This paper is accepted for publication in IEEE Sensors Journal, vol.-, no.-, 2016

A Novel Semi-supervised Learning Approach in Artificial Olfaction for E-Nose Application

Lei Zhang, *Member, IEEE*, David Zhang, *Fellow, IEEE*, Xin Yin, and Yan Liu

***Abstract*—Artificial olfaction data is usually represented by a sensor array embedded in an electronic nose system (E-Nose), such that each observation can be expressed as a feature vector for pattern recognition. The concerns of this paper are threefold: *first*, each feature can be represented by multiple different modalities; *second*, manual labeling of sensory data in real application is difficult and hardly impossible, which results in an issue of insufficient labeled data; *third*, classifier learning is generally independent of feature engineering, such that the recognition capability of E-Nose is restricted due to the unilateral suboptimum. Motivated by these concerns, in this paper, from a new perspective of multi-task learning, we aim at proposing a unified semi-supervised learning framework nominated as MFKS, and the merits are composed of three points: 1) A multi-feature joint classifier learning with low-rank constraint is developed for exploiting the structural information of multiple feature modalities. The relatedness of sub-classifiers *w.r.t.* feature modalities is preserved by imposing a low-rank constraint on the group classifier. 2) With a manifold assumption, a Laplacian graph manifold regularization is incorporated for capturing the intrinsic geometry of unlabeled data. 3) The features and classifiers are learned simultaneously in a unified framework, such that the optimality and robustness are improved. Experiments on two datasets including a large-scale 16-sensor data with 36-month drift and a small-scale temperature modulated sensory data demonstrate that the proposed approach has 4% improvement in classification accuracy than others.**

***Index Terms*—Artificial olfactory system, electronic nose, multi-feature learning, semi-supervised learning**

I. Introduction

Numerous progress of artificial olfaction system has been made over the last two decades in many respects such as sensor array, hardware, data processing, and pattern recognition algorithms. In particular, electronic noses, as a typical machine olfaction application, have been widely developed [1]. Briefly, electronic nose (E-Nose) is a multi-sensor system consisting of pattern recognition algorithms and a metal oxide semiconductor sensor array with cross-sensitivity and selectivity. An excellent overview of E-Nose and sensor data processing can be found in [1, 2].

A. Background

The latest progress of artificial olfaction and E-Nose are reviewed in this section. Some good examples are given as follows. Adiguzel and Kulah [3] proposed a breath sensor system for lung cancer diagnosis in medical applications. Sunny *et al* [4] proposed to detect binary gas mixtures by using a thick-film sensor array. Hassan *et al*. [5] proposed a robust rank-order based classifier for gases classification by an E-Nose. Pardo and Sberveglieri [6] proposed a SVM method for E-Nose data classification. Zhang *et al*. [7] proposed a hybrid support vector machine (HSVM) coupled with linear discriminant analysis for classification of multiple indoor air contaminants. Brudzewski *et al*. [8] proposed a novel differential electronic nose for recognition of coffee by using SVM, which has an advantage of automatic sensor baseline drift elimination. Martinelli *et al*. [9] proposed a neuro-adaptive E-Nose for odor recognition, in which the merit is the significant stability of their developed adaptive and unsupervised neural network. Hossein-Babaei and Amini [10] attempted to recognize complex odors by using one tin oxide gas sensor with temperature modulation-alike techniques. Recently, Fonollosa et al. [11] proposed a reservoir computing (RC) algorithm for achieving accurate and continuous prediction to fast varying gas concentrations. These works focus on methodology of artificial olfaction in decision-level.

Another important topic in artificial olfaction is the data processing in feature level. It's known that good features selected from sensors can better represent the odor patterns such that the discriminative performance of the designed E-Nose classifier would be more significant. Researchers have investigated different data processing and feature selection methods. For example, Osuna *et al*. [12] proposed a *k*-nearest neighbor cross-validation estimate for evaluating the performance of data processing. Krutzler *et al*. [13] presented a study how the sensor production influences the recognition performance of an E-Nose. Jha and Yadava [14] proposed a denoising method based on singular value decomposition (SVD) for E-Nose data processing and showed a good performance in feature dimension reduction. Wang *et al*. [15] proposed a novel information-theoretic based feature selection method for human breath-print recognition. Sunil *et al*. [16] proposed a

Manuscript received 20 Mar 2016, accepted 5 Apr 2016. This work was supported in part by National Natural Science Foundation of China (Grant 61401048) and the Research Fund Project of Central Universities.

- L. Zhang is with the College of Communication Engineering, Chongqing University, Chongqing 400044, China and also with Department of Computing, The Hong Kong Polytechnic University, Hong Kong. (e-mail: leizhang@cqu.edu.cn)
- D. Zhang is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: csdzhang@comp.polyu.edu.hk)
- X. Yin and Y. Liu are with the College of Communication Engineering, Chongqing University, Chongqing 400044, China. (e-mail: yxsuixinliutang@sina.com; jolly817@sina.com)

novel sensor selection algorithm for selecting an optimal set of surface acoustic wave (SAW) sensors for E-Nose by similarity metric. Additionally, feature fusion and processing of redundant sensors and features for robust gases classification have also been proposed in the latest works [17, 18, 19].

B. Problem Statement

From the existing work of artificial olfaction and E-Nose, we can observe that different methods in feature-level and classifier-level (model-level) have been proposed by many researchers. Although the discriminative capability has been gradually improved, the data processing methods proposed in feature-level are independent of pattern recognition unit in classifier-level, such that features are not adapted to the classification model if the data structure becomes more complex (e.g. high-dimensional data and sensor drift data) [20]. Briefly, sensor drift is commonly caused by poisoning, aging, environmental variation (ambient temperature, humidity, pressure, etc.), which has deteriorated the performance of E-Nose in both feature and classifier levels [21, 22]. One salient feature of drift is that the conditional probability distribution $P(\mathbf{X}_{no\ drift}|\mathbf{Y}) \neq P(\mathbf{X}_{drift}|\mathbf{Y})$, where $\mathbf{X}_{no\ drift} \in \Re^{d\times N}$ and $\mathbf{X}_{drift} \in \Re^{d\times N}$ denote the data before and after drift, and $\mathbf{Y} \in \Re^{N}$ denotes the labels. The distribution differences also imply that the statistical property differences such as mean and covariance change. Therefore, some independent feature-level data processing methods or model-level classification models would not adapt to the drifted and noisy E-Nose data in real applications due to the temporal data structure variation.

C. Motivation

From the perspective of machine learning, both feature and classifier learning in E-Nose are supervised and enforced separately. However, with the change of data structure caused by complex and long-term E-Nose drift in time series, it is difficult to manually label each observation. Consequently, the model learning in E-Nose would face with a new dilemma that the labeled data is insufficient.

On the one hand, in classifier-level learning, for addressing the issue of insufficient labeled data, semi-supervised learning has been used in artificial olfaction [23, 24]. Semi-supervised learning [25] was initially proposed for handling the problem of insufficient labeled data based on two important assumptions: manifold assumption and cluster assumption [26]. Manifold assumption implies that the data spans a low-dimensional manifold space where the structure information of unlabeled data can be preserved and the nearby data points are more likely to have the same label. The cluster assumption implies that the data points lying in the same cluster are more likely to have the same label. The difference between the two assumptions lies in that cluster assumption is local while manifold assumption is global. Therefore, manifold regularization has become a mainstream of semi-supervised learning in computer vision community for image classification and multimedia application [27, 28]. Laplacian graph manifold was generally proposed for dimension reduction and graph embedding, which implies the manifold structure preservation [29-33]. However, manifold learning suffered from a fact that there is no explicit mapping matrix for low-dimensional projection. Instead of dimension reduction, in this paper, we tend to propose a multi-task semi-supervised learning framework with Laplacian graph manifold regularization for handling noisy E-Nose data.

On the other hand, for feature-level data processing, we aim at proposing multi-feature joint learning by integrating the classifier learning together, which is different from the existing feature selection and feature fusion algorithms that are carried out separately with classifier learning. In other words, the motivation of this paper is to learn the multi-feature and semi-supervised classifier jointly and simultaneously, for the pursuit of the optimal performance in system-level.

D. Paper Contribution

In this paper, with the idea of multi-feature joint learning and Laplacian graph based semi-supervised learning, we propose a multi-feature kernel semi-supervised joint learning model (MFKS) in artificial olfaction system for improving the robustness of an E-Nose. The model, optimization algorithm, convergence and complexity of the proposed MFKS are well formulated theoretically in this paper. To our best knowledge, there is no report of simultaneous multi-feature joint learning and graph based semi-supervised classifier learning in E-Nose community. The contributions of this paper are fourfold.

- From the perspective of multi-task and semi-supervised learning, we propose a unified semi-supervised learning model (MFKS) for classification tasks/scenarios.
- An idea of multi-feature joint learning with low-rank constraint for automatic feature adaptation and structural information sharing is proposed.
- The proposed MFKS method can counteract long-term sensor drift and noise in large-scale olfaction datasets.
- The proposed MFKS can improve the robustness of the temperature-modulated E-Nose system.

E. Paper Organization

This rest of this paper is organized as follows. Section II illustrates the related work for handling drifted E-Nose dataset. Section III presents the model, optimization, convergence and complexity of the proposed MFKS method. The experiments on a large-scale artificial olfaction E-Nose dataset with sensor drift are conducted in Section IV. The experiments on a temperature-modulated E-Nose dataset are conducted in Section V. Finally, Section VI concludes this paper.

II. Related Works

In E-Nose system, machine learning based methods including representation learning, ensemble learning, transfer learning and deep learning have been proposed for drift compensation in very recent years. Specifically, the literature reviews are divided into four categories.

- In representation learning based methods, Ziyatdinov *et al*. [34] proposed a principal component analysis (PCA) method for learning a transformation such that the drift direction can be captured. Carlo *et al*. [35] proposed to learn a data shift transformation matrix by using evolutionary algorithm with

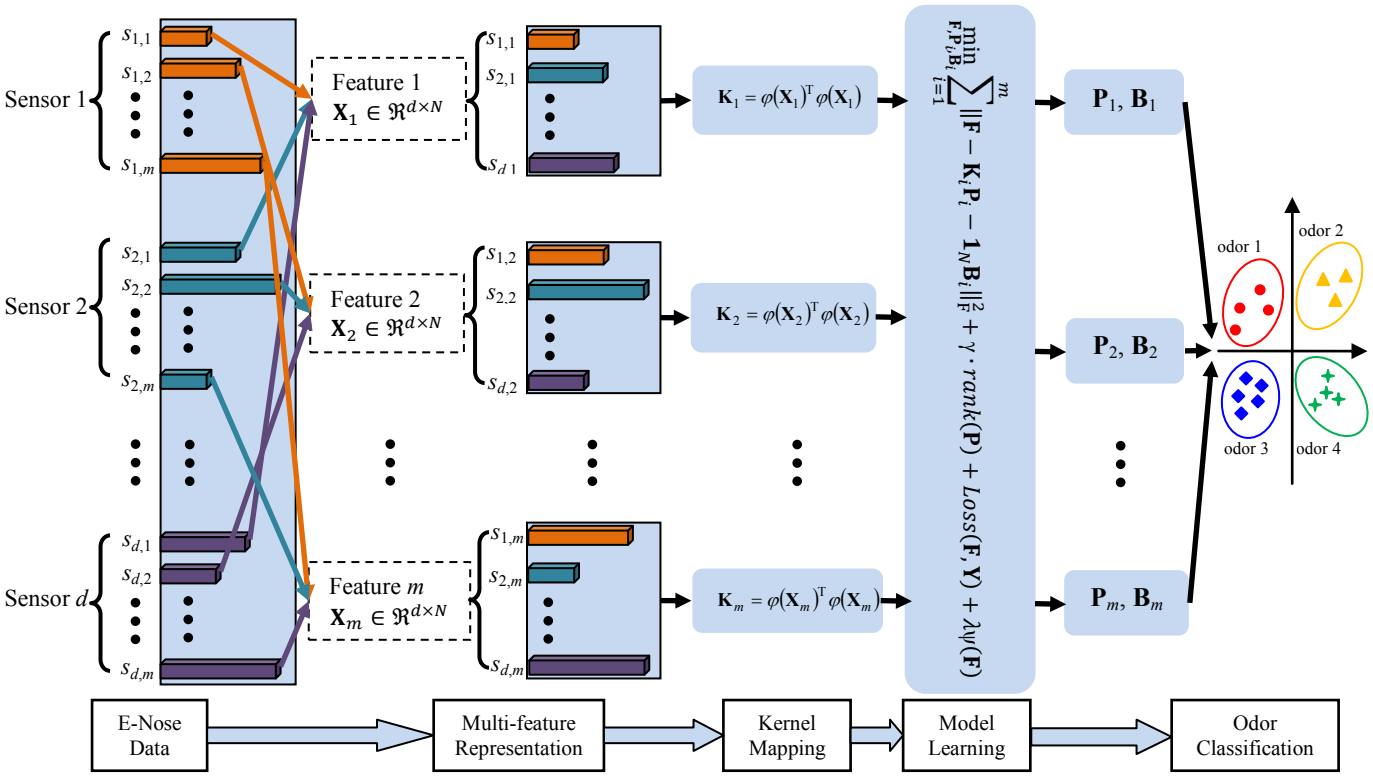

Fig. 1. Overview of the proposed MFKS framework for artificial olfactory system

an objective of improving the classification accuracy of E-Nose. Padilla *et al.* [36] proposed an orthogonal signal correction (OSC) method for drift compensation by removing the components orthogonal to sensor data.

- In ensemble learning based methods, Vergara *et al.* [20], Dang *et al.* [37] and Liu *et al.* [38] proposed classifier ensemble method based on multiple SVM models, that aim at improving the robustness of gases recognition with drift data. However, the weight of each SVM is obtained independent from the model learning, which would result in a local optimal solution. Besides, how to guarantee the diversity of multiple SVM classifiers is still an open problem.
- In transfer learning based methods, Liu *et al.* [39] proposed a novel idea of semi-supervised domain adaptation method for drift compensation in electronic nose. Zhang *et al.* [40, 41] proposed a transfer extreme learning machine based domain adaptation for fast classifier learning and drift compensation. Domain adaptation based transfer learning has been proved to be effective in improving the generalization of classifier with drift-knowledge adaptation.
- In deep learning based methods, Martin *et al.* [42] proposed to learn deep feature in unsupervised manner by using deep restricted Boltzmann machines for bacteria identification in blood by an E-Nose. Similarly, Liu *et al.* [43] proposed for gas recognition with concept drift by using deep learning techniques such as deep Boltzmann machine and sparse auto-encoder. A good advantage of deep learning is the feature representation ability, but depends on a large-scale training data which is still comparatively deficient in E-Nose.

III. Proposed Multi-feature Kernel Semi-supervised Joint Learning Model (MFKS)

A. Notations

Let $\mathbf{X}_i = [\mathbf{X}_i^l, \mathbf{X}_i^u] = [\mathbf{x}_i^1, \mathbf{x}_i^2, \cdots, \mathbf{x}_i^N] \in \Re^{d\times N}, i = 1, \cdots, m$ denotes the training data of the i-th feature modality, where $\mathbf{X}_i^l \in \Re^{d\times N_l}$ and $\mathbf{X}_i^u \in \Re^{d\times N_u}$ represent the labeled data and unlabeled data, respectively, $\mathbf{x}_i^j \in \Re^d$ denotes the j-th observation sample of the i-th feature modality, $N = N_l + N_u$ denotes the number of samples, d denotes the dimension (i.e. the number of sensors) of each observation, and m denotes the number of feature types. Let $\mathbf{Y} = [\mathbf{Y}_l, \mathbf{Y}_u] \in \Re^{C\times N}$ denote the label matrix, where $\mathbf{Y}_j \in \Re^C, j = 1, \cdots, N$ and C denotes the number of classes. $\|\cdot\|_{\mathrm{F}}$, $\|\cdot\|_2$ and $\|\cdot\|_*$ denote the Frobenius norm, ℓ_2-norm and nuclear norm, respectively. $Tr(\cdot)$ denotes the trace operator, $rank(\cdot)$ denotes the rank function, $\langle\cdot\rangle$ denotes inner product operator, and $(\cdot)^{\mathrm{T}}$ denotes the transpose operator. The labels are defined as follows.

- For labeled data, $\mathbf{Y}_l \in \Re^{C\times N_l}$, where $Y_{l,j}^c = 1$ if the j-th sample belongs to class c (c=1,...,C), and -1 otherwise.
- For unlabeled data, $\mathbf{Y}_u = \mathbf{0} \in \Re^{C\times N_u}$.

Considering the nonlinear property of multi-sensor system, the raw data $\mathbf{X}_i$ can be mapped into a high-dimensional space (e.g. reproduced kernel Hilbert space $\mathcal{H}$, RKHS) by using a nonlinear transformation φ. Therefore, we introduce kernel matrix $\mathbf{K}_i \in \Re^{N\times N}, i = 1, \cdots, m$ for representing the training data $\mathbf{X}_i$ in RKHS $\mathcal{H}$ as follows

$$\mathbf{K}_i = \langle \varphi(\mathbf{X}_i), \varphi(\mathbf{X}_i) \rangle = \varphi(\mathbf{X}_i)^{\mathrm{T}} \varphi(\mathbf{X}_i) = \kappa(\mathbf{X}_i, \mathbf{X}_i) \qquad (1)$$

where $\kappa(\cdot)$ denotes the kernel function (e.g. Gaussian RBF, sigmoid, polynomial, etc.). In this paper, the Gaussian RBF function represented as $\kappa(\mathbf{x}_i,\mathbf{x}_j)=\exp\left(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|_2^2/2\sigma^2\right)$ is used. The kernel parameter σ can be tuned.

B. Model Formulation of MFKS

For realizing multi-feature joint learning with m feature kernel matrix in semi-supervised manner, the proposed MFKS is to solve the following general minimization problem.

$$\begin{aligned}&\min_{\mathbf{F},\mathbf{P}_i,\mathbf{B}_i}\sum_{i=1}^{m}\left\|\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+\gamma\cdot rank(\mathbf{P})\\&+Loss(\mathbf{F},\mathbf{Y})+\lambda\cdot\psi(\mathbf{F})\end{aligned}\tag{2}$$

where $\mathbf{P}_i\in\Re^{N\times C}$ denotes the classifier parameter matrix, $\mathbf{B}_i\in\Re^{1\times C}$ denotes the classifier bias, $\mathbf{P}=[\mathbf{P}_1,\mathbf{P}_2,\cdots,\mathbf{P}_m]\in\Re^{N\times mC}$ denotes the group classifier consisting of m sub-classifiers, $\mathbf{1}_N\in\Re^N$ is a full-one vector, $\mathbf{F}\in\Re^{N\times C}$ represents the predicted label matrix, $Loss(\cdot)$ is the least square-alike loss function, γ and λ are regularization parameters.

Specifically, the rational explanations of the four terms behind in the proposed model (2) are presented as follows.

- The first term represents global label prediction based on m feature modalities by simultaneously learning classifier $\mathbf{P}_i$ $(i=1,\cdots,m)$ for each type of feature.
- The second term is the regularization of the learned group classifier $\mathbf{P}$ which is restricted to be low rank, behind the rational is that we expect the relatedness among the learned $\mathbf{P}_1,\mathbf{P}_2,\cdots,\mathbf{P}_m$ to be well preserved by imposing a low-rank constraint, such that the structural information of multiple features can be shared in the proposed model. A good work of low-rank representation is referred to as [47].
- The third term denotes the semi-supervised loss function, formulated based on least-square alike loss as follows

$$Loss(\mathbf{F},\mathbf{Y})=Tr(\mathbf{F}-\mathbf{Y})^{\mathrm{T}}\mathbf{U}(\mathbf{F}-\mathbf{Y})\tag{3}$$

  where $\mathbf{U}$ is diagonal selection matrix defined as follows

$$\mathbf{U}_{jj}=\begin{cases}\infty, \text{if } \mathbf{x}_i^j \text{ is labeled}\\ 0, \text{otherwise}\end{cases};j=1,\cdots,N;i=1,\cdots,m\tag{4}$$

- The last term denotes the manifold regularization on the predicted label $\mathbf{F}$ for exploring label consistency in semi-supervised learning. With manifold assumption, we expect that nearby points are more likely to have the same labels, which can be mathematically formulated as follows,

$$\begin{aligned}&\min_f\sum_{p,q}A_{p,q}\left\|f(\mathbf{x}_p)-f(\mathbf{x}_q)\right\|_2^2\\&=\min_f\sum_{p,q}\left(\left\|f(\mathbf{x}_p)\right\|_2^2+\left\|f(\mathbf{x}_q)\right\|_2^2-2f(\mathbf{x}_p)f(\mathbf{x}_q)^{\mathrm{T}}\right)A_{p,q}\\&=\min_f\sum_p\left(\left\|f(\mathbf{x}_p)\right\|_2^2\sum_q A_{p,q}\right)+\sum_q\left(\left\|f(\mathbf{x}_q)\right\|_2^2\sum_p A_{p,q}\right)\\&\qquad-2\sum_{p,q}f(\mathbf{x}_p)f(\mathbf{x}_q)^{\mathrm{T}}A_{p,q}\\&=\min_{\mathbf{F}}Tr(2\mathbf{F}^{\mathrm{T}}\mathbf{D}\mathbf{F})-Tr(2\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F})\\&=\min_{\mathbf{F}}Tr(2\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F})\end{aligned}\tag{5}$$

where $f(\cdot)$ denotes the label predictor, $\mathbf{D}$ is a diagonal matrix with entries $D_{pp}=\sum_q A_{pq}$, $\mathbf{L}=\mathbf{D}-\mathbf{A}$ is the Laplacian graph matrix, and $\mathbf{A}$ is the affinity matrix computed as

$$A_{p,q}=\begin{cases}1, \text{if } \mathbf{x}_p\in N_k(\mathbf{x}_q) \text{or } \mathbf{x}_q\in N_k(\mathbf{x}_p)\\ 0, \text{otherwise}\end{cases}\tag{6}$$

where $N_k(\mathbf{x})$ represents the k nearest neighbors of sample $\mathbf{x}$.

Therefore, in terms of Eq.(5), the last term $\psi(\mathbf{F})$ in the proposed model (2) can be written as follows

$$\psi(\mathbf{F})=Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right)\tag{7}$$

where the Laplacian matrix $\mathbf{L}=\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i$ is represented as weighted summation of $\mathbf{L}_i$ for multi-feature learning, $0<\alpha_i^r<1$ represents the coefficient of the i-th feature, $\sum_{i=1}^{m}\alpha_i=1$, and $r>1$. Note that the setting of $r>1$ is to better exploit the complementary information of multiple features and avoid that trivial solution with only the best feature used (e.g. $\alpha_i=1$).

In summary, by combining (2), (3) and (7), the proposed MFKS model can be completely written as follows

$$\begin{aligned}&\min_{\mathbf{F},\mathbf{P}_i,\mathbf{B}_i,\alpha_i}\sum_{i=1}^{m}\left\|\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+\gamma\cdot rank(\mathbf{P})\\&+Tr(\mathbf{F}-\mathbf{Y})^{\mathrm{T}}\mathbf{U}(\mathbf{F}-\mathbf{Y})+\lambda\cdot Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right)\\&\text{s.t.}\sum_{i=1}^{m}\alpha_i=1,0<\alpha_i<1\end{aligned}\tag{8}$$

Note that the *rank* of $\mathbf{P}$ is a non-convex operator, and it is generally addressed by using its convex surrogate of the rank [47], that is, the nuclear norm or trace-norm defined as $\|\mathbf{P}\|_*$ (i.e. the sum of singular values of $\mathbf{P}$), represented as

$$\left\|\mathbf{P}\right\|_*=Tr\left(\mathbf{P}^{\mathrm{T}}\left(\mathbf{P}\mathbf{P}^{\mathrm{T}}\right)^{-\frac{1}{2}}\mathbf{P}\right)\tag{9}$$

Finally, by substituting (9) into (8), the proposed MFKS model can be reformulated as follows.

$$\begin{aligned}&\min_{\mathbf{F},\mathbf{P}_i,\mathbf{B}_i,\alpha_i}\sum_{i=1}^{m}\left\|\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+\frac{\gamma}{2}\cdot\left\|\mathbf{P}\right\|_*\\&+Tr(\mathbf{F}-\mathbf{Y})^{\mathrm{T}}\mathbf{U}(\mathbf{F}-\mathbf{Y})+\lambda\cdot Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right)\\&\text{s.t.}\sum_{i=1}^{m}\alpha_i=1,0<\alpha_i<1\end{aligned}\tag{10}$$

C. Model Optimization and Algorithm Solver

From the structure of the proposed MFKS model (10), we observe that the objective function is non-convex *w.r.t.* four variables $\mathbf{F}$, $\mathbf{P}_i$, $\mathbf{B}_i$ and α_i. However, when fix other three variables (e.g. $\mathbf{P}_i$, $\mathbf{B}_i$ and α_i), it is convex *w.r.t.* another variable (i.e. $\mathbf{F}$). Therefore, the solutions $\mathbf{F}$, $\mathbf{P}_i$, $\mathbf{B}_i$ and α_i can be efficiently solved by a variable-alternative optimization approach. Specifically, the optimization process is as follows.

- *Initialize* $\mathbf{F}^{(0)}$, $\alpha_i^{(0)}$, $\mathbf{P}_i^{(0)}$ *and* $\mathbf{B}_i^{(0)}$

First, we initialize $\mathbf{P}_i^{(0)}=\mathbf{0}$ and $\mathbf{B}_i^{(0)}=\mathbf{0}$. With $\sum_{i=1}^{m}\alpha_i=1$, we initialize $\alpha_i^{(0)}=1/m,\forall\, i$, the initialized $\mathbf{F}$ can be solved analytically by setting the derivative of (11) *w.r.t.* $\mathbf{F}$ as 0,

$$\min_{\mathbf{F}} Tr\left(\mathbf{F}-\mathbf{Y}\right)^{\mathrm{T}}\mathbf{U}\left(\mathbf{F}-\mathbf{Y}\right)+\lambda\cdot Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right) \quad (11)$$

Then, $\mathbf{F}$ can be initialized as

$$\mathbf{F}^{(0)}=\left(\mathbf{U}+\lambda\cdot\sum_{i=1}^{m}\left(\alpha_i^{(0)}\right)^r\mathbf{L}_i\right)^{-1}\mathbf{U}\mathbf{Y} \quad (12)$$

- *Update* $\mathbf{P}_i^{(t)}$ *and* $\mathbf{B}_i^{(t)}$

Second, after fixing $\mathbf{F}$ and α_i, the optimization problem of model (10) becomes

$$\begin{aligned}&\min_{\mathbf{P}_i,\mathbf{B}_i}\sum_{i=1}^{m}\left\|\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+\frac{\gamma}{2}\cdot\left\|\mathbf{P}\right\|_*\\&=\min_{\mathbf{P}_i,\mathbf{B}_i}\sum_{i=1}^{m}\left\|\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+\frac{\gamma}{2}\cdot Tr\left(\mathbf{P}^{\mathrm{T}}\left(\mathbf{P}\mathbf{P}^{\mathrm{T}}\right)^{-\frac{1}{2}}\mathbf{P}\right)\end{aligned} \quad (13)$$

By setting the derivative of the objective function $J\left(\mathbf{P}_i,\mathbf{B}_i\right)$ in (13) *w.r.t.* $\mathbf{P}_i$ and $\mathbf{B}_i$ to be 0, respectively, one can obtain

$$\begin{aligned}&\frac{\partial J\left(\mathbf{P}_i,\mathbf{B}_i\right)}{\partial\mathbf{P}_i}=-\mathbf{K}_i^{\mathrm{T}}\left(\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right)+\frac{\gamma}{2}\left(\mathbf{P}\mathbf{P}^{\mathrm{T}}\right)^{-\frac{1}{2}}\mathbf{P}_i=0\\&\frac{\partial J\left(\mathbf{P}_i,\mathbf{B}_i\right)}{\partial\mathbf{B}_i}=-\mathbf{1}_N^{\mathrm{T}}\left(\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right)=0\end{aligned} \quad (14)$$

Therefore, for iteration t, $\mathbf{P}_i^{(t)}$and $\mathbf{B}_i^{(t)}$ *w.r.t.* the i-th feature can be solved in close-form as follows

$$\begin{aligned}&\mathbf{P}_i^{(t)}=\left(\mathbf{K}_i^{\mathrm{T}}\mathbf{K}_i+\gamma\cdot\mathbf{D}_p^{(t-1)}\right)^{-1}\left(\mathbf{K}_i^{\mathrm{T}}\mathbf{F}^{(t-1)}-\mathbf{K}_i^{\mathrm{T}}\mathbf{1}_N\mathbf{B}_i^{(t-1)}\right)\\&\mathbf{B}_i^{(t)}=\frac{\mathbf{1}_N^{\mathrm{T}}\mathbf{F}-\mathbf{1}_N^{\mathrm{T}}\mathbf{K}_i\mathbf{P}_i^{(t)}}{\mathbf{1}_N^{\mathrm{T}}\mathbf{1}_N}\end{aligned} \quad (15)$$

where $\mathbf{D}_p^{(t-1)}=0.5\left(\mathbf{P}^{(t-1)}\left(\mathbf{P}^{(t-1)}\right)^{\mathrm{T}}\right)^{-0.5}$ is a diagonal matrix and $\mathbf{P}^{(t-1)}=\left[\mathbf{P}_1^{(t-1)},\cdots,\mathbf{P}_m^{(t-1)}\right]$ is the group classifier.

- *Update* $\mathbf{F}^{(t)}$

Third, after fixing α_i, $\mathbf{P}_i$ and $\mathbf{B}_i$, the optimization problem (10) becomes

$$\begin{aligned}\min_{\mathbf{F}}\sum_{i=1}^{m}\left\|\mathbf{F}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+Tr\left(\mathbf{F}-\mathbf{Y}\right)^{\mathrm{T}}\mathbf{U}\left(\mathbf{F}-\mathbf{Y}\right)\\+\lambda\cdot Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right)\end{aligned} \quad (16)$$

From (16), it is easy to solve $\mathbf{F}$ by setting the derivative *w.r.t.* $\mathbf{F}$ to be 0, as follows

$$\begin{aligned}\frac{\partial J\left(\mathbf{F}\right)}{\partial\mathbf{F}}&=\left(m\mathbf{I}+\mathbf{U}+\lambda\cdot\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}-\mathbf{U}\mathbf{Y}-\sum_{i=1}^{m}\left(\mathbf{K}_i\mathbf{P}_i+\mathbf{1}_N\mathbf{B}_i\right)\\&=0\end{aligned} \quad (17)$$

where $\mathbf{I}\in\Re^{N\times N}$ is an identity matrix. Then, from (17), $\mathbf{F}^{(t)}$ in iteration t can be updated as

$$\mathbf{F}^{(t)}=\left(m\mathbf{I}+\mathbf{U}+\lambda\cdot\sum_{i=1}^{m}\left(\alpha_i^{(t-1)}\right)^r\mathbf{L}_i\right)^{-1}\left(\begin{array}{l}\mathbf{U}\mathbf{Y}+\\\sum_{i=1}^{m}\left(\mathbf{K}_i\mathbf{P}_i^{(t)}+\mathbf{1}_N\mathbf{B}_i^{(t)}\right)\end{array}\right) \quad (18)$$

- *Update* $\alpha_i^{(t)}$

Fourth, after obtaining $\mathbf{F}$, $\mathbf{P}_i$ and $\mathbf{B}_i$, the optimization problem (10) becomes the following minimization

Algorithm 1. MFKS

Input:

The training data of m feature matrix $\mathbf{X}_i\in\mathbb{R}^{d\times N},i=1,\cdots,m$;

The training labels $\mathbf{Y}\in\mathbb{R}^{C\times N}$;

Parameters λ, γ, and r;

Procedure:

1. Compute the kernel matrix $\mathbf{K}_i$ using (1);
2. Compute the graph Laplacian matrix $\mathbf{L}_i$ with (5) and (6);
3. Compute the selection matrix $\mathbf{U}$ using (4);
4. Initialize $\alpha_i^{(0)}\leftarrow 1/m$, $\mathbf{P}_i^{(0)}\leftarrow\mathbf{0}$, $\mathbf{B}_i^{(0)}\leftarrow\mathbf{0},i=1,\cdots,m$;
5. Initialize $\mathbf{P}^{(0)}=\left[\mathbf{P}_1^{(0)},\cdots,\mathbf{P}_m^{(0)}\right]$ and $\mathbf{D}_p^{(0)}=\frac{1}{2}\left(\mathbf{P}^{(0)}\left(\mathbf{P}^{(0)}\right)^{\mathrm{T}}\right)^{-\frac{1}{2}}$;
6. Initialize $\mathbf{F}^{(0)}$ according to (12);
7. **repeat**

   Compute $\mathbf{P}_i^{(t)}$ and $\mathbf{B}_i^{(t)}$ according to (15);

   Update $\mathbf{F}^{(t)}$ according to (18);

   Update $\alpha_i^{(t)}$ according to (22);

   Update $\mathbf{P}^{(t)}=\left[\mathbf{P}_1^{(t)},\cdots,\mathbf{P}_m^{(t)}\right]$ and $\mathbf{D}_p^{(t)}=\frac{1}{2}\left(\mathbf{P}^{(t)}\left(\mathbf{P}^{(t)}\right)^{\mathrm{T}}\right)^{-\frac{1}{2}}$

   until convergence;
8. **Output:** $\mathbf{P}_i$ and $\mathbf{B}_i$, i=1,…,m;

$$\begin{aligned}&\min_{\alpha_i} Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right)\\&\text{s.t.}\sum_{i=1}^{m}\alpha_i=1,0<\alpha_i<1\end{aligned} \quad (19)$$

The Lagrange equation of (19) can be written as

$$J\left(\alpha,\mu\right)=Tr\left(\mathbf{F}^{\mathrm{T}}\left(\sum_{i=1}^{m}\alpha_i^r\mathbf{L}_i\right)\mathbf{F}\right)-\mu\cdot\left(\sum_{i=1}^{m}\alpha_i-1\right) \quad (20)$$

By setting the derivative of (20) *w.r.t.* α_i and μ to 0, respectively, we have

$$\begin{cases}r\alpha_i^{r-1}Tr\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}_i\mathbf{F}\right)-\mu=0\\\sum_{i=1}^{m}\alpha_i-1=0\end{cases} \quad (21)$$

By solving the equation group (21), $\alpha_i^{(t)},i=1,\cdots,m$ in iteration t can be easily updated as

$$\alpha_i^{(t)}=\left(\frac{1}{Tr\left(\left(\mathbf{F}^{(t)}\right)^{\mathrm{T}}\mathbf{L}_i\mathbf{F}^{(t)}\right)}\right)^{\frac{1}{r-1}}\Bigg/\sum_{i=1}^{m}\left(\frac{1}{Tr\left(\left(\mathbf{F}^{(t)}\right)^{\mathrm{T}}\mathbf{L}_i\mathbf{F}^{(t)}\right)}\right)^{\frac{1}{r-1}} \quad (22)$$

The detail of solving Eq. (21) is illustrated in **Appendix**.

Finally, the optimization of the proposed MFKS is illustrated as **Algorithm 1.**

D. Classification

The classification denotes the testing phase after MFKS training. The classification parameters $\{\mathbf{P}_i\}_{i=1}^m$ and $\{\mathbf{B}_i\}_{i=1}^m$ can be obtained by solving the proposed model (10) using the proposed **Algorithm 1** on the provided training data $\mathbf{X}=\{\mathbf{X}_i\}_{i=1}^m$. For real-time application, we perform classification and predict the label of a new observation $\mathbf{z}$ represented by m features $\mathbf{z}=\{\mathbf{z}_i\}_{i=1}^m$ using the following decision function,

$$label\left(\mathbf{z}\right)=\arg\max_{j\in\{1,..,C\}}\sum_{i=1}^{m}\left(\kappa\left(\mathbf{z}_i,\mathbf{X}_i\right)\cdot\mathbf{P}_i+\mathbf{B}_i\right) \quad (23)$$

Algorithm 2. Classification of MFKS

Input:

Training set $\{\mathbf{X}_i\}_{i=1}^m$, training labels $\mathbf{Y}$, and one test sample $\{\mathbf{z}_i\}_{i=1}^m$ with m kinds of features;

Procedure:

1. Obtain the optimal $\{\mathbf{P}_i\}_{i=1}^m$ and $\{\mathbf{B}_i\}_{i=1}^m$ by solving the proposed MFKS model (10) using **Algorithm 1**.
2. Compute the kernel mapping $\mathbf{K}_{\mathbf{z}_i} = \kappa(\mathbf{z}_i, \mathbf{X}_i)$ of $\{\mathbf{z}_i\}_{i=1}^m$.

Output:

$identity(\mathbf{z}) \leftarrow \arg\max_{j\in\{1,\cdots,C\}} \left[\sum_{i=1}^m \left(\mathbf{K}_{\mathbf{z}_i} \cdot \mathbf{P}_i + \mathbf{B}_i\right)\right]_j$

where $\kappa(\cdot)$ is the kernel function. The classification procedure of the proposed MFKS is summarized in **Algorithm 2**.

E. Convergence

To explore the convergence behavior of the proposed Algorithm 1, we first provide a lemma as follows.

Lemma 1: *For alternative optimization, when update one variable with other variables fixed, it will not increase the objective function value. Four claims with short proofs are given as follows:*

Claim 1. $J\left(\mathbf{P}_i^{(t)}, \mathbf{B}_i^{(t)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right) \geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right)$

Proof. When fix $\mathbf{B}_i^{(t)}$, $\mathbf{F}^{(t)}$, $\alpha_i^{(t)}$ and update $\mathbf{P}_i^{(t+1)}$, the objective function is convex *w.r.t.* $\mathbf{P}_i$. As shown in (14) by setting the derivative of the objective function *w.r.t.* $\mathbf{P}_i$ to be 0, then it's clear that the expression of *claim 1* holds.

Claim 2. $J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right) \geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right)$

Proof. Similar to the proof of *claim 1*, the objective function becomes convex *w.r.t.* $\mathbf{B}_i$ when fix $\mathbf{P}_i^{(t+1)}$, $\mathbf{F}^{(t)}$, $\alpha_i^{(t)}$ and update $\mathbf{B}_i^{(t+1)}$. Then, *claim 2* is proven.

Claim 3. $J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right) \geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t+1)}, \alpha_i^{(t)}\right)$

Proof. When $\mathbf{P}_i^{(t+1)}$, $\mathbf{B}_i^{(t+1)}$, $\alpha_i^{(t)}$ are fixed, and update $\mathbf{F}^{(t)}$, the optimization problem becomes (16) which is convex *w.r.t.* $\mathbf{F}$. By setting the derivative of the objective function (16) *w.r.t.* $\mathbf{F}$ to be 0, its solution in (17) makes *claim 3* hold.

Claim 4. $J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t+1)}, \alpha_i^{(t)}\right) \geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t+1)}, \alpha_i^{(t+1)}\right)$

Proof. As can be seen from (21), with $\mathbf{P}_i^{(t+1)}$, $\mathbf{B}_i^{(t+1)}$, $\mathbf{F}^{(t+1)}$ fixed, the update rule of $\alpha_i^{(t+1)}$ is obtained by setting the derivatives of objective function (20) *w.r.t.* α_i to be 0. Also, since the second-order derivative *w.r.t.* α_i are positive with the condition that $r > 1, \alpha_i > 0$, i.e. there is

$$\frac{\partial^2 J(\alpha_i, \mu)}{\partial \alpha_i^2} = r(r-1)\alpha_i^{r-2} Tr\left(\mathbf{F}^{\mathrm{T}} \mathbf{L}_i \mathbf{F}\right) > 0$$

Thus, the update rule (22) of α_i can make the objective function (20) decrease, and *claim 4* is proven. In summary, the convergence of the proposed **Algorithm 1** is summarized as the following theorem.

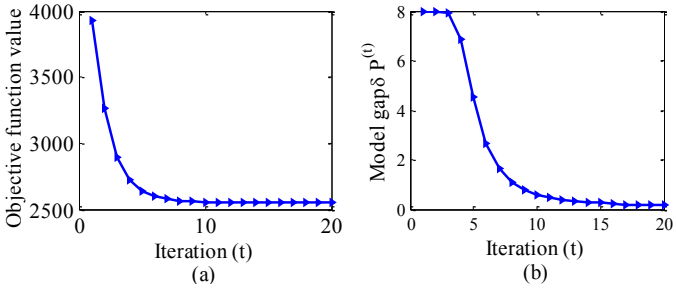

Fig. 2. Convergence analysis for optimality condition

Theorem 1: *The objective function (10) monotonically decreases until convergence after several iterations by using Algorithm 1.*

Proof. Suppose the updated $\mathbf{P}_i^{(t)}$, $\mathbf{B}_i^{(t)}$, $\mathbf{F}^{(t)}$, and $\alpha_i^{(t)}$ are $\mathbf{P}_i^{(t+1)}$, $\mathbf{B}_i^{(t+1)}$, $\mathbf{F}^{(t+1)}$, and $\alpha_i^{(t+1)}$, respectively. In terms of the *claim 1*, *claim 2*, *claim 3* and *claim 4* presented in **Lemma 1**, we have the following inequalities,

$$\begin{aligned} J\left(\mathbf{P}_i^{(t)}, \mathbf{B}_i^{(t)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right) &\geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right) \\ &\geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t)}, \alpha_i^{(t)}\right) \\ &\geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t+1)}, \alpha_i^{(t)}\right) \\ &\geq J\left(\mathbf{P}_i^{(t+1)}, \mathbf{B}_i^{(t+1)}, \mathbf{F}^{(t+1)}, \alpha_i^{(t+1)}\right) \end{aligned}$$

Therefore, the **Theorem 1** is proven.

F. Computational Complexity

We briefly analyze the computational complexity of the MFKS involving T iterations and m modalities. Before stepping into the learning phase, the complexity of computing the Laplacian matrices is $O(mN^3)$. In learning, each iteration involves three update steps, and the complexity in T iterations is $O(m^2N^2T)$. Hence, the computational complexity of our method is $O(mN^3)+O(m^2N^2T)$. Note that the Laplacian matrices are not involved in iterations and therefore the computational complexity of computing Laplacian matrix is $O(mN^3)$.

G. Remarks on Optimality Condition

The convergence of MFKS shown in **Algorithm 1** is demonstrated in **Theorem 1**. For insight of the optimality condition, the objective function value of (10) over 20 iterations on E-nose dataset used in this paper for six kinds of gases recognition is described in Fig. 2(a). One can observe that after 10 iterations, the algorithm can converge to a stable value. Furthermore, the classifier gap $\delta\mathbf{P}^{(t)}$ is introduced for optimality condition check. The classifier gap is defined as

$$\delta\,\mathbf{P}^{(t)} = \sum_{i=1}^m \left\|\mathbf{P}_i^{(t)} - \mathbf{P}_i^{(t-1)}\right\|_{\mathrm{F}} \Big/ \left\|\mathbf{P}_i^{(t)}\right\|_{\mathrm{F}} \tag{24}$$

The convergence of $\delta\mathbf{P}^{(t)}$ over 20 iterations is described in Fig. 2(b). It is clearly seen that model gap can quickly converge to a small and stable value after 10 iterations. Therefore, we can conclude that the optimality condition of the proposed MFKS model can be achieved during 10 iterations by using the proposed optimization **Algorithm 1**.

From the viewpoint of machine learning, the proposed semi-supervised MFKS model and algorithm can also be used

TABLE I
LONG-TERM DRIFT DATASET

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

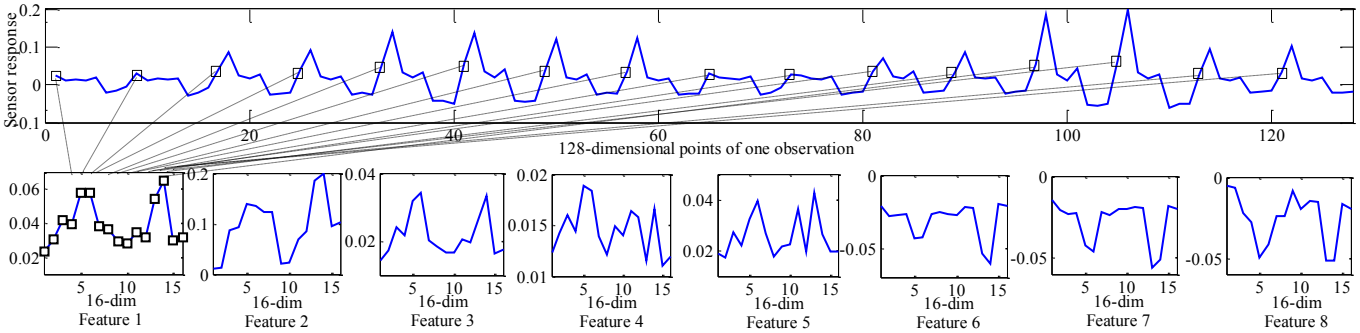

Fig. 3. Multi-feature representation of one observation in batch 1 with 16 sensors under Acetone. Each square denotes the first feature of each sensor.

for supervised learning by setting **F**=**Y** in Eq.(10), such that the first two terms are kept for purely supervised learning, i.e.

$$\min_{\mathbf{P}_i,\mathbf{B}_i} \sum_{i=1}^{m} \left\|\mathbf{Y}-\mathbf{K}_i\mathbf{P}_i-\mathbf{1}_N\mathbf{B}_i\right\|_{\mathrm{F}}^2+\frac{\gamma}{2}\cdot\left\|\mathbf{P}\right\|_* \qquad (25)$$

From (25), the supervised MFKS can be recognized as a multi-feature learning framework with low-rank constraint on the group classifier **P**.

IV. EXPERIMENTS ON DRIFTED E-NOSE DATA

In this section, the recently released E-nose olfactory dataset with drift in UCI Machine Learning Repository [20, 44, 48] is used to evaluate the proposed MFKS model and algorithm.

A. Experimental Data

The artificial olfaction dataset contains 13,910 measurements (observation samples) of 6 kinds of gases including acetone, acetaldehyde, ethanol, ethylene, ammonia, and toluene at different concentration levels. The dataset is divided into 10 batches versus time. The dataset was sampled by an E-Nose system with 16 gas sensors. For each sensor, 8 kinds of features were extracted from each sensor (i.e. m=8, in this paper). We refer interested readers to as [20, 48] for details on how to determine the 8 features. Fig.3 illustrates the normalized sensor response and the multi-feature formulation in this paper. For each feature type, a 16-dimensional feature vector of each measurement is formulated consequently. The details of the dataset are presented in Table I, which shows the number of samples of each gas.

B. Experimental Setup

For evaluating the proposed semi-supervised learning framework, we adopt the following experimental setting.

Semi-supervised setting: Take batch 1 as fixed training set and tested on batch K, K=2,…,10.

Description: in semi-supervised setting, the batch 1 is used as labeled training data, while the data from batch K (K=2,…,10) is used as unlabeled data for semi-supervised learning. Besides, we randomly select L=10, 20, 30, 40, and 50 labeled data from batch K (K=2,…,10), respectively, and the remaining are still unlabeled for semi-supervised learning.

C. Parameter Setting

In default, the RBF kernel function is used for kernel matrix computing. For finding out the best results, the kernel parameter σ is adjusted from the given set $\{2^{-4},2^{-3},\cdots,2^{4}\}$ and the regularization coefficient γ is adjusted from the given set $\{10^{-4},10^{-3},\cdots,10\}$. The regularization coefficient of graph manifold regularization λ is set as 1. The number of maximum iterations in MFKS model is set as 10 throughout the paper. The raw sensor features are normalized appropriately within the interval (-1, 1) for easier convergence in optimization.

D. Compared Methods

Following the semi-supervised experimental setting and parameter setting, we compare the proposed semi-supervised MFKS model with three SVM methods including RBF kernel based SVM (SVM-rbf), geodesic flow kernel based SVM (SVM-gfk) and combination kernel based SVM (SVM-comgfk), two semi-supervised methods including RBF kernel based manifold regularization (ML-rbf) and combination geodesic flow kernel based manifold regularization (ML-comgfk) [39], and extreme learning machine (ELM) based on RBF function (ELM-rbf) [45].

E. Experimental Results

- **Performance comparisons**

The recognition accuracies of compared methods including SVM-rbf, SVM-gfk, SVM-comgfk, ML-rbf, ML-comgfk, ELM-rbf, and the proposed MFKS on the testing data from the

TABLE II
COMPARISONS OF RECOGNITION ACCURACY (%) WITH SEMI-SUPERVISED SETTING

Batch ID	Batch 2	Batch 3	Batch 4	Batch 5	Batch 6	Batch 7	Batch 8	Batch 9	Batch 10	Average
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
MFKS (10)	**80.79**	**80.64**	**86.75**	**79.14**	**80.69**	36.19	**68.30**	63.04	37.10	**68.07**
MFKS (20)	**85.45**	**77.96**	**88.65**	**83.61**	**89.38**	**68.80**	**84.67**	**78.66**	**42.54**	**77.75**

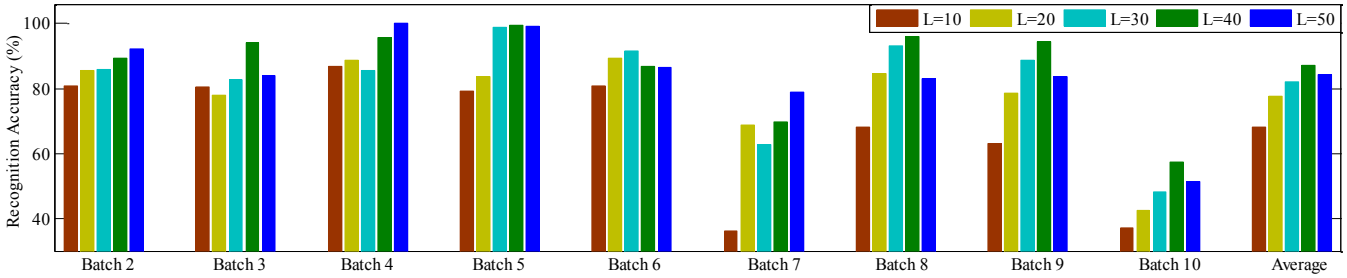

Fig. 4. Recognition Accuracy with Semi-supervised setting by using the proposed MFKS model with different number L of labeled data from batch K (K=2,…,10)

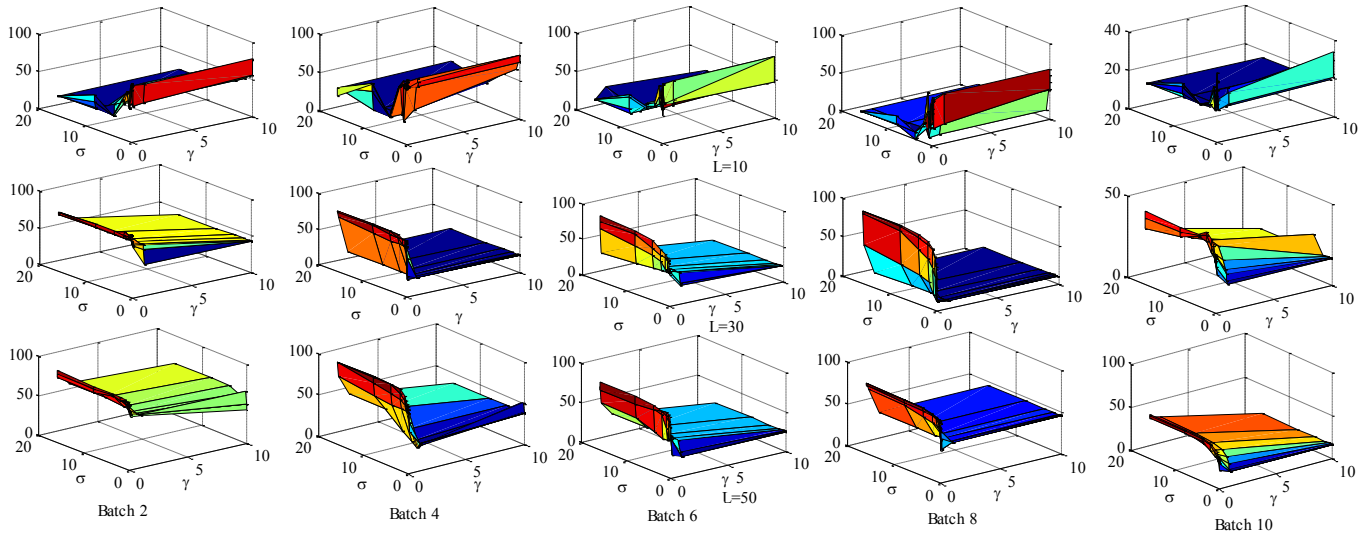

Fig. 5. Parameter sensitivity analysis (σ and γ) of MFKS with different number L=10, 30, 50 of labeled data for batch K (K=2, 4, 6, 8, 10)

batch 2 to batch 10 have been reported in Table II. From the results, we can see that the proposed MFKS outperforms other methods, but for batch 7, batch 9 and batch 10 SVM-comgfk and ML-comgfk show a comparative performance. The average recognition accuracies from batch 2 to batch 10 for MFKS (10) and MFKS (20) are 68.07% and 77.75%, which are superior to other methods. Note that L in MFKS (L) denotes the number of labeled data leveraged from batch K (K=2,…,10). From the comparisons, the effectiveness of the proposed MFKS is clearly demonstrated. This also implies that semi-supervised learning can be used to improve the drift-counteraction property of an E-Nose system.

- **Performance with increasing number *L***

In training process on batch 1, a few number L of labeled data from batch K (K=2,…,10) are leveraged for semi-supervised learning. Therefore, we discuss the performance with different number L of leveraged labeled samples for each test batch, respectively. The recognition accuracies for each batch with different number L from 10 to 50 by using the proposed MFKS model are illustrated in Fig. 4. From the results, we observe that the performance is positively improved with increasing number L, however, when L=50, the performance becomes weak. From the results, we can see the accuracy change *w.r.t.* L. Specifically, we get that for L=10, 20, 30, 40 and 50, the average recognition accuracy is 68.1%, 77.8%, 82.0%, 87.0% and 84.4%, respectively. Note that the accuracies when L=10 and 20 have been reported in Table II.

- **Parameter sensitivity analysis**

Two parameters including kernel parameter σ from $\{2^{-4}, 2^{-3}, \cdots, 2^{4}\}$ and regularization coefficient γ from $\{10^{-4}, 10^{-3}, \cdots, 10\}$ of MFKS model are discussed. The 3-D surface plots of σ (x-axis), γ (y-axis) and accuracy (z-axis) for batch K (K=2, 4, 6, 8, 10) under different L (L=10, 30, 50) are illustrated in Fig. 5. From the 3D surface, we can observe that a small σ produces a better performance for L=10 and the γ has less slight influence, while a small γ produces significantly

good performance when L=30 and 50, and the σ has little influence. From Fig. 5, it is easy to determine the best model parameters during semi-supervised training process.

V. Experiments on Modulated E-Nose Data

In this section, we adopt the temperature modulated E-Nose data for evaluating the proposed MFKS model and algorithm.

A. Experimental Data

The temperature modulated E-Nose system is developed by our group. Generally, the system consists of three parts: voltage control (i.e. sensor temperature control), sensor array and data acquisition. The printed circuit board (PCB) comprises of a gas sensor array (i.e. TGS2620 and TGS 2602), DC power supply interface, voltage control interface and data acquisition interface. We claim that the data is "modulated" due to that the sensors' heating voltage changes linearly from 3v to 5v continuously with a frequency of 20 mHz during sampling.

In sampling experiments, three gases including formaldehyde (HCHO), nitrogen dioxide (NO_2) and carbon monoxide (CO) are experimented by using our temperature modulated E-Nose system, separately. The modulated gas sensor response curves in a steady-state period of one observation for each gas are illustrated in Fig. 6. Note that 25 points are *uniformly sampled* as features from each sensor curve of 5000 points shown in Fig. 6. Therefore, for each observation, a 50-dimensional feature vector (2 sensors×25) is obtained. Inspired by the proposed multi-task learning, the 50-dimensional vector is then represented as 5 feature modalities, for each the dimension is 10 as shown in Fig.7.

Totally, the number of HCHO, NO_2 and CO samples is 100, 113 and 96, respectively. Note that for each feature type, the input dimension is 12 consisting of 10 points for gas sensors in Fig.7 and 2 extra values for the ambient temperature and humidity in sampling. The temperature range changes from 10℃ to 35℃ and the humidity range is from 40% to 70%.

B. Experimental Setup

In this experiment, we adopt the following experimental setting for evaluating different methods.

Experimental setting: for each class, different number L of labeled samples is randomly selected from the dataset of each class for semi-supervised learning. In this paper, L=1, 2, 3,…, 10 is considered. The remaining samples are recognized to be unlabeled data. The performance with increasing number L of labeled data is further explored.

For this modulated E-Nose data, we compare with two popular classifiers such as support vector machine (SVM) and extreme learning machine (ELM) [46]. Considering that SVM and ELM belong to single-task learning framework, for fair comparison, we therefore train SVM or ELM on feature type i (i=1,…,m) and obtain a classifier f_i at first, and then the final result of SVM or ELM is calculated as $1/m \cdot \sum_{i=1}^{m} f_i(\mathbf{x})$, which implies that each feature type makes equal contribution. Note that for this dataset, 5-fold cross-validation is used.

C. Parameter Setting

For ELM, two parameters including penalty coefficient C and the number H of hidden neurons are involved. Therefore, for obtaining the best performance, H is tuned from the set of

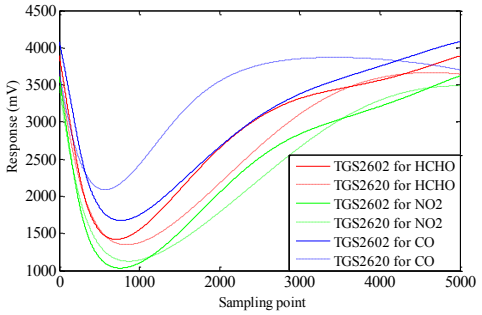

Fig. 6. Modulated steady-state response curve for one observation under HCHO, NO2 and CO, respectively

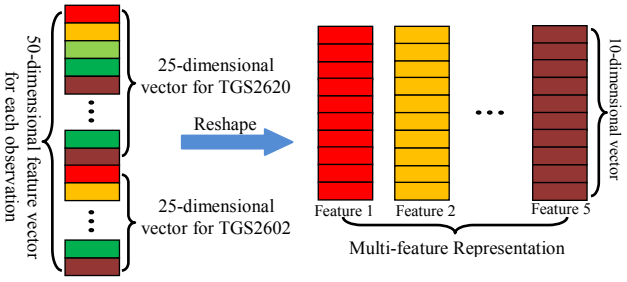

Fig. 7. Multi-feature representation of the extracted feature points for each observation

$\{10, 10^2, 10^3, 10^4\}$ and C is tuned from the set $\{10^{-5}, 10^{-4}, \cdots, 10^{10}\}$. For SVM, two parameters including penalty coefficient C and kernel parameter σ are involved. For the pursuit of the best performance, both C and σ are adjusted from the set $\{2^{-8}, 2^{-7}, \cdots, 2^{8}\}$. For MFKS, the kernel parameter σ is adjusted from the set $\{2^{-5}, 2^{-3}, \cdots, 2^{5}\}$ and the regularization coefficient γ is tuned from the set $\{10^{-8}, 10^{-7}, \cdots, 10\}$. The coefficient λ of graph manifold regularization is set as 1 in default. Note that the best results for each method are reported in this paper.

D. Experimental Results

- **Performance comparisons**

With the temperature-modulated E-Nose data, experimental setup and the parameter setting presented above, the average recognition accuracies of HCHO, NO_2, and CO for all methods are reported in Table III, in which the results with different number L of labeled data for each class are discussed. Note that SVM (*sl*), ELM (*sl*) and MFKS (*sl*) denote purely *supervised learning*, while MFKS (*ssl*) denotes *semi-supervised learning* with unlabeled data considered. For MFKS (*sl*), the model of Eq.(25) is used. From the results, we can observe that the proposed MFKS model and algorithm outperform SVM and ELM to a large extent for supervised learning. The average result of MFKS in semi-supervised setting is 83.47% which is higher than SVM (77.97%) and ELM (81.95%), but lower than that of MFKS in supervised learning (86.08%).

- **Parameter sensitivity analysis**

The parameter sensitivity analysis is presented for SVM, ELM and MFKS, respectively. The number L of labeled training data is set as 1, 3, 5, 7, and 9. The performance of SVM with C and σ is illustrated in Fig. 8(a) and we see that the kernel parameter σ plays a key role in recognition than penalty parameter C. The performance of ELM with C and H is shown

TABLE III
RECOGNITION ACCURACY (%) ON TEST DATA WITH DIFFERENT NUMBER OF LABELED DATA FROM TRAINING SET

Method	*L*=1	*L* =2	*L* =3	*L* =4	*L* =5	*L* =6	*L* =7	*L* =8	*L* =9	*L* =10	Average
SVM (*sl*)	68.69	75.91	76.73	80.4	78.64	78.42	78.61	79.16	79.15	83.94	77.97
ELM (*sl*)	72.94	77.69	78.00	83.03	82.86	82.96	82.99	84.63	85.39	88.96	81.95
MFKS (*ssl*)	66.66	78.87	**83.33**	**88.88**	85.37	85.56	86.45	85.61	86.52	87.45	83.47
MFKS (*sl*)	**75.16**	**81.51**	82.00	83.50	**88.77**	**88.65**	**91.66**	**86.66**	**89.36**	**93.54**	**86.08**

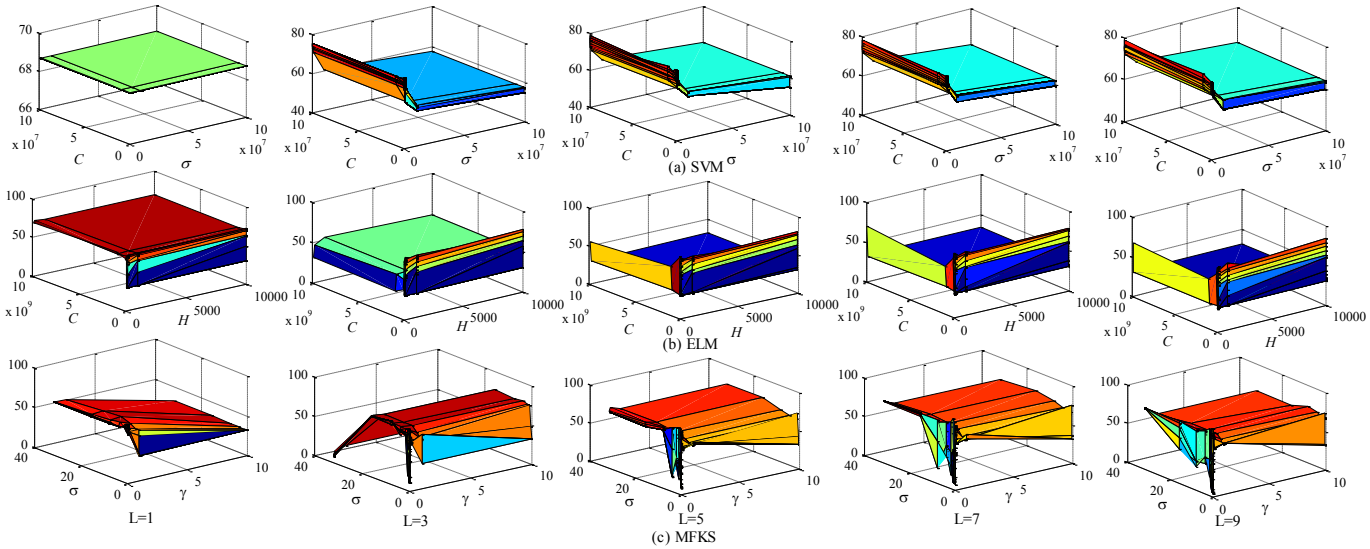

Fig. 8. Parameter sensitivity analysis of SVM, ELM and MFKS with different number *L*=1, 3, 5, 7, and 9 of labeled data. z-axis denotes the accuracy (%)

in Fig. 8 (b) and we observe that the penalty parameter *C* has larger influence than the number *H* of hidden neurons. The recognition performance of MFKS with kernel parameter σ and regularization coefficient γ is illustrated in Fig. 8(c). We can observe that for different number of labeled training data, the relation between the recognition accuracy (z-axis) and the model parameters (x-axis and y-axis) is similar. Specifically, a very small or large σ would produce weak recognition performance, and the influence of γ is not very obvious. However, a larger γ would produce more stable performance.

VI. CONCLUSION

In this paper, we aim at introducing a new machine learning perspective in artificial olfaction system. Sensor array feature extraction is an important topic in electronic nose community. Instead of traditional feature selection, we attempt to learn the potential features from the viewpoint of machine learning and achieve an insight of the "best" feature that mostly contributes to pattern classification. In general, feature extraction and pattern classification in E-Nose are conducted independently, such that a "hard" learning of the pattern classifier is weakly conducted for adapting to the hand-crafted features and the classification capability is restricted. Motivated by this issue, we aim at proposing a unified learning framework in artificial olfaction system, which tends to learn multiple features and multiple sub-classifiers simultaneously. In this way, the "hard" learning automatically evolves to "soft" learning and achieves an optimal cooperative learning between features and classifiers. Specifically, a multi-feature joint semi-supervised learning with kernel mapping is proposed, and it is the so-called MFKS. The proposed model, optimization algorithm, convergence and complexity have been completely formulated and discussed in this paper. Furthermore, for validating the effectiveness of MFKS in odor classification by an electronic nose, two experiments on an existing large-scale artificial olfaction dataset with sensor drift of 36 months and an existing small-scale temperature modulated E-Nose data are conducted, respectively. Experimental results on the two existing artificial olfaction datasets demonstrate that the proposed MFKS model outperforms other algorithms in recognition performance.

In the future, we would like to focus on the multi-task learning and transfer learning for robust domain adaptation in the pursuit of a drift-countered artificial olfaction system. Large-scale transfer learning would be a competitive research direction for challenging E-Nose data.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Vergara, Dr. J. Fonollosa and their team from UCSD for their provided long-term artificial olfaction data. The authors would also like to thank the AE and anonymous reviewers for their insightful comments and concerns.

APPENDIX

To solve α_i in the equation group (21) in the paper, we describe the detail as follows.

For convenience, we rewrite the equation group (21) as

$$\begin{cases} r\alpha_i^{r-1} Tr\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}_i\mathbf{F}\right) - \mu = 0 \\ \sum_{i=1}^{m} \alpha_i - 1 = 0 \end{cases} \tag{26}$$

To the first equation in (26), there is

$$\left[r\alpha_i^{r-1}Tr\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}_i\mathbf{F}\right)\right]^{\frac{1}{r-1}} = \mu^{\frac{1}{r-1}}$$

$$\Downarrow$$

$$\alpha_i = \mu^{\frac{1}{r-1}} \Big/ \left(r^{\frac{1}{r-1}} \cdot Tr\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}_i\mathbf{F}\right)^{\frac{1}{r-1}} \right) \tag{27}$$

$$\Downarrow$$

$$\sum_{i=1}^{m}\alpha_i = \mu^{\frac{1}{r-1}} \Big/ r^{\frac{1}{r-1}} \cdot \sum_{i=1}^{m}\left(1 \Big/ Tr\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}_i\mathbf{F}\right)^{\frac{1}{r-1}} \right)$$

Consider (26) and (27), we have

$$\mu^{\frac{1}{r-1}} \Big/ r^{\frac{1}{r-1}} = 1 \Big/ \sum_{i=1}^{m}\left(1 \Big/ Tr\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}_i\mathbf{F}\right)^{\frac{1}{r-1}} \right) \tag{28}$$

Substitute (28) into (27), we can obtain α_i as Eq.(22).

REFERENCES

[1] S. Marco, A. Gutierrez-Galvez, "Signal and Data Processing for Machine Olfaction and Chemical Sensing: A Review," IEEE Sensors Journal, vol. 12, no. 11, pp. 3189-3214, 2012.

[2] J.W. Gardner and P.N. Bartlett, "Electronic Noses: Principles and Applications," Oxford, U.K.: Oxford Univ. Press, 1999.

[3] Y. Adiguzel and H. Kulah, "Breath sensors for lung cancer diagnosis," Biosensors and Bioelectronics, vol. 65, pp. 121-138, 2015.

[4] Sunny, V. Kumar, V.N. Mishra, R. Dwivedi, R.R. Das, "Classification and Quantification of Binary Mixtures of Gases/Odors Using Thick-Film Gas Sensor Array Responses," IEEE Sensors Journal, vol. 15, no. 2, pp. 1252-1260, 2015.

[5] M. Hassan, S.B. Belhaouari, and A. Bermak, "Probabilistic Rank Score Coding: A Robust Rank-Order Based Classifier for Electronic Nose Applications," IEEE Sensors Journal, vol. 15, no. 7, pp.3934-3946, 2015.

[6] M. Pardo and G. Sberveglieri, "Classification of electronic nose data with support vector machines," Sensors and Actuators B: Chemical, vol. 107, no. 2, pp. 730-737, 2005.

[7] L. Zhang, F. Tian, H. Nie, L. Dang, G. Li, Q. Ye, and C. Kadri, "Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine," Sensors and Actuators B: Chemical, vol. 174, pp. 114-125, 2012.

[8] K. Brudzewski, S. Osowski, and A. Dwulit, "Recognition of Coffee Using Differential Electronic Nose," IEEE Transactions on Instrumentation and Measurement, vol. 61, no. 6, pp. 1803-1810, 2012.

[9] E. Martinelli, G. Magna, D. Polese, A. Vergara, D. Schild, and C. Di Natale, "Stable Odor Recognition by a neuro-adaptive Electronic Nose," Scientific Report, 2015.

[10] F. Hossein-Babaei and A. Amini, "Recognition of complex odors with a single generic tin oxide gas sensor," Sensors and Actuators B: Chemical, vol. 194, pp. 156-163, 2014.

[11] J. Fonollosa, S. Sheik, R. Huerta, and S. Marco, "Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring," Sensors and Actuators B: Chemical, vol. 215, pp. 618-629, 2015.

[12] R. Guterrez-Osuna and H.T. Nagle, "A method for evaluation data-preprocessing techniques for odor classification with an array of gas sensors," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 29, no. 5, pp. 626-632, 1999.

[13] C. Krutzler, A. Unger, H. Marhold, T. Fricke, T. Conrad, A. Schütze, "Influence of MOS Gas-Sensor Production Tolerances on Pattern Recognition Techniques in Electronic Noses," IEEE Transactions on Instrumentation and Measurement, vol. 61, no. 1, pp. 276-283, 2012.

[14] S.K. Jha, R.D.S. Yadava, "Denoising by singular value decomposition and its application to electronic nose data processing," IEEE Sensors Journal, vol. 11, no. 1, pp. 35-44, 2011.

[15] X.R. Wang, J.T. Lizier, A.Z. Berna, F.G. Bravo, and S.C. Trowell, "Human breath-print identification by E-nose, using information-theoretic feature selection prior to classification," Sensors and Actuators B: Chemical, vol. 217, pp. 165-174, 2015.

[16] T.T. Sunil, S. Chaudhuri, and V. Mishra, "Optimal selection of SAW sensors for E-Nose applications," Sensors and Actuators B: Chemical, vol. 219, pp. 238-244, 2015.

[17] Y. Dai, R. Zhi, L. Zhao, H. Gao, B. Shi, H. Wang, "Longjing tea quality classification by fusion of features collected from E-nose," Chemometrics and Intelligent Laboratory Systems, 2015.

[18] A.K. Bag, B. Tudu, N. Bhattacharyya, R. Bandyopadhyay, "Dealing With Redundant Features and Inconsistent Training Data in Electronic Nose: A Rough Set Based Approach," IEEE Sensors Journal, vol. 14, no. 3, pp. 758-767, 2014.

[19] L. Fernandez, S. Marco, and A. Gutierrez-Galvez, "Robustness to sensor damage of a highly redundant gas sensor array," Sensors and Actuators B: Chemical, vol. 218, pp. 296-302, 2015.

[20] A. Vergara, S. Vembu, T. Ayhan, M.A. Ryan, M.L. Homer, "Chemical gas sensor drift compensation using classifier ensembles," Sensors and Actuators B: Chemical, vol. 166, pp. 320-329, 2012.

[21] M. Holmberg, F.A.M. Davide, C.D. Natale, A.D. Amico, F. Winquist, and I. Lundström, "Drift counteraction in odour recognition applications: Lifelong calibration method," Sensors and Actuators B: Chemical, vol. 42, no. 3, pp. 185-194, 1997.

[22] S.D. Carlo and M. Falasconi, "Drift correction methods for gas chemical sensors in artificial olfaction systems: Techniques and Challenges," Adv. Chem. Sensors, pp. 305-326, 2012.

[23] G. Fattoruso, S. De Vito, M. Pardo, F. Tortorella, G. Di Francia, "A Semi-Supervised Learning Approach to Artificial Olfaction," Lecture Notes in Electrical Engineering, vol. 109, pp. 157-162, 2012.

[24] S. De Vito, G. Fattoruso, M. Pardo, F. Tortorella, G. Di Francia, "Semi-Supervised Learning Techniques in Artificial Olfaction: A Novel Approach to Classification Problems and Drift Counteraction," IEEE Sensors Journal, vol. 12, no. 11, Nov 2012.

[25] O. Chapelle, A. Zien, and B. Sholkopf, Semi-Supervised Learning. Boston, MA: MIT Press, 2006.

[26] D. Zhou, O. Bousquet, T. Navin Lal, J. Weston, and B. Schölkopf, "Learning with Local and Global Consistency," NIPS, pp. 321-328, 2004.

[27] Y. Luo, D. Tao, B. Geng, C. Xu, "Manifold Regularized Multitask Learning for Semi-Supervised Multilabel Image Classification," IEEE Transactions on Image Processing, vol. 22, no. 2, pp. 523-536, 2013.

[28] Y. Yang, Z. Ma, A.G. Hauptmann, N. Sebe, "Feature Selection for Multimedia Analysis by Sharing Information Among Multiple Tasks," IEEE Transactions on Multimedia, vol. 15, no. 3, pp. 661-669, 2013.

[29] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol. 290, no. 22, pp. 2323-2326, 2000.

[30] J. Tenenbaum, V. Silva, and J. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, vol. 290, no. 22, pp. 2319-2323, 2000.

[31] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.

[32] S. Yan, D. Xu, B. Zhang, H.J. Zhang, Q. Yang, and S. Lin, "Graph Embedding and Extensions: A General Framework for Dimensionality Reduction," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 40-51, 2007.

[33] T. Xia, T. Mei, Y. Zhang, "Multiview Spectral Embedding," IEEE Transactions on Systems, Man, and Cybernetics-part B, vol. 40, no. 6, pp. 1438-1446, 2010.

[34] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, A. Perera, "Drift compensation of gas sensor array data by common principal component analysis," Sensors and Actuators B: Chemcial, vol. 146, no. 2, pp. 460-465, 2010.

[35] S.D. Carlo, M. Falasconi, E. Sanchez, A. Scionti, G. Squillero, and A. Tonda, "Increasing pattern recognition accuracy for chemical sensing by evolutionary based drift compensation," Pattern Recognit. Lett., vol. 32, no. 13, pp. 1594-1603, 2011.

[36] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, and S. Marco, "Drift compensation of gas sensor array data by Orthogonal Signal Correction," vol. 100, no. 1, pp. 28-35, 2010.

[37] L.J. Dang, F. Tian, L. Zhang, C. Kadri, X. Yin, X. Peng, and S. Liu, "A novel classifier ensemble for recognition of multiple indoor air contaminants by an electronic nose," Sensors and Actuators A: Physical, vol. 207, pp. 67-74, 2014.

[38] H. Liu, R. Chu, and Z. Tang, "Metal Oxide Gas Sensor Drift Compensation Using a Two-Dimensional Classifier Ensemble," Sensors, vol. 15, no. 5, pp. 10180-10193, 2015.

[39] Q. Liu, X. Li, M. Ye, S.S. Ge, and X. Du, "Drift Compensation for Electronic Nose by Semi-Supervised Domain Adaption," IEEE Sensors

Journal, vol. 14, no. 3, pp. 657-665, 2014.
[40] L. Zhang and D. Zhang, "Domain Adaptation Transfer Extreme Learning Machine," Proceedings in Adaptation, Learning and Optimization, vol. 3, pp. 103-119, 2015.
[41] L. Zhang and D. 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, 2015.
[42] L. Martin, L. Amy, "Unsupervised feature learning for electronic nose data applied to Bacteria Identification in Blood," NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.
[43] Q. Liu, X. Hu, M. Ye, X. Cheng, and F. Li, "Gas Recognition under Sensor Drift by Using Deep Learning," International Journal of Intelligent Systems, vol. 30, no. 8, pp. 907-922, 2015.
[44] http://archive.ics.uci.edu/ml/datasets/Gas+Sensor+Array+Drift+Dataset+at+Different+Concentrations.
[45] D.A.P. Daniel, K. Thangavel, R. Manavalan, and R.S.C. Boss, "ELM-Based Ensemble Classifier for Gas Sensor Array Drift Dataset," Adv. Intell. Syst. Comput. vol. 246, pp. 89-96, 2014.
[46] 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, 2012.
[47] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust Recovery of Subspace Structure by Low-Rank Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 171-184, 2013.
[48] I. Rodriguez-Lujan, J. Fonollosa, A. Vergara, M. Homer, and R. Huerta, "On the calibration of sensor arrays for pattern recognition using the minimal number of experiments," Chemometrics and Intelligent Laboratory Systems, vol. 130, pp. 123-134, 2014.

Lei Zhang received his Ph.D degree in Circuits and Systems from the College of Communication Engineering, Chongqing University, Chongqing, China, in 2013. He is currently a Professor/Distinguished Research Fellow with Chongqing University. He was selected as a Hong Kong Scholar in China in 2013, and worked as a Post-Doctoral Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2013 to 2015. He has authored more than 40 scientific papers in top journals, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, the IEEE SENSORS JOURNAL, INFORMATION FUSION, SENSORS & ACTUATORS B, and ANALYTICA CHIMICA ACTA. His current research interests include electronic olfaction, machine learning, pattern recognition, computer vision and multi-sensor system. Dr. Zhang was a recipient of Outstanding Doctoral Dissertation Award of Chongqing, China, in 2015, Hong Kong Scholar Award in 2014, Academy Award for Youth Innovation of Chongqing University in 2013 and the New Academic Researcher Award for Doctoral Candidates from the Ministry of Education, China, in 2012.

David Zhang graduated in Computer Science from Peking University. He received his MSc in 1982 and his PhD in 1985 in Computer Science from the Harbin Institute of Technology (HIT), respectively. From 1986 to 1988 he was a Postdoctoral Fellow at Tsinghua University and then an Associate Professor at the Academia Sinica, Beijing. In 1994 he received his second PhD in Electrical and Computer Engineering from the University of Waterloo, Ontario, Canada. He is a Chair Professor since 2005 at the Hong Kong Polytechnic University where he is the Founding Director of the Biometrics Research Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He also serves as Visiting Chair Professor in Tsinghua University, and Adjunct Professor in Peking University, Shanghai Jiao Tong University, HIT, and the University of Waterloo. He is the Founder and Editor-in-Chief, International Journal of Image and Graphics (IJIG); Book Editor, Springer International Series on Biometrics (KISB); Organizer, the International Conference on Biometrics Authentication (ICBA); Associate Editor of more than ten international journals including IEEE Transactions and so on; and the author of more than **10** books, over **300** international journal papers and **30** patents from USA/Japan/HK/China. Professor Zhang is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and a Fellow of both IEEE and IAPR.

Xin Yin received his BSc and MSc degree in Circuits and Systems from College of Communication Engineering, Chongqing University, Chongqing, China, in 2012 and 2015. He is now working in China Electronics Technology Group Corporation No.29 Research Institute. His research interests include hardware design, electronic nose technology and pattern recognition.

Yan Liu received her Bachelor degree in Information Engineering in 2014 from Chengdu Polytechnic University, China. Since September 2014, she is currently pursuing the MSc degree in Chongqing University. Her research interests include electronic nose and intelligent algorithm.