MULTI-VIEW MULTI-TASK FEATURE LEARNING: APPLICATION TO TELEMONITORING DATA ANALYSIS OF PARKINSON'S DISEASE

Lin Li *Yali Wang* *Yu Qiao*

Shenzhen Key Lab of Computer Vision & Pattern Recognition
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
{lin.li,yl.wang,yu.qiao}@siat.ac.cn

ABSTRACT

The Unified Parkinson's Disease Rating Scale (UPDRS) plays an important role to assess Parkinson's Disease (PD) symptoms. To avoid time-consuming and expensive medical examinations in clinic, telemonitoring analysis of UPDRS has been spotlighted recently by using the simple and rapid tests of patient's speech recordings. However, different types of UPDRS scores in the telemonitoring data are strongly-correlated tasks, and different types of bio-medical voice features represent complex multi-view characteristics. This significantly restricts the accuracy and robustness of PD telemonitoring analysis performance. In this paper, we propose a novel Multi-View Multi-Task Feature Learning (MVMTFL) approach in which multi-view representation and multi-task modeling are flexibly incorporated in a non-trivial unifying framework. By sharing a feature subspace for each view within multiple tasks, our MVMTFL can capture the strong correlations between different tasks, and also represent the complex multi-view information of voice features. Experimental results on a challenging Parkinson's disease telemonitoring dataset show that our MVMTFL significantly outperforms the baseline approaches, and also exhibits high accuracy and robustness to different sizes of training data and different partitions of feature views.

***Index Terms*—** Parkinson's Disease (PD), Multi-View Multi-Task Feature Learning (MVMTFL), Telemonitoring Analysis

1. INTRODUCTION

During the past decades, Parkinson's disease (PD) has been significantly investigated to alleviate the neurological disorder symptoms of patients [1, 2, 3, 4, 5, 6]. The Unified Parkinsons Disease Rating Scale (UPDRS) is a crucial factor to track PD progression. However, the traditional PD progression monitoring is time-consuming and costly, since the patients are required to take a number of complex medical examinations in clinic. Alternatively, with the development of signal processing, Internet and multimedia technologies, telemonitoring analysis of UPDRS has become a more convenient choice for PD symptoms assessment [3]. By processing the simple and noninvasive tests of patient's speech recordings, one can use a number of biomedical voice features to predict different scores of UPDRS for different PD patients.

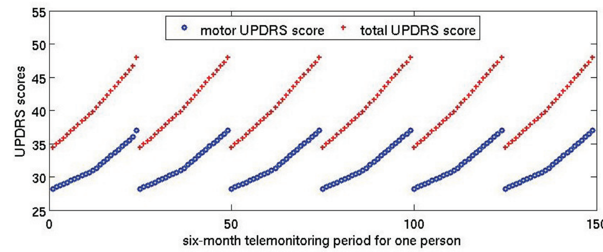

Fig. 1. Total and motor UPDRS scores during six-month telemonitoring period for one person. It is shown that these two UPDRS scores are strongly correlated.

However, as shown in Fig. 1, different types of UPDRS are strongly-correlated tasks in the telemonitoring data[3]. The predictive performance for each single UPDRS score is often limited if we learn each task independently. To capture the dependence between different tasks, Multi-Task Feature Learning (MTFL) is often a reasonable model by commonly sharing a low-dimensional latent subspace [7, 8, 9]. Unfortunately, MTFL simply treats multi-view features [10] as a single view or treats different views independently [11]. This often reduces the prediction accuracy, especially when different types of features (such as multiple biomedical voice features in our case) represent complex multi-view properties.

In this paper, we propose a novel Multi-View Multi-Task Feature Learning (MVMTFL) approach where multi-view representation and multi-task modeling are jointly learned in a non-trivial unifying framework. By sharing a latent subspace for each feature view within multiple tasks, our MVMTFL does not only capture the strong correlations between different UPDRS tasks but also represents multiple views of biomedical voice features. Additionally, we utilize trace-norm regularization to further improve effectiveness and efficiency of our MVMTFL approach. Finally we perform our MVMTFL on a challenging Parkinson's disease

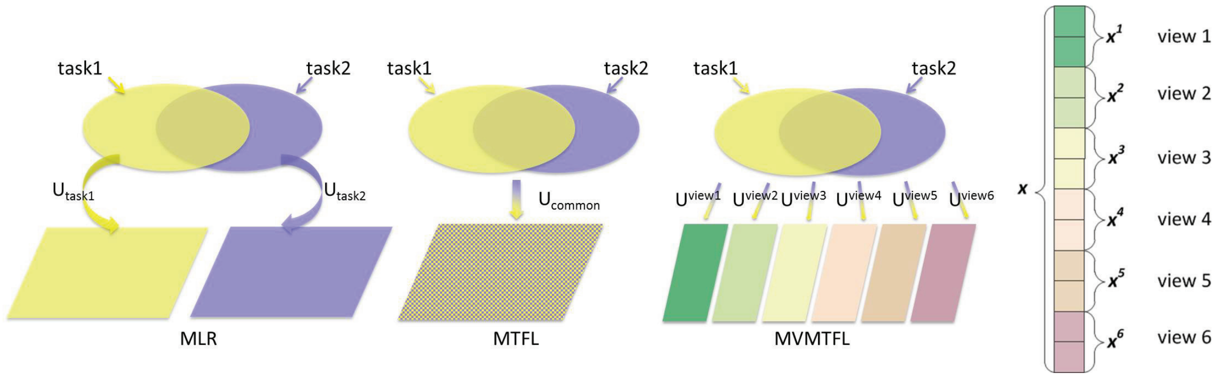

Fig. 2. Illustration of MLR (left), MTFL (center), and MVMTFL (right). Task1 and Task2 are two correlated tasks where the input feature x for both tasks have 6 feature views $(x^1, x^2, ..., x^6)$. Conceptionally, MLR learns two independent latent subspaces U_{task1}, U_{task2} for each individual task. Hence it ignores the dependencies between different tasks. Alternatively, MTFL learns a common subspace U_{common} between two tasks. However this design ignores multiple views of input features. On the contrary, our MVMTFL constrains the common subspaces $(U^1, U^2, ..., U^6)$ to each feature view among multiple tasks. Hence, multi-view and multi-task can be naturally incorporated into our approach.

telemonitoring dataset. The experimental results show that our MVMTFL significantly outperforms other baseline approaches, and also show that our MVMTFL is quite robust to different sizes of data and different partitions of feature views.

2. OUR MVMTFL APPROACH

2.1. Problem Formulation

Suppose that there are T correlated tasks, and the t-th task is modeled as $y_t = f_t(x) =< w_t, x >$. The input feature $x \in R^d$ contains K different views of features $[x^1, \ldots, x^K]$ where the k-th view is $x^k \in R^{d_k}$. Additionally, the output of the t-th task is $y_t \in R$; the model parameter of the t-th task is w_t; the operation $< \cdot, \cdot >$ is inner product.

To achieve multi-view representation within multiple tasks, we assume that, for each feature view x^k, all T tasks share a common latent subspace U^k where U^k is a $d_k \times d_k$ orthogonal matrix. In this case, we reformulate the model of the t-th task by taking different feature views into account

$$y_t = f_t(x) =< w_t, x >= \sum_{k=1}^{K} < a_t^k, {U^k}^T x^k > \quad (1)$$

where a_t^k is the parameter in the projected subspace U^k for the t-th task. Consequentially, the model parameter w_t can be represented as a multi-view form, i.e., $w_t = [w_t^1, \ldots, w_t^K]$ and $w_t^k = (U^k a_t^k)^T$. For convenience, we denote $W^k = [w_1^k, ..., w_T^k]^T$, $W = [W^1; ...; W^K]$, $U = \text{Diag}(U^1, ..., U^K)$, $A^k = [a_1^k, ..., a_T^k]^T$ and $A = [A^1; ...; A^K]$. Fig. 2 demonstrates the model of our Multi-View Multi-Task Feature Learning (MVMTFL). Different from the classical multiple linear regression and multi-task feature learning, our MVMTFL shares a common subspace for each feature view within multiple tasks.

Given M training examples $\{(x_{tm}, y_{tm})\}_{m=1}^{M}$ for each task (M could be different for different tasks), we minimize the following objective function to jointly learn multiple tasks and capture multiple views

$$\begin{aligned} E(A, U) = &\sum_{t=1}^{T} \sum_{i=1}^{M} L(y_{ti}, \sum_{k=1}^{K} < a_t^k, {U^k}^T x_{ti}^k >) \\ &+ \gamma \sum_{k=1}^{K} \left\|A^k\right\|_{2,1}^2 \end{aligned} \quad (2)$$

where L is a square loss function, and $\gamma > 0$ is a weight to balance between the loss L and the $l_{2,1}$-norm regularizer.

2.2. Optimization

Eq. (2) is a non-convex problem. Inspired by [7], we reformulate Eq. (2) as the following convex auxiliary function with a perturbation term ε,

$$\begin{aligned} R_\varepsilon(W, D) = &\sum_{t=1}^{T} \sum_{i=1}^{M} L(y_{ti}, \sum_{k=1}^{K} < w_t^k, x_{ti}^k >) \\ &+ \gamma \sum_{k=1}^{K} \text{trace}((D^k)^{-1}(W^k {W^k}^T + \varepsilon I)) \end{aligned} \quad (3)$$

where I is an identity matrix, D^k is a rank-d_k semi-definite matrix and $D = \text{Diag}(\text{D}^1, ..., \text{D}^\text{K})$.

In this case, we can iteratively optimize model parameter matrix W^k and feature matrix D^k. With a fixed D^k, the problem in Eq. (3) reduces to the following standard 2-norm regularization optimization for each task, and thus can be solved straightforwardly [7]

$$\underset{W^k}{\arg\min} \sum_{i=1}^{M} L(y_{ti}, \sum_{k=1}^{K} < w_t^k, x_{ti}^k >) + \gamma \sum_{k=1}^{K} < w_t^k, (D^k)^{-1} w_t^k > \tag{4}$$

Then, with a fixed W^k, Eq.(3) can be reduced to the following problem for each view [7]

$$\underset{D^k}{\arg\min} \sum_{t=1}^{T} < w_t^k, (D^k)^{-1} w_t^k > + \varepsilon trace((D^k)^{-1}) \tag{5}$$

Minimizing over D^k achieves a closed form solution

$$\hat{D}_{\varepsilon}^{k} = \frac{(W^k {W^k}^T + \varepsilon I)^{1/2}}{\text{trace}(W^k {W^k}^T + \varepsilon I)^{1/2}} \tag{6}$$

As a result, optimization for W^k and D^k in our MVMTFL can be efficiently performed in an iterative manner.

3. EXPERIMENTS AND RESULTS

3.1. Parkinsons Telemonitoring Data Set

In the Parkinsons telemonitoring data set[1], there are six-month trials of biomedical voice recordings from different patients with early-stage Parkinson's disease. The data size for one patient is roughly 150. These recordings are obtained automatically by a telemonitoring device to remotely assess Parkinson's disease symptoms progression.

Each voice recording is characterized by 16 vocal features which represent 6 semantic views of the voice signals. Specifically, (i) The 5 jitter features (Jitter(%), Jitter(Abs), Jitter:RAP, Jitter:PPQ5, Jitter:DDP) represent the range of variation in speech fundamental frequency of different recording cycles. (ii) The 6 Shimmer features (Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA) depict the range of cycle to cycle variation in amplitude of voice signals. (iii) The NHR and HNR features describe noise-to-harmonics ratio and harmonics-to-noise ratio respectively. (iv) The RPDE feature is recurrence period density entropy. (v) The DFA feature is detrended fluctuation analysis. (vi) The PPE feature is pitch period entropy. More details of different features can be found in [3].

The goal is to predict two strongly-correlated clinician's PD symptoms scores on UPDRS (i.e., motor and total UPDRS scores) by learning a mapping from 16 voice features to these two UPDRS scores. In the following experiments, we show our multi-view multi-task feature learning (MVMTFL) approach can successfully capture the strong-dependencies between motor and total UPDRS scores, and also represent the complex multi-view information of different voice features. For all the experiments, we normalize the voice features along each feature dimension to make each dimension as zero mean and unit variance. We also subtract the mean respectively for two UPDRS scores. Furthermore, the baselines for comparison are multiple linear regression (MLR) [3] and a related multi-task feature learning (MTFL) approach [7]. The predictive performance for all approaches is measured by mean squared error (MSE).

[1] http://archive.ics.uci.edu/ml/index.html

3.2. Prediction with Different Sizes of Data

We firstly evaluate our approach with different sizes of training/test data settings. In our MVMTFL, we utilize the abovementioned 6 semantic views of voice features for multi-view consideration.

To show the robustness of our MVMTFL, we run all the approaches on the Parkinson's telemonitoring data of 6, 10, 20 people. For each of 6, 10, 20 people settings, we perform the following strategy to choose training/test sets. Firstly, we randomly choose 3/3, 5/5, 10/10 people as the training/test set for the task of motor UPDRS score. Then we perform 3/3, 5/5, 10/10 random selection procedure again for the task of total UPDRS score. In this case, the training set of motor and total UPDRS tasks can be quite different, which makes the prediction problem even harder.

Taking one of our training/test settings for example, people with (ID: 3, 5, 6)/(ID: 1, 2, 4) are chosen as the training/test data for motor UPDRS score. People with (ID: 2, 5, 6)/(ID: 1, 3, 4) are chosen as the training/test data for total UPDRS score. In this setting, learning correlations between different UPDRS tasks is quite crucial due to the fact that people with ID: 2/ ID:3 are respectively missing in the training set of motor/total UPDRS tasks.

For each of 6, 10, 20 people settings, we randomly partition training/test data five times and show average MSE in Table.1. It is shown that our MVMTFL achieves the best results among all these methods. Moreover, from the standard deviation of MSE, we can see that our MVMTFL are the most robust approach to all different training/test data settings.

3.3. Prediction with Different Partitions of Feature Views

We now evaluate our approach with different partitions of feature views. We respectively investigate one data set for each of 6, 10, 20 people settings. To examine multiple views of the voice feature, we use one data set of 6 people setting as an example to show the covariance matrix of different features.

The resulting matrix is shown in Fig. 3. It is clear that the 5 jitter features are highly-correlated. It illustrates that these 5 features roughly follow their semantic meaning. This finding resembles our suggestion for semantic feature view partition of jitter features. Similarly, the 6 shimmer features also follow their semantic view partition. However, the NHR and

Table 1. Mean square error (mean ± std) of total and motor UPDRS scores with different sizes of training/test data.

train/test	3 people/3 people			5 people/5 people			10 people/10 people		
method	MLR	MTFL	MVMTFL	MLR	MTFL	MVMTFL	MLR	MTFL	MVMTFL
motor UPDRS	139.13 ±71.28	64.27 ±12.11	**61.62 ±10.99**	109.10 ±30.83	89.12 ±10.17	**76.51 ±5.97**	77.16 ±12.26	67.60 ±12.24	**65.80 ±10.82**
total UPDRS	227.80 ±100.18	111.13 ±36.16	**110.29 ±35.06**	164.13 ±65.12	154.16 ±67.77	**131.46 ±44.23**	109.26 ±20.37	97.13 ±17.67	**94.80 ±16.17**

Table 2. Mean square error (mean ± std) of total and motor UPDRS scores with different partitions of feature views.

train/test	method	MLR	MTL	MVMTL				
				semantic	clustering			
				6	2	4	6	8
3 /3 people	motor UPDRS	170.36	82.95	78.42	80.15± 0.01	**76.93 ± 0.91**	77.31 ± 0.65	78.43 ± 0.75
	total UPDRS	209.78	76.02	74.11	75.62 ± 0.01	74.83 ± 0.05	74.48 ± 0.28	**73.98 ± 0.32**
5 /5 people	motor UPDRS	116.97	81.89	**77.22**	80.73±0.32	79.74±0.01	79.31±0.51	80.76±0.29
	total UPDRS	231.93	256.56	196.13	231.89±0.01	209.11±0.01	191.53±2.63	**181.12±2.49**
10 /10 people	motor UPDRS	81.11	71.27	**69.20**	70.69±0.04	69.97±0.11	69.35±0.29	69.94±0.21
	total UPDRS	106.17	103.34	99.64	102.09±0.38	100.79±0.16	99.58±0.18	**99.09±0.15**

HNR features do not follow their semantic partition in which these two features are considered as a common view. Fig. 3 clearly shows that HNR is not quite related with all other features including NHR. Additionally, the RPDE, DFA and PPE features are treated as different views according to their semantic meaning, while they are actually weakly correlated to each other in the covariance matrix.

All these findings inspires us to further explore multiple views of biomedical voice features in this Parkinson's disease. Hence, we perform kmeans clustering for each data with the number of clusters 2,4,6,8 five times to eliminate the randomness effect of clustering. Then, we run the baseline approaches MLR, MTFL and our MVMTFL with semantic multi-view partition (i.e., 16 vocal features are considered as 6 semantic views). Average MSE results for all the approaches are summarized in Table.2.

In Table.2, our MVMTFL approach with 2 views (clusters) is even better than MTFL for all data sets. This illustrates that the multi-view property is a quite crucial factor in this Parkinson's disease data set. Furthermore, our MVMTFL with 4, 6, 8 clustering views is generally better than our MVMTFL with 6 semantic views. This is credited to the fact that we correctly explore the feature correlation using the similarities between different features.

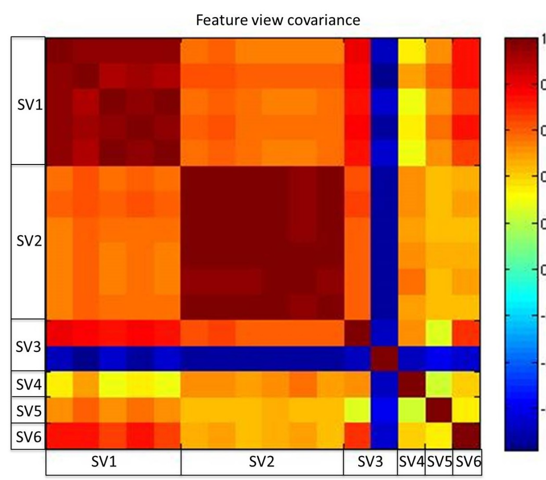

Fig. 3. Illustration of feature covariance matrix (6 people). For comparison, we also show 6 semantic views (SV): SV1 are 5 jitter features, SV2 are 6 shimmer features, SV3 are NHR and HNR, SV4 is RPDE, SV5 is DFA, SV6 is PPE. It is shown that similarities between features in the covariance matrix do not quite match to the ones in their semantic views.

4. CONCLUSIONS

In this paper, we propose a novel Multi-View Multi-Task Feature Learning (MVMTFL) approach to predict UPDRS scores for Parkinson's Disease telemonitoring data. By jointly learning the latent subspace of each feature view within multi-task, our MVMTFL can successfully capture the strong-task-correlations between different UPDRS scores, and also flexibly represent the complex multi-view characteristics of remote voice features of patients. Our experiments show that our MVMTFL significantly outperforms the baselines, and thus improve the accuracy and robustness to telemonitoring prediction of Parkinson's disease symptoms progression. In the future, it would be interesting to extend our approach to classify which stage of Parkinson's disease is for one patient.

5. REFERENCES

[1] Anthony E Lang and Andres M Lozano, "Parkinson's disease," *New England Journal of Medicine*, vol. 339, pp. 1044–1053, 1998.

[2] Stephen K Van Den Eeden, Caroline M Tanner, Allan L Bernstein, Robin D Fross, Amethyst Leimpeter, Daniel A Bloch, and Lorene M Nelson, "Incidence of parkinsons disease: variation by age, gender, and race/ethnicity," *American journal of epidemiology*, vol. 157, pp. 1015–1022, 2003.

[3] Athanasios Tsanas, Max Little, Patrick E McSharry, Lorraine O Ramig, et al., "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *Biomedical Engineering, IEEE Transactions on*, vol. 57, pp. 884–893, 2010.

[4] Aileen K Ho, Robert Iansek, Caterina Marigliani, John L Bradshaw, and Sandra Gates, "Speech impairment in a large sample of patients with parkinsons disease," *Behavioural neurology*, vol. 11, pp. 131–137, 1999.

[5] Jeri A Logemann, Hilda B Fisher, Benjamin Boshes, and E Richard Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," *Journal of Speech and Hearing Disorders*, vol. 43, pp. 47–57, 1978.

[6] Max A Little, Patrick E McSharry, Eric J Hunter, Jennifer Spielman, and Lorraine O Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *Biomedical Engineering, IEEE Transactions on*, vol. 56, pp. 1015–1022, 2009.

[7] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil, "Convex multi-task feature learning," *Machine Learning*, 2008.

[8] Rich Caruana, "Multitask learning," *Machine learning*, 1997.

[9] Ali Jalali, Sujay Sanghavi, Chao Ruan, and Pradeep K Ravikumar, "A dirty model for multi-task learning," in *Advances in Neural Information Processing Systems*, 2010.

[10] Chang Xu, Dacheng Tao, and Chao Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[11] Qiang Zhou, Gang Wang, Kui Jia, and Qi Zhao, "Learning to share latent tasks for action recognition," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013.