Poster Abstract: A Maximal Correlation Embedding Method for Multilabel Human Context Recognition

Lu Li
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University
lilou16@mails.tsinghua.edu.cn

Yang Li
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University
yangli@sz.tsinghua.edu.cn

Xiangxiang Xu
Dept. of Electronic Engineering
Tsinghua University
xuxx14@mails.tsinghua.edu.cn

Lin Zhang
Tsinghua-Berkeley Shenzhen Institute
Tsinghua University
linzhang@tsinghua.edu.cn

ABSTRACT

Real-time human context recognition is one of the most exciting emerging technologies in sensing nowadays. Compared with most recognition problems in machine learning, the challenge lies in the complexity and incompleteness of labels, in other words, each sample can have several label concepts simultaneously but some of them could be missing. This poster proposes an effective approach for multilabel human context recognition with signals from sensors embedded in the wearable devices. The proposed algorithm demonstrates to be very robust to incomplete labels.

ACM Reference Format:
Lu Li, Yang Li, Xiangxiang Xu, and Lin Zhang. 2019. Poster Abstract: A Maximal Correlation Embedding Method for Multilabel Human Context Recognition. In *The 18th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2019) (IPSN '19), April 16–18, 2019, Montreal, QC, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3302506.3312601

1 INTRODUCTION

The health industry brings a boom in human context recognition. A person's context includes their location, activity and so on. Traditionally, such information are collected either by self-report or by the caregivers. Automatic recognition technology greatly reduces manpower requirements. More importantly, it can provide feedback at arbitrary time so that interventions could be taken in time.

In this work, we aim to solve the multilabel human context recognition problem. To reflect the real situation, daily-used wearable devices are used to obtain data. The dataset used in this paper was collected by Vaizman et al. [2]. We adopted signals from six core sensors and used the same input features to be consistent with the data preprocessing steps in [3].

The positive labels of each instance are selected from a large candidate label set by the user themselves. This could result in incomplete labeling, so in the poster we mitigate classification bias due to incomplete labels with a maximal correlation type regularization. Our work was compared with the original work in [3], i.e. Multiple Layer Perceptron(MLP).

2 PROPOSED METHOD

Low rank label transformation is usually adopted in the multilabel classification problems, in which labels are embedded into

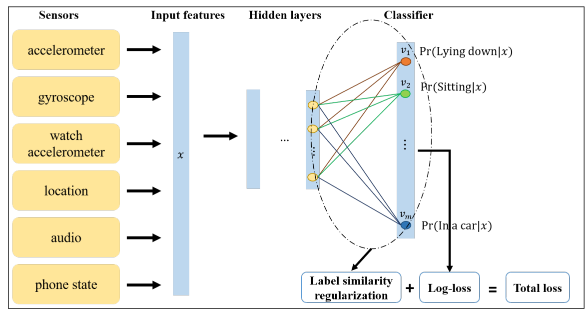

Figure 1: The Framework of Human Context Recognition

dense vectors with lower dimension than the original label space. While previous works implement label transformation using matrix decomposition or alternating optimization, both are limited in nonlinear expressiveness and scalability for large datasets. To overcome these flaws, we integrate the idea of low rank label transformation into a network shown in Figure 1. We restrict the dimension of the last hidden layer to be smaller than the label cardinality, and the weights connected between the last hidden layer and the output layer could be regarded as embedding vectors for labels. For example, in our problem, the first unit in the output layer stands for the probability of lying down, and all the weights connected between this unit and the units in last hidden layer compose a vector denoted as v_1, then v_1 is the embedding vector for the label "Lying down".

In the case with incomplete labels, the missing labels are mixed up with negative ones. This could bring noise to the training process. Consider the following scenario: a training instance is labeled with "Cooking", but missed with a positive label "At home". Then the training optimizer would treat the label "At home" as negative and compute the gradients in a totally wrong direction. However, the label "Cooking" and "At home" often co-occur in the same instance in the whole training data. In other words, these two labels are highly correlated. The correlations among labels can be utilized to mitigate the problem caused by missing labels. Based on the claim, we proposed a label similarity regularization on the embedding vectors for labels.

Huang et al. [1] proposed the Generalized Maximal Correlation (GMC) to maximize the total correlation among multiple variables. We transform the original optimization problem with constraints

Table 1: Dataset Description

Dataset	Number of Instance	Dimension of Features	Label Cardinality	Average Positive Labels per Sample
ExtraSensory	377346	175	51	3.4

Table 2: Performance in Two Settings for Missing Labels

	Setting I		Setting II	
	MLP	Ours	MLP	Ours
sensitivity	0.773	0.763	0.145	0.629
specificity	0.773	0.905	0.959	0.953
accuracy	0.773	0.885	0.935	0.933
balanced accuracy	0.773	0.834	0.552	0.791

into a relaxed GMC problem:

$$\underset{g_1,g_2,\dots g_n}{\text{maximize}} \quad \mathbb{E}\left[\sum_{i\neq j}(g_i(Y_i))^{\mathrm{T}}(g_j(Y_j))\right] - \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\|g_i(Y_i)\|^2\right] \qquad (1)$$

in which $g_1, g_2, \dots g_n$ stand for the optimal transformation functions for variables $Y_1, Y_2, \dots Y_n$. In our scenario, we wish to obtain optimal embedding vectors for multiple labels so that the labels frequently co-occur in the same instance will be closer in the embedded label space. Then we weighted the first term of the optimization objective in (1) using the coexist counts between label pairs:

$$\underset{v_1,v_2,\dots v_n}{\text{maximize}} \quad \frac{1}{m}\left(\sum_{i=1}^{n}\sum_{j\neq i}S_{i,j}\cdot v_i^{\mathrm{T}}v_j - \frac{1}{2}\sum_{i=1}^{n}S_{i,i}\,\|v_i\|^2\right) \qquad (2)$$

in which m,n are the number of instances and labels respectively, $v_1, v_2, \dots v_n$ stands for the embedding vectors for labels, and $S_{i,j}$ indicates how many times that label i and label j co-occur. We negated the objective function in (2) to form the label similarity regularization term Ω. Minimizing Ω is equivalent to maximize the total correlation among labels.

We consider two common missing label settings in real-world applications: 1)Setting I, i.e. the positions of missing labels are known; 2)Setting II, i.e. the positions of missing labels are hidden. In the former setting, we can eliminate loss components in the missing positions with a mask matrix. In particular, $\text{mask}_j^{(i)} = 0$ means that the label j is missing for the i^{th} sample, otherwise $\text{mask}_j^{(i)} = 1$. We use the summation of sigmoid-cross-entropy loss for each label as the basic classification error, and denote it as a log-loss defined by:

$$\text{log-loss} = -\sum_{j=1}^{n}\sum_{i=1}^{m}\text{mask}_j^{(i)}\cdot l(y_j^{(i)}, \hat{y}_j^{(i)}) \qquad (3)$$

More precisely, $l(y_j^{(i)}, \hat{y}_j^{(i)})$ is the log-loss between prediction $\hat{y}_j^{(i)}$ and ground truth $y_j^{(i)}$. Since different losses have different scales, we use a hyper-parameters α to weight label similarity regularization. Then the total loss function is

$$\text{total loss} = \text{log-loss} + \alpha\Omega \qquad (4)$$

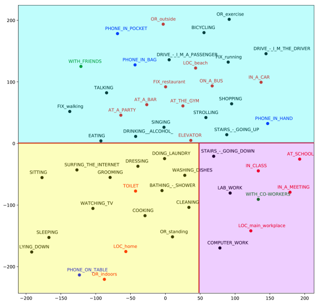

Figure 2: t-SNE Visualization for Label Embeddings

3 EXPERIMENTS

To demonstrate the effectiveness of our approach, we did 5-hold cross validation experiments on the ExtraSensory dataset. The statistic description about the dataset is shown in Table 1.

Four evaluation metrics are chosen, including sensitivity, specificity, balanced accuracy and accuracy, where the balanced accuracy is most revealing and important. We implement the network with two hidden layers with relu activation and batch normalization unit, and select the optimal value for hyper parameter α using grid search strategy. Our method is compared with the original MLP method in [3] under two settings mentioned above. As shown in Table 2, our method performs much better in both settings. Moreover, our approach shows strong stability in Setting II with hidden missing positions: the balanced accuracy only drops **0.043** while the compared method drops **0.221**.

Figure 2 displays the t-SNE visualization for all the label embeddings. Interestingly, those labels could be divided into three parts with only two linear classifiers. And labels in the same block show strong correlations, i.e. the labels in the yellow block are mostly indoor-related. Some labels such as "stairs going down" and "stairs going up" are very close, but almost impossible to coexist, we can infer that the input features of these labels may be very similar.

4 CONCLUSION AND FUTURE WORK

We proposed a multilabel classification algorithm, which is proved to be robust in case with hidden missing labels. In the future, we will validate our approach in multimedia data such as images and texts, and compare it with other methods in the related field.

ACKNOWLEDGMENTS

This work is funded by Shenzhen Municipal Scientific Program GJHZ20170314112258560, JCYJ20170818094022586.

REFERENCES

[1] Huang, S.-L., Zhang, L., and Zheng, L. An information-theoretic approach to unsupervised feature selection for high-dimensional data. In *Information Theory Workshop (ITW), 2017 IEEE* (2017), IEEE, pp. 434–438.

[2] Vaizman, Y., Ellis, K., and Lanckriet, G. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Computing 16*, 4 (2017), 62–74.

[3] Vaizman, Y., Weibel, N., and Lanckriet, G. Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1*, 4 (2018), 168.