SFB TRR 169 Crossmodal Learning

Project A4: Crossmodal joint sparse feature learning in robot tasks

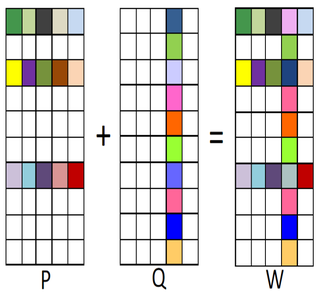
Changshui Zhang &Jianwei Zhang , Tsinghua University

Aims and Strategies

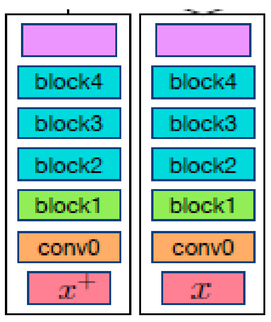
- 1. Establishing a multimodal feature learning method that is adaptive to the robot scenario using the modalities of image, sound and tactile signals as input.
- 2. Designing classifiers that are adaptive to new unknown categories.
- 3. Applying the trained model to the robot system and complete the task of assisting elderly people in taking medicine.

Previous Work and Pilot Data

- **robust multi-task feature learning** method, which decomposes the weight matrix into two matrices with different sparse regularization added.
- **the model with capped-L1 norm**, which could learn shared features and the specific feature of each task. In general, the capped-L1 norm is a better approximation than the L0 norm and achieves a better performance.
- **the metric Learning algorithm symmetry Match-Net**, which could mearsure the metric between two audio clips using two same structure of deep neural network.



Weight matrix can be decomposed into a row sparse matrix P and a column sparse matrix Q.



Match-Net

Methods

- **Multimodal feature learning method.**
- Image features——SIFT,HOG.
- Sound features——MFCC.
- Tactile features——A matrix signal is received tactile sensors.
- **Classifiers adaptive to unknown categories.**
- Methods based on sparse representation—— Establish a dictionary based on typical samples from all categories and for each sample, seek the sparse representation in all dictionaries.
- Methods based on learning the category boundaries.——Learn an estimation of the category border through training samples, the test samples outside of the border can be taken as samples from a new category.

Gong, P., Zhang, C., Lu, Z., Huang, J., & Ye, J. (2013). A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. *Machine learning: proceedings of the International Conference on Machine Learning.* Vol. 28. No. 2. NIH Public Access.

Gong, P., & Zhang, C. (2011). A fast dual projected Newton Method for L1-regularized least squares. *The 22nd International Joint Conference on Artificial Intelligence(IJCAI),* Barcelona, Spain, July 16-22.

He, J., & Zhang, J. (2014). In-hand haptic perception in dexterous manipulations. *SCIENCE CHINA-INFORMATION SCIENCES* Volume: 57 Issue: 12 Article Number: 120207 .

Hypotheses

The categories of medicine can be distinguished by different kinds of signal such as image, audio or tactile information.

The multimodal feature could fetch better than monomodal one for the categories information **is shared by different kinds of feature**.

The joint feature of the sample is supposed to be the **sparse representation** so that the learned joint feature has the separability for different categories of samples.

Work Programme & First Results

- **Task 1: Acquisition and annotation of image, sound and tactile data**
- For this task we will exploit the robot platform set up by project Z3/II-R. Here, we take the robot for elderly assistance as a background, but our methods are general crossmodal learning methods.
- **Task 2: Algorithm for multimodal joint feature learning**
- we will mainly focus on multi-task sparse learning algorithms. The robot processes different tasks using multimodal features. While these tasks are different, they share some features. Such as Robust Multi-Task Learning. And Multi-Stage Multi-Task Learning Algorithm.
- **Task 3: Classifiers suitable for new object categories**
- Having received information about the medicine (e.g. name, shape, colour), the robot needs to confirm whether the medicine it selected corresponds to the required medicine. In this context, the selected medicine could be unknown and therefore could belong to unknown categories. The robot should be able to recognize objects from new categories, which is also a main problem to be solved.
- **Task 4: Application of the models and methods to a robot system**
- Based on the feature extraction method and classifier design method in Task 2 and Task 3, the models will be applied to the robot-assisted elderly medication task in order to verify our approach. The model will be tested in different environments and with different medicine, in order to observe whether the robot can perform correctly, i.e., whether the robot can deliver the right medication according to the joint multimodal learning methods.

Role within the SFB

- This project studies and develops new machine learning methods and algorithms which directly serve the main goal of the TRR.
- The main task of this project is to study the joint multimodal learning methods in the scenario of robotic assistance of elderly people in taking medicine.