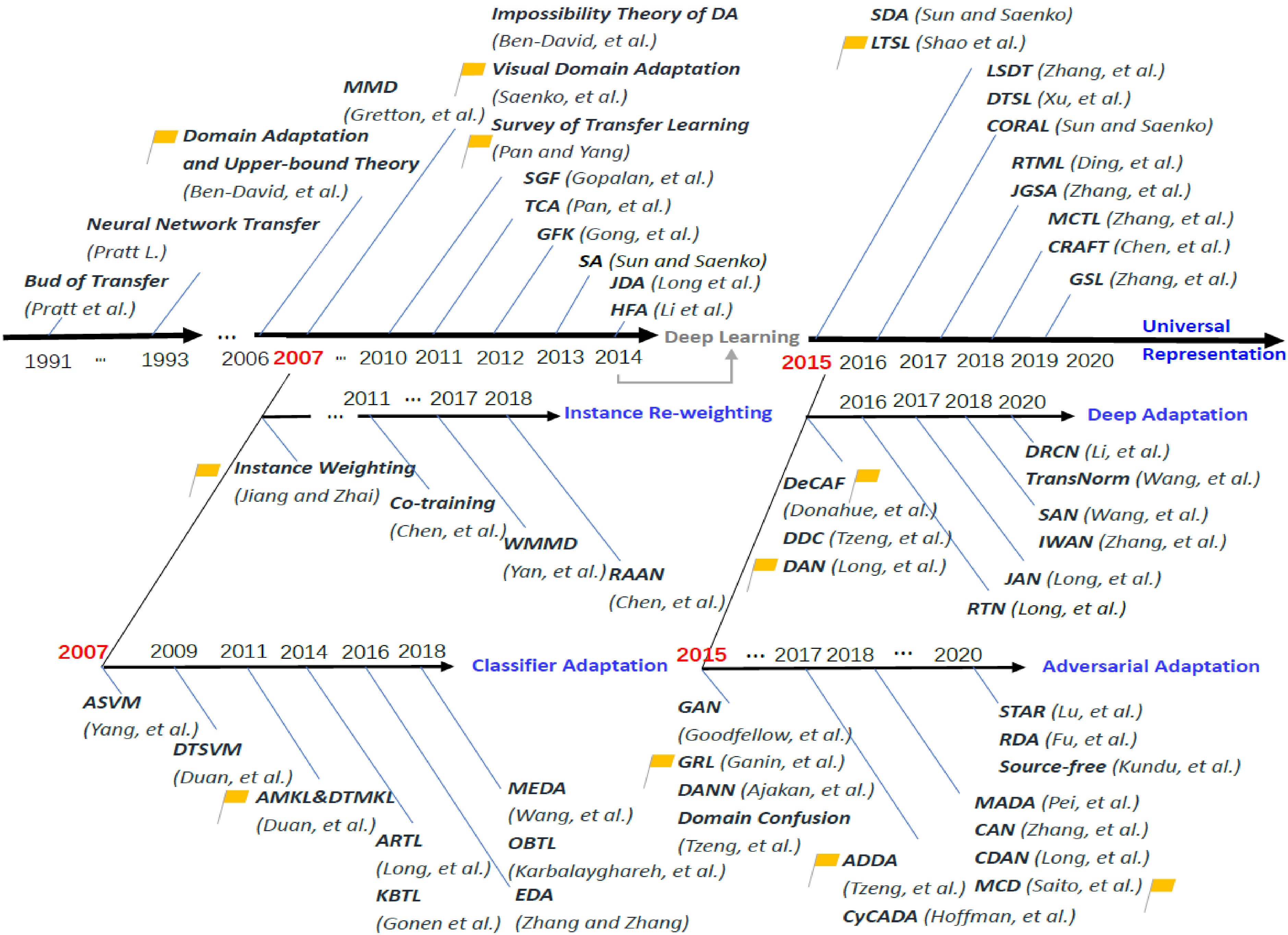
1) Source error $\epsilon_S(h)$
2) Domain discrepancy $\hat{d}_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{U}_S, \mathcal{U}_T)$
3) Combined error $\lambda = \min_{h \in \mathcal{H}} \epsilon_S(h) + \epsilon_T(h)$

Transfer Learning Algorithm

How to design TL/DA models and algorithms?

- A taxonomy:





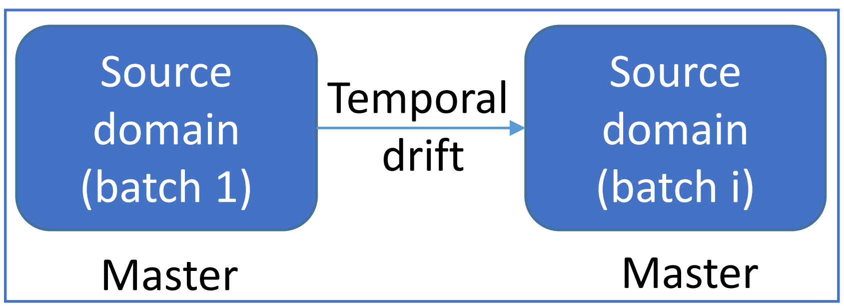
L. Zhang and X. Gao, Transfer Adaptation Learning: A Decade Survey, IEEE TNNLS 2022.

Contents

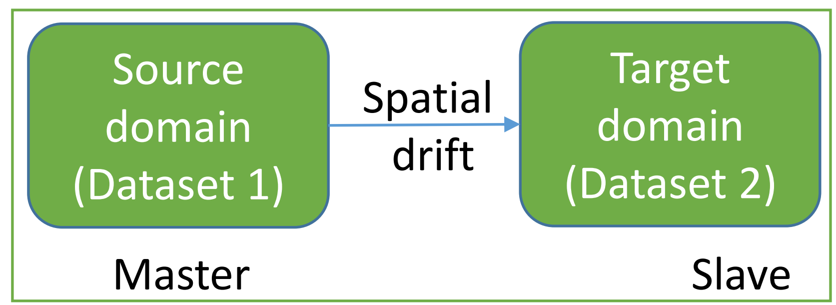
- Part I: Background and Preliminary
- Part II: Transfer learning: Concept, Theory and Algorithms
- Part III: Transfer Learning Algorithms for E-nose
- Summary

TL Algorithm for E-nose

Since 2015, transfer learning/domain adaptation methodology has been widely used for drift adaptation(a)/instrumental calibration (b) in E-noses.



(a) TL for drift compensation



(b) TL for instrumental calibration

TL Algorithm for E-nose

- Two branches in recent years:
- Classifier transfer (parameter adaptation): *How to learn domain-common classifier adapted to two domains?*

- Feature transfer (subspace adaptation): *How to learn domain-common feature representation (projection) between two domains?*

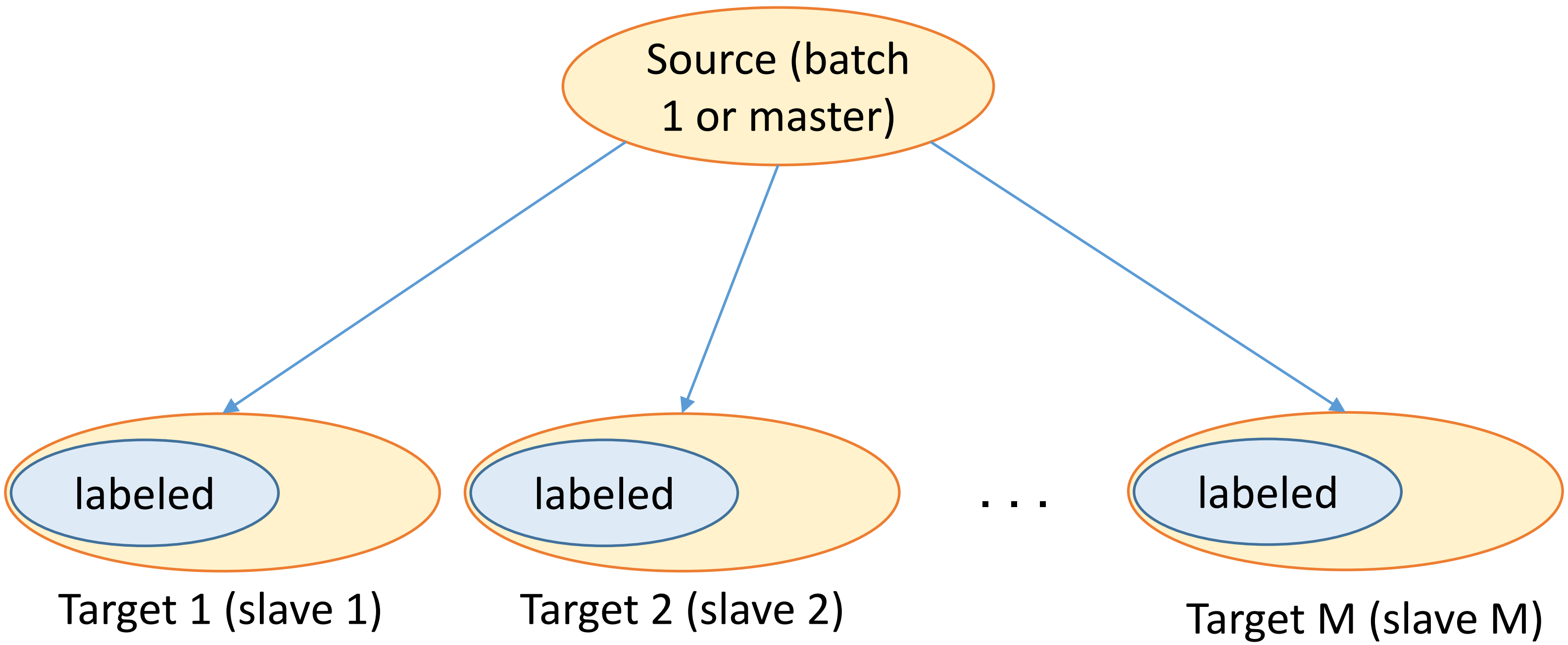
TL Algorithm for E-nose

- Two branches in recent years:
- Classifier transfer (parameter adaptation): *How to learn domain-common classifier adapted to two domains?*

- Feature transfer (subspace adaptation): *How to learn domain-common feature representation (projection) between two domains?*

Domain Adaptive ELM (DAELM)

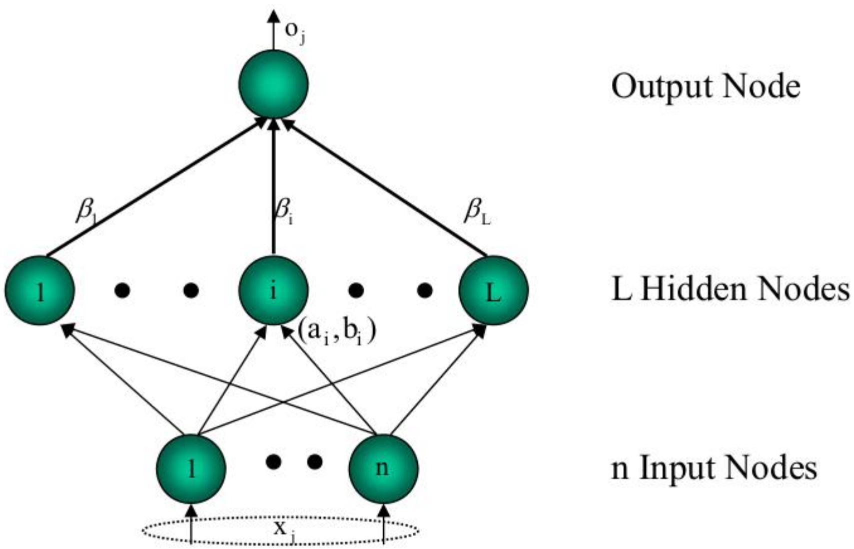
Consider M target tasks with few labeled data by learning ELMs on a sufficiently labeled source domain.



L. Zhang and D. Zhang, “Domain adaptation extreme learning machines for Drift Compensation in E-nose Systems,” IEEE Trans. Instru. Meas., vol. 64, no. 7, 2015.

Extreme Learning Machine

ELM was first proposed for solving a single-layer feed-forward network (SLFN) by Huang et al. 2004



G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012

Extreme Learning Machine

Model and Algorithm:

- Given N samples $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ and their corresponding ground truth $[\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N]$, a general ELM model,

$$\begin{cases} \min\limits_{\boldsymbol{\beta}} \quad \mathcal{L}_{\mathrm{ELM}} = \frac{1}{2}\left\|\boldsymbol{\beta}\right\|^2 + C \cdot \frac{1}{2} \cdot \sum\limits_{i=1}^{N} \left\|\boldsymbol{\xi}_i\right\|^2 \\ \text{s.t.} \quad \mathcal{H}(\mathbf{x}_i)\,\boldsymbol{\beta} = \mathbf{t}_i - \boldsymbol{\xi}_i, \quad i = 1, \ldots, N \end{cases}$$

- $\boldsymbol{\beta}$ is the output weights to be solved. $\mathcal{H}$ is the random hidden layer. C is the penalty and ξ is the prediction error.

Extreme Learning Machine

Model and Algorithm:

- Analytically determined solution (closed-form soultion)
- Overdetermined problem (N>L)

$$\beta^* = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \frac{\mathbf{I}_L}{C}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T}$$

- Underdetermined problem (N<L)

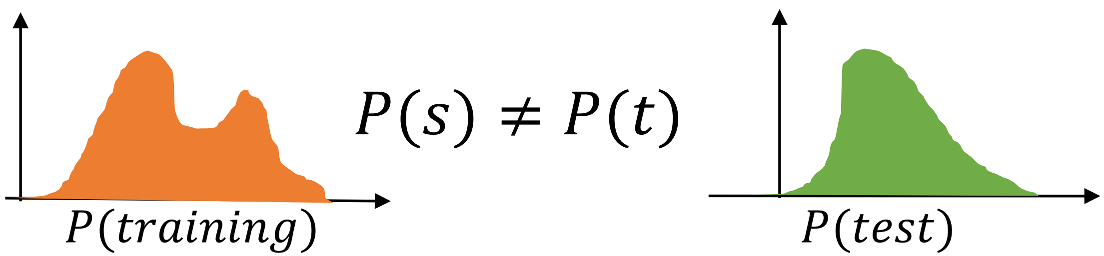
$$\beta^* = \mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{H}^{\mathrm{T}} + \frac{\mathbf{I}_N}{C}\right)^{-1}\mathbf{T}$$

- **H** is the random weights in hidden layer
- ELM does not need to be tuned like Back-propagation.

Common Problem of ELM

Similar to general machine learning, ELM is not transferable:

- For classification problem, labels are needed for ELM.
- ELM is also conditional on the distributions (i.i.d.).
- ELM classifier is not transferable to other domains.

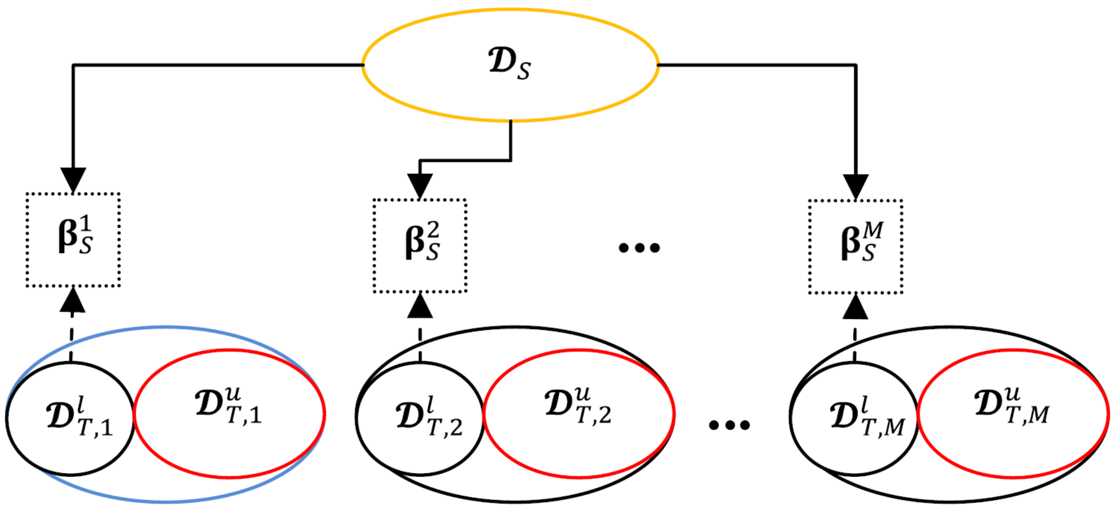


ELM is also not distribution adaptive

- Therefore, we expect some advanced versions of ELM by introducing domain adaptation and transfer learning.

Domain Adaptive ELM (DAELM)

DAELM.v1 (unlabeled target data is not used)



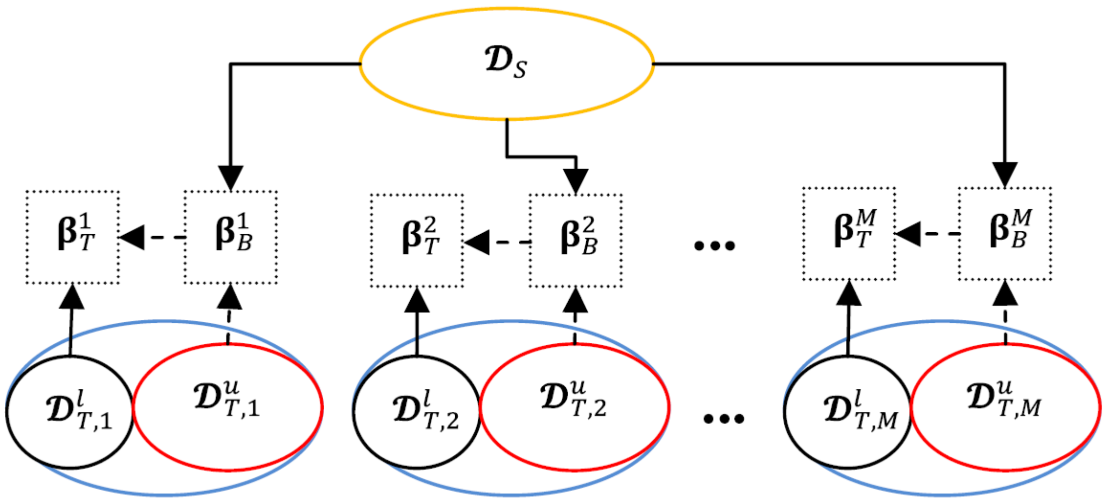
$$\min_{\boldsymbol{\beta}_S, \boldsymbol{\xi}_S^i, \boldsymbol{\xi}_T^i} \frac{1}{2}\|\boldsymbol{\beta}_S\|^2 + C_S\frac{1}{2}\sum_{i=1}^{N_S}\left\|\boldsymbol{\xi}_S^i\right\|^2 + C_T\frac{1}{2}\sum_{j=1}^{N_T}\left\|\boldsymbol{\xi}_T^j\right\|^2$$

$$\text{s.t.} \begin{cases} \mathbf{H}_S^i\boldsymbol{\beta}_S = \mathbf{t}_S^i - \boldsymbol{\xi}_S^i, & i = 1, \ldots, N_S \\ \mathbf{H}_T^j\boldsymbol{\beta}_S = \mathbf{t}_T^j - \boldsymbol{\xi}_T^j, & j = 1, \ldots, N_T \end{cases}$$

Note that \beta is classifier

Domain Adaptive ELM (DAELM)

DAELM.v2 (unlabeled target data is used with pseudo labels)



$$\min_{\boldsymbol{\beta}_T} L_{\mathrm{DAELM-T}}(\boldsymbol{\beta}_T) = \frac{1}{2}\|\boldsymbol{\beta}_T\|^2 + C_T\frac{1}{2}\|\mathbf{t}_T - \mathbf{H}_T\boldsymbol{\beta}_T\|^2 + C_{Tu}\frac{1}{2}\|\mathbf{H}_{Tu}\boxed{\boldsymbol{\beta}_B} - \mathbf{H}_{Tu}\boldsymbol{\beta}_T\|^2$$

Pre-trained source base classifier

Note that \beta is classifier

Domain Adaptive ELM (DAELM)

Prediction/inference stage of unlabeled target data:

DAELM.v1:

$$\mathbf{y}_{Tu}^{k} = \mathbf{H}_{Tu}^{k} \cdot \boldsymbol{\beta}_{S}, \quad k = 1, \ldots, N_{Tu}$$

DAELM.v2:

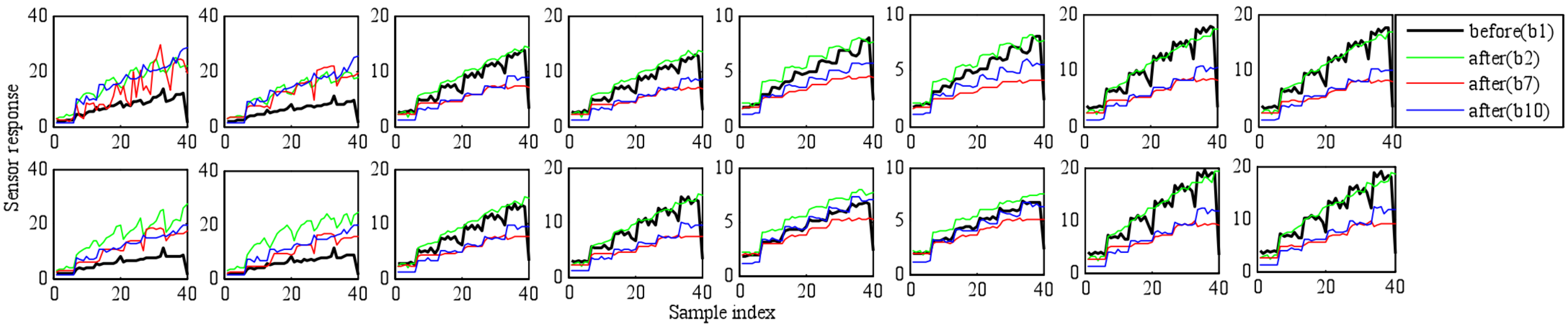
$$\mathbf{y}_{Tu}^{k} = \mathbf{H}_{Tu}^{k} \cdot \boldsymbol{\beta}_{T}, \quad k = 1, \ldots, N_{Tu}$$

Note that \beta is classifier

Experiment

E-nose drift data via (10 batches, 16 gas sensors, acquired from 2008 to 2011 containing 36 months by Vergara et al. 2012):

Batch ID	Month	Acetone	Acetaldehyde	Ethanol	Ethylene	Ammonia	Toluene	Total
Batch 1	1, 2	90	98	83	30	70	74	445
Batch 2	3~10	164	334	100	109	532	5	1244
Batch 3	11, 12, 13	365	490	216	240	275	0	1586
Batch 4	14, 15	64	43	12	30	12	0	161
Batch 5	16	28	40	20	46	63	0	197
Batch 6	17, 18, 19, 20	514	574	110	29	606	467	2300
Batch 7	21	649	662	360	744	630	568	3613
Batch 8	22, 23	30	30	40	33	143	18	294
Batch 9	24, 30	61	55	100	75	78	101	470
Batch 10	36	600	600	600	600	600	600	3600



L. Zhang and D. Zhang, "Domain adaptation extreme learning machines for Drift Compensation in E-nose Systems," IEEE Trans. Instru. Meas., vol. 64, no. 7, 2015.

Experiment

Gas classification performance

Batch 1 is fixed as source domain

Batch ID	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6	Batch 7	Batch 8	Batch 9	Batch 10	Average
CC-PCA	67.00	48.50	41.00	35.50	55.00	31.00	56.50	46.50	30.50	45.72
SVM-rbf	74.36	61.03	50.93	18.27	28.26	28.81	20.07	34.26	34.47	38.94
SVM-gfk	72.75	70.08	60.75	75.08	73.82	54.53	55.44	69.62	41.78	63.76
SVM-comgfk	74.47	70.15	59.78	75.09	73.99	54.59	55.88	70.23	41.85	64.00
ML-rbf	42.25	73.69	75.53	66.75	77.51	54.43	33.50	23.57	34.92	53.57
ML-comgfk	80.25	74.99	78.79	67.41	77.82	71.68	49.96	50.79	53.79	67.28
ELM-rbf	70.63	66.44	66.83	63.45	69.73	51.23	49.76	49.83	33.50	57.93
Our DAELM-S(20)	87.57	96.53	82.61	81.47	84.97	71.89	78.10	87.02	57.42	**80.84**
Our DAELM-S(30)	87.98	95.74	85.16	95.99	94.14	83.51	86.90	100.0	53.62	**87.00**
Our DAELM-T(40)	83.52	96.34	88.20	99.49	78.43	80.93	87.42	100.0	56.25	**85.62**
Our DAELM-T(50)	97.96	95.34	99.32	99.24	97.03	83.09	95.27	100.0	59.45	**91.86**

Batch k is fixed as source domain and batch k+1 is target domain

Batch ID	1→2	2→3	3→4	4→5	5→6	6→7	7→8	8→9	9→10	Average
SVM-rbf	74.36	87.83	90.06	56.35	42.52	83.53	91.84	62.98	22.64	68.01
SVM-gfk	72.75	74.02	77.83	63.91	70.31	77.59	78.57	86.23	15.76	68.56
SVM-comgfk	74.47	73.75	78.51	64.26	69.97	77.69	82.69	85.53	17.76	69.40
ML-rbf	42.25	58.51	75.78	29.10	53.22	69.17	55.10	37.94	12.44	48.17
ML-comgfk	80.25	98.55	84.89	89.85	75.53	91.17	61.22	95.53	39.56	79.62
Ensemble	74.40	88.00	92.50	94.00	69.00	69.50	91.00	77.00	65.00	80.04
ELM-rbf	70.63	40.44	64.16	64.37	72.70	80.75	88.20	67.00	22.00	63.36
Our DAELM-S(20)	87.57	96.90	85.59	95.89	80.53	91.56	88.71	88.40	45.61	**84.53**
Our DAELM-S(30)	87.98	96.58	89.75	99.04	84.43	91.75	89.83	100.0	58.44	**88.64**
Our DAELM-T(40)	83.52	96.41	81.36	96.45	85.13	80.49	85.71	100.0	56.81	**85.10**
Our DAELM-T(50)	97.96	95.62	99.63	98.17	97.13	83.10	94.90	100.0	59.88	**91.82**

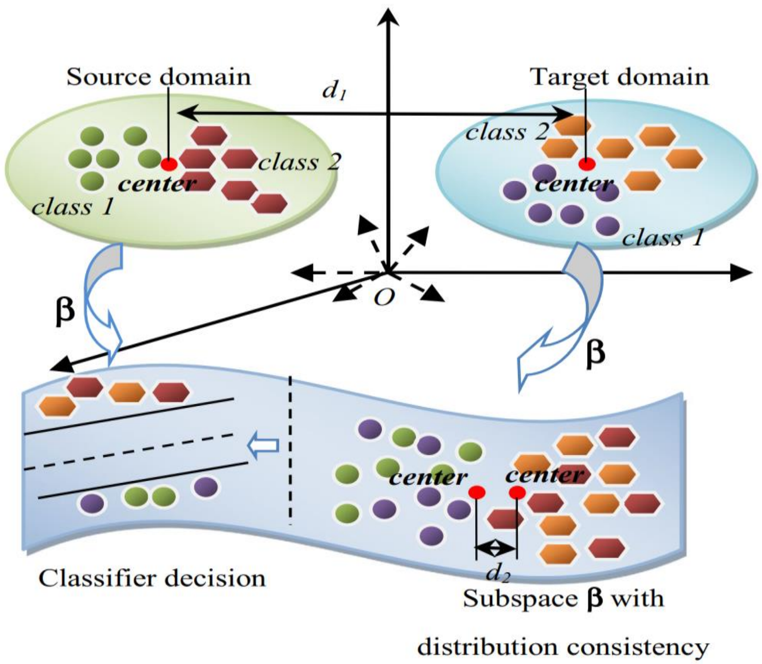
TL Algorithm for E-nose

- Two branches in recent years:
- Classifier transfer (parameter adaptation): *How to learn domain-common classifier adapted to two domains?*

- Feature transfer (subspace adaptation): *How to learn domain-common feature representation (projection) between two domains?*

Cross-domain ELM

- In DAELM model, ELM is used as a classifier (β). Here, we consider it as a feature projector (β).



Step 1: Project data into a unified subspace with discrimination
Step 2: Minimize the distribution shift with mean distribution discrepancy
Step 3: Energy preservation

$$\min_{\beta} \frac{Tr\left(\beta^T S_W^S \beta\right) + \lambda_0 \|\beta\|_F^2 + \lambda_1 \left\| \frac{1}{N_S}\sum_{i=1}^{N_S} \beta^T h_S^i - \frac{1}{N_T}\sum_{j=1}^{N_T} \beta^T h_T^j \right\|_F^2}{Tr\left(\beta^T S_B^S \beta\right) + \lambda_2 Tr\left(\beta^T H_T H_T^T \beta\right)}$$

Final model (objective function)

Y. Liu, L. Zhang, P. Deng and Z. He, "Common Subspace Learning via Cross-domain Extreme Learning Machine," Cognitive Computation, 2017.

Experiment

Gas classification performance on our E-nose data (3 systems)

E-nose System	Formaldehyde	Benzene	Toluene	Carbon monoxide	Nitrogen dioxide	Ammonia	Total
Master	126	72	66	58	38	60	420
Slave 1	108	108	106	98	107	81	608
Slave 2	108	108	94	95	108	84	576

Table 2 Recognition accuracy (%) with sensor calibration under setting 1

Cross-domain tasks	SVM	ELM (sigmoid)	ELM (rbf)	KernelELM	PCA	LDA	LPP	NPE	NCA	MDS	LFDA	SGF	CdELM (sigmoid)	CdELM (rbf)
Master → slave 1	51.97	55	54.59	53.63	55.05	55.56	53.95	53.62	41.28	51.15	61.84	55.10	64.90	**66.39**
Master → slave 2	60.59	59.83	61.16	61.93	60.88	61.09	57.81	54.69	33.85	58.51	61.63	57.49	68.11	**68.45**

Table 3 Recognition accuracy (%) with sensor calibration under setting 2 (task 1)

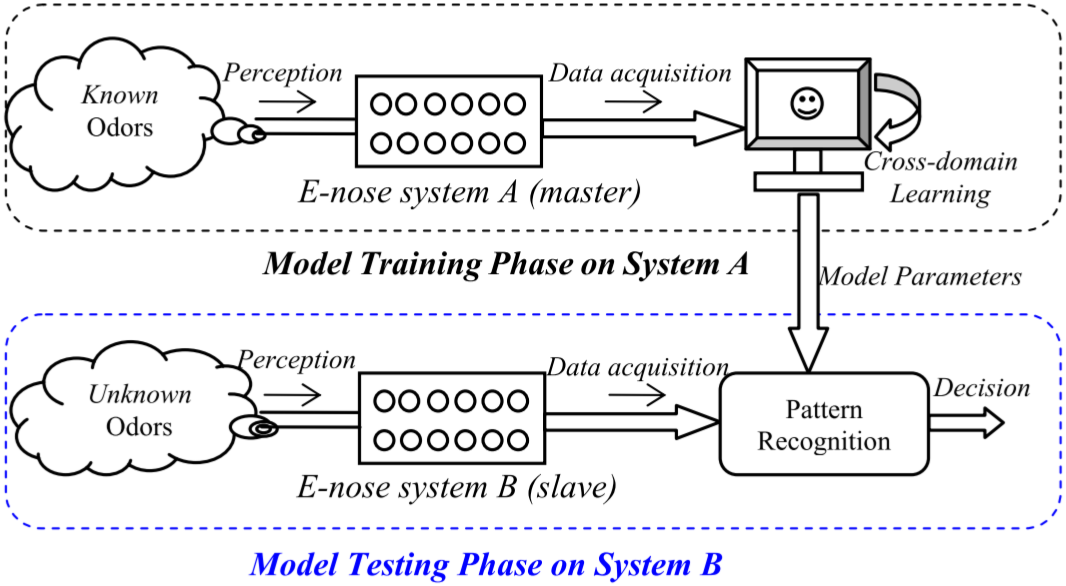
Methods	n_t					Average
	1	3	5	7	9	
SVM	59.14	63.22	62.80	70.49	70.76	65.28
ELM(sigmoid)	59.05	65.10	69.52	71.20	72.18	67.41
ELM(rbf)	62.46	65.83	67.61	69.36	69.87	67.03
PCA	63.92	67.32	70.02	73.36	73.83	69.69
LDA	63.84	67.83	70.33	71.48	73.45	69.39
LPP	65.46	69.83	71.45	72.08	71.48	70.06
NPE	64.78	64.07	63.49	71.55	71.84	67.15
NCA	52.49	50.85	53.81	50.00	63.00	54.03
MDS	61.13	64.75	65.57	70.32	72.92	66.94
LFDA	62.13	67.12	71.63	76.86	74.91	70.53
SGF	66.61	66.95	67.13	70.14	72.74	68.71
CdELM(sigmoid)	67.88	71.53	73.50	**75.34**	76.36	72.92
CdELM(rbf)	**68.34**	**73.32**	**74.42**	75.07	**77.35**	**73.70**

Table 4 Recognition accuracy (%) with sensor calibration under setting 2 (task 2)

Methods	n_t					Average
	1	3	5	7	9	
SVM	69.65	72.76	73.63	74.16	74.90	73.02
ELM(sigmoid)	64.18	67.35	67.91	68.09	69.29	67.36
ELM(rbf)	65.35	68.41	68.63	68.71	69.60	68.14
PCA	65.84	69.27	70.18	72.23	72.15	69.93
LDA	65.23	68.76	69.85	71.74	73.07	69.73
LPP	69.82	74.19	73.63	72.85	76.82	73.46
NPE	71.05	71.15	71.43	72.28	74.14	72.01
NCA	56.32	47.49	52.01	55.62	58.03	53.90
MDS	72.11	73.48	73.81	75.28	75.29	74.00
LFDA	65.26	70.61	73.08	75.47	77.01	72.29
SGF	68.07	71.51	73.08	73.03	73.56	71.85
CdELM(sigmoid)	**74.47**	75.35	75.78	77.56	78.09	76.25
CdELM(rbf)	74.13	**76.06**	**77.69**	**78.95**	**79.18**	**77.21**

Cross-domain Discriminative Subspace Learning (CDSL)

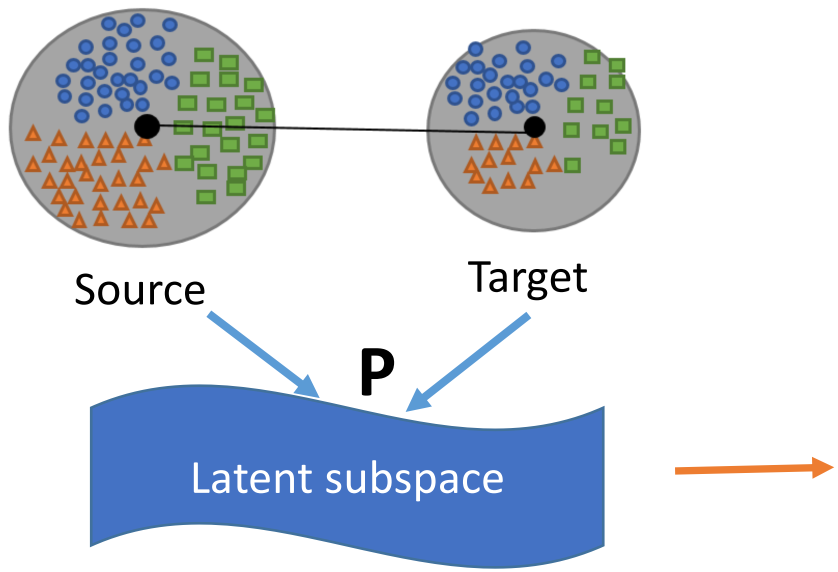
- Feature transfer aims to learn a domain-invariant mapping, such that the domain shift caused by instrumental drift is removed in the latent subspace.



L. Zhang, Y. Liu and P. Deng, "Odor recognition in multiple E-nose systems with cross-domain discriminative subspace learning," IEEE Trans. Instr. Meas., 2017.

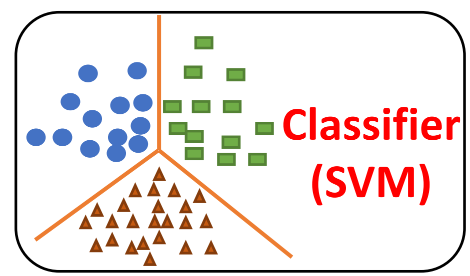
Cross-domain Discriminative Subspace Learning (CDSL)

- How to learn the feature mapping **P** (linear transform)?



$$\min_{\mathbf{w},\xi_i} \ \frac{1}{2}\|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^{N} \xi_i$$

$$\text{s.t. } \xi_i \geq 0, \quad y_i[\mathbf{w}^T \varphi(\mathbf{x}_i) + b] \geq 1 - \xi_i$$



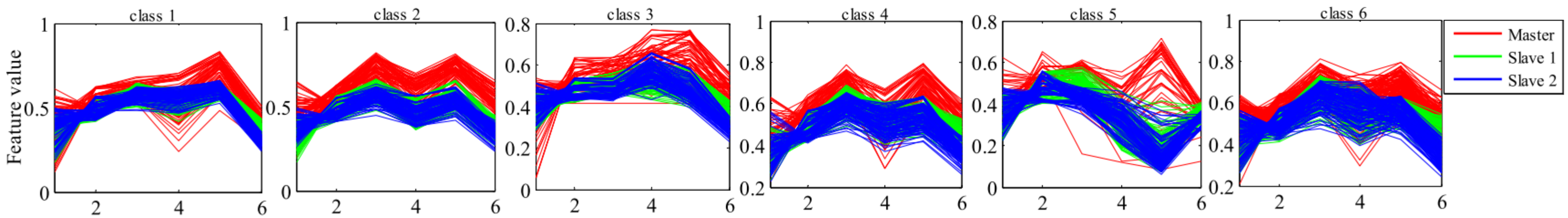
$$\max_{\mathbf{P}} \frac{\mathrm{Tr}\left(\mathbf{P}^T \mathbf{S}_B^S \mathbf{P}\right) + \lambda_0 \cdot \mathrm{Tr}\left(\mathbf{P}^T \mathbf{X}_T \mathbf{X}_T^T \mathbf{P}\right)}{\mathrm{Tr}\left(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}\right) + \lambda_1 \cdot \sum_{i=1}^{M} w_i \left\| \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{P}^T \mathbf{x}_S^{ij} - \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{P}^T \mathbf{x}_T^{ik} \right\|_2^2}$$

L. Zhang, Y. Liu and P. Deng, "Odor recognition in multiple E-nose systems with cross-domain discriminative subspace learning," IEEE Trans. Instr. Meas., 2017.

Experiment

Gas classification performance on our E-nose data (3 systems)

E-nose systems	*Dimensionality*	*Toluene*	*Benzene*	*Ammonia*	*Carbon monoxide*	*Nitrogen dioxide*	*Formaldehyde*	*Total*
Source domain	6	66	72	60	58	38	126	420
Target domain 1	6	106	108	81	98	107	108	608
Target domain 2	6	94	87	84	95	108	108	576



Cross-domain task	SVM	PCA	LDA	LPP	NPE	NCA	MDS	LFDA	GFK	SGF	SA	OSC	DS	GLSW	CDSL
Source domain → target domain 1	51.97	51.97	51.97	53.95	53.62	41.28	51.15	61.84	33.88	55.10	41.10	34.38	45.00	40.46	**71.88**
Source domain → target domain 2	60.59	60.59	56.77	57.81	54.69	33.85	58.51	61.63	32.81	57.49	31.12	36.46	42.62	53.65	**71.88**

10% improvement on the average classification accuracy with domain transfer

Domain Regularized Component Analysis (DRCA)

- **Motivation:** Principle component analysis (PCA) is used for dimension reduction, but only on single domain, and does not work on multiple domains with shift.
- *We therefore consider regularize the PCA with target domain and simultaneously reduce domain shift.*

PCA	+	Regularization	+	Shift reduction
source		target		alignment

L. Zhang, et al., “Anti-drift in E-nose: A Subspace Projection Approach with Drift Reduction,” Sensors and Actuators B: Chemical, 2017.

Domain Regularized Component Analysis (DRCA)

PCA ➕ Regularization ➕ Shift reduction

source target alignment

$$\max_{\mathbf{P}} Tr\left(\mathbf{P}^{\mathrm{T}}\mathbf{X}_S\mathbf{X}_S^{\mathrm{T}}\mathbf{P}\right) + \lambda \cdot Tr\left(\mathbf{P}^{\mathrm{T}}\mathbf{X}_T\mathbf{X}_T^{\mathrm{T}}\mathbf{P}\right)$$

$$\min_{\mathbf{P}} \left\| \frac{1}{N_S}\sum_{i=1}^{N_S}\mathbf{P}^{\mathrm{T}}\mathbf{x}_S^i - \frac{1}{N_T}\sum_{j=1}^{N_T}\mathbf{P}^{\mathrm{T}}\mathbf{x}_T^j \right\|_2^2$$

$$\max_{\mathbf{P}} \frac{Tr\left(\mathbf{P}^{\mathrm{T}}\left(\mathbf{X}_S\mathbf{X}_S^{\mathrm{T}} + \lambda \cdot \mathbf{X}_T\mathbf{X}_T^{\mathrm{T}}\right)\mathbf{P}\right)}{\left\| \frac{1}{N_S}\sum_{i=1}^{N_S}\mathbf{P}^{\mathrm{T}}\mathbf{x}_S^i - \frac{1}{N_T}\sum_{j=1}^{N_T}\mathbf{P}^{\mathrm{T}}\mathbf{x}_T^j \right\|_2^2}$$

L. Zhang, et al., "Anti-drift in E-nose: A Subspace Projection Approach with Drift Reduction," Sensors and Actuators B: Chemical, 2017.

Domain Regularized Component Analysis (DRCA)

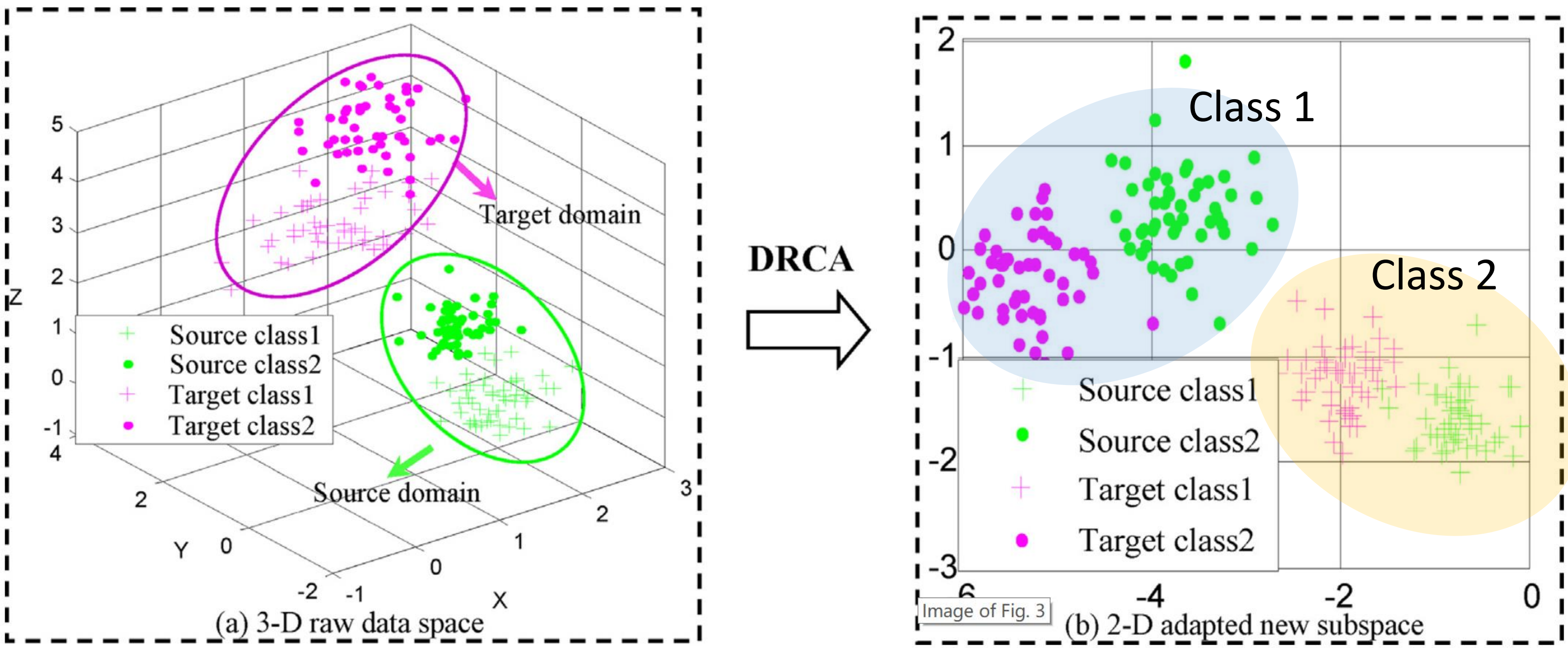


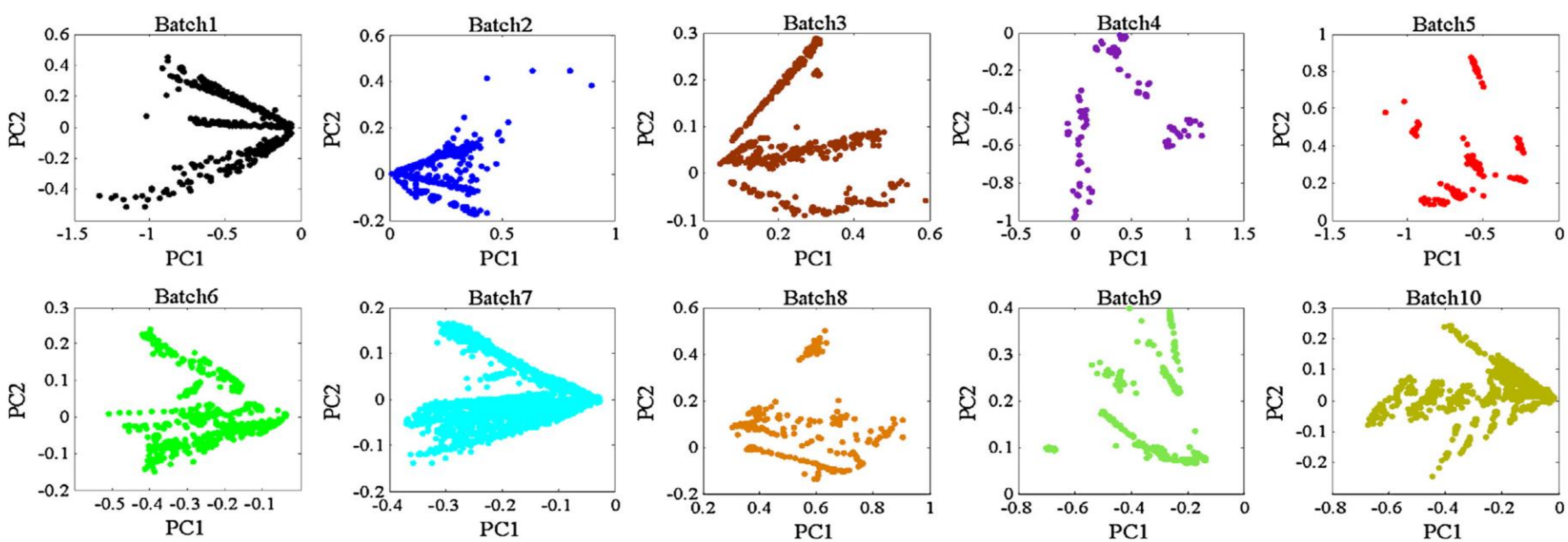
Fig. 3. Subspace adaptation of synthetic data by using the proposed DRCA method.

L. Zhang, et al., "Anti-drift in E-nose: A Subspace Projection Approach with Drift Reduction," Sensors and Actuators B: Chemical, 2017.

Experiment

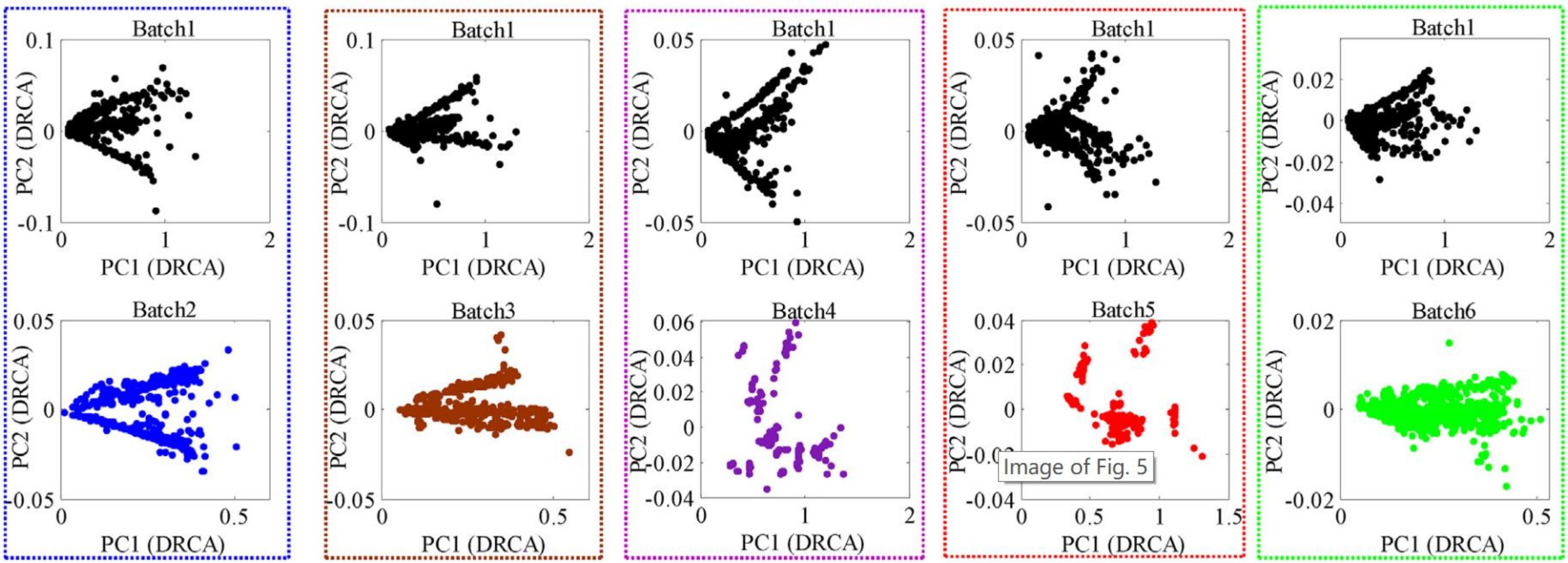
E-nose drift data via (10 batches, 16 gas sensors, acquired from 2008 to 2011 containing 36 months by Vergara et al. 2012):

Batch ID	Month	Acetone	Acetaldehyde	Ethanol	Ethylene	Ammonia	Toluene	Total
Batch 1	1, 2	90	98	83	30	70	74	445
Batch 2	3~10	164	334	100	109	532	5	1244
Batch 3	11, 12, 13	365	490	216	240	275	0	1586
Batch 4	14, 15	64	43	12	30	12	0	161
Batch 5	16	28	40	20	46	63	0	197
Batch 6	17, 18, 19, 20	514	574	110	29	606	467	2300
Batch 7	21	649	662	360	744	630	568	3613
Batch 8	22, 23	30	30	40	33	143	18	294
Batch 9	24, 30	61	55	100	75	78	101	470
Batch 10	36	600	600	600	600	600	600	3600



Experiment

E-nose drift data via (10 batches, 16 gas sensors, acquired from 2008 to 2011 containing 36 months by Vergara et al. 2012):



Task 1: P_{12} Task 2: P_{13} Task 3: P_{14} Task 4: P_{15} Task 5: P_{16}

With drift adaptation by DRCA, the distribution is successfully aligned

Experiment

Recognition Accuracy (%) Under Experimental **Setting 1**.

Batch ID	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6	Batch 7	Batch 8	Batch 9	Batch 10	Average
PCA_{SVM}	82.40	84.80	80.12	75.13	73.57	56.16	48.64	67.45	49.14	68.60
LDA_{SVM}	47.27	57.76	50.93	62.44	41.48	37.42	**68.37**	52.34	31.17	49.91
CC-PCA	67.00	48.50	41.00	35.50	55.00	31.00	56.50	46.50	30.50	45.72
SVM-rbf	74.36	61.03	50.93	18.27	28.26	28.81	20.07	34.26	34.47	38.94
SVM-gfk	72.75	70.08	60.75	75.08	73.82	54.53	55.44	69.62	41.78	63.76
SVM-comgfk	74.47	70.15	59.78	**75.09**	73.99	54.59	55.88	70.23	41.85	64.00
ML-rbf	42.25	73.69	75.53	66.75	77.51	54.43	33.50	23.57	34.92	53.57
ML-comgfk	80.25	74.99	78.79	67.41	**77.82**	**71.68**	49.96	50.79	**53.79**	67.28
ELM-rbf	70.63	66.44	66.83	63.45	69.73	51.23	49.76	49.83	33.50	57.93
OSC	**88.10**	66.71	54.66	53.81	65.13	**63.71**	36.05	40.21	40.08	56.50
GLSW	78.38	69.36	**80.75**	74.62	69.43	44.28	48.64	67.87	46.58	64.43
DS	69.37	46.28	41.61	58.88	48.83	32.83	23.47	**72.55**	29.03	46.98
DRCA	**89.15**	**92.69**	**87.58**	**95.94**	**86.52**	60.25	**62.24**	**72.34**	**52.00**	**77.63**

Recognition Accuracy (%) Under Experimental **Setting 2**.

Batch ID	$1 \to 2$	$2 \to 3$	$3 \to 4$	$4 \to 5$	$5 \to 6$	$6 \to 7$	$7 \to 8$	$8 \to 9$	$9 \to 10$	Average
PCA_{SVM}	82.40	98.87	83.23	72.59	36.70	74.98	58.16	84.04	30.61	69.06
LDA_{SVM}	47.27	46.72	70.81	85.28	48.87	75.15	77.21	62.77	30.25	60.48
SVM-rbf	74.36	87.83	90.06	56.35	42.52	83.53	91.84	62.98	22.64	68.01
SVM-gfk	72.75	74.02	77.83	63.91	70.31	77.59	78.57	86.23	15.76	68.56
SVM-comgfk	74.47	73.75	78.51	64.26	69.97	77.69	82.69	85.53	17.76	69.40
ML-rbf	42.25	58.51	75.78	29.10	53.22	69.17	55.10	37.94	12.44	48.17
ML-comgfk	80.25	**98.55**	84.89	**89.85**	**75.53**	**91.17**	61.22	**95.53**	**39.56**	**79.62**
ELM-rbf	70.63	40.44	64.16	64.37	72.70	80.75	88.20	67.00	22.00	63.36
GLSW	78.38	97.04	81.99	73.60	36.57	74.48	60.54	81.91	26.31	67.87
DS	69.37	53.59	67.08	37.56	36.30	26.57	49.66	42.55	25.78	45.38
DRCA	**89.15**	98.11	**95.03**	69.54	50.87	78.94	65.99	84.04	36.31	**74.22**

Classification accuracy by training SVM on transformed data

Experiment

Gas classification performance on our E-nose data (3 systems)

E-nose System	*DoF*	Formaldehyde	Benzene	Toluene	Carbon monoxide	Nitrogen dioxide	Ammonia	Total
Master (no drift)	6	126	72	66	58	38	60	420
Slave 1 (drift + shift)	6	108	108	106	98	107	81	608
Slave 2 (drift + shift)	6	108	87	94	95	108	84	576

Setting 1 (Drift): Due to that there is also sensor discreteness between slaves and master, the sensor calibration between slaves and master is made by using linear regression according to [38]. Therefore, only drift exists between the source and target data.

Setting 2 (Drift + Shift): The sensor calibration step is omitted, which implies that both the sensor drift and shift exist between the source and target data.

Recognition Accuracy (%) With Sensor Shift Calibration (Setting 1).

Task	SVM	PCA	LDA	GLSW	DS	DRCA
master → slave1	45.89	46.22	42.11	41.45	40.30	**57.07**
master → slave2	31.08	41.84	41.32	48.09	39.76	**52.95**

Recognition Accuracy (%) Without Sensor Shift Calibration (Setting 2).

Task	SVM	PCA	LDA	GLSW	DS	DRCA
master → slave1	51.97	51.97	51.97	47.53	40.46	**58.55**
master → slave2	60.59	60.59	56.77	59.38	40.63	**61.63**

Contents

- Part I: Background and Preliminary
- Part II: Transfer learning: Concept, Theory and Algorithms
- Part III: Transfer Learning Algorithms for E-nose
- Summary vs. Future

Summary vs. Future

In this talk, a systematic introduction of transfer learning and its application for chemical sensors is given.

- Transfer learning plays a vital role in E-noses with robustness.
- Drift compensation still faces a challenge for real application.
- Since pattern recognition is a key unit for E-nose instrument, advanced transfer learning technique is wonderful.
- E-nose faces real application scenario in the future.
- Gas sensors are still the core component of E-nose.

References

1. **Lei Zhang*** and David Zhang, "Domain Adaptation Extreme Learning Machines for Drift Compensation in E-nose Systems," **IEEE Transactions on Instrumentation and Measurement** , vol. 64, no. 7, pp. 1790-1801, July 2015.
2. **Lei Zhang*** and Fengchun Tian, "Performance Study of Multilayer Perceptrons in Low-Cost Electronic Nose,“ **IEEE Transactions on Instrumentation and Measurement**, vol.63, pp. 1670-1679, 2014.
3. **Lei Zhang*** and Pingling Deng, "Abnormal Odor Detection in Electronic Nose via Self-expression Inspired Extreme Learning Machine," **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, vol. 49, no. 10, pp. 1922-1932, 2019.
4. **Lei Zhang*** and David Zhang, "Efficient Solutions for Discreteness, Drift, and Disturbance (3D) in Electronic Olfaction," **IEEE Transactions on Systems, Man, and Cybernetics: Part A**, vol. 48, no. 2, pp. 242-254, 2018.
5. **Lei Zhang***, Yan Liu and Pingling Deng, "Odor Recognition in Multiple E-nose Systems with Cross-domain Discriminative Subspace Learning," **IEEE Transactions on Instrumentation and Measurement**, vol. 66, no. 7, pp. 1679-1692, July 2017.
6. **Lei Zhang***, David Zhang, Xin Yin, and Yan Liu, "A Novel Semi-supervised Learning Approach in Artificial Olfaction for E-Nose Application," **IEEE Sensors Journal**, vol. 16, no. 12, pp. 4919-4931, 2016.
7. **Lei Zhang***, Xuehan Wang, Guang-bin Huang, Tao Liu, and Xiaoheng Tan, "Taste Recognition in E-Tongue using Local Discriminant Preservation Projection," **IEEE Transactions on Cybernetics**, vol. 49, no. 3, pp. 947-960, Mar 2019.
8. **Lei Zhang***, Yan Liu, Zhenwei He, Ji Liu, Pingling Deng, "Anti-Drift in E-nose: A Subspace Projection Approach with Drift Reduction," **Sensors and Actuators B: Chemical**, vol. 253, pp. 407-417, 2017.
9. **Lei Zhang***, Fengchun Tian, Shouqiong Liu, Lijun Dang,Xiongwei Peng, Xin Yin,"Chaotic time series prediction of e-nose sensor drift in embedded phase space," **Sensors and Actuators B: Chemical**, vol.182, pp.71-79,2013.
10. **Lei Zhang**, F.C. Tian*, C. Kadri, B. Xiao, H.J. Li, L. Pan, H. Zhou, "On-Line Sensor Calibration Transfer Among Electronic Nose Instruments for Monitoring Volatile Organic Chemicals in Indoor Air Quality," **Sensors and Actuators B: Chemical**, vol. 160, pp.899-909, 2011.

Thank you for attention

Lei Zhang

School of Microelectronics and Communication Engineering,
Chongqing University, Chongqing, China

Email: leizhang@cqu.edu.cn

Website: https://www.leizhang.tk

Google scholar: https://scholar.google.com/citations?user=Nt9es7kAAAAJ&hl=zh-CN