# Multi-sensor classification via Consensus-based Multi-View Maximum Entropy Discrimination


*Tianpei Xie[1], Nasser M. Nasrabadi[2], Alfred O. Hero[1]*

[1]Electrical Engineering, Systems, University of Michigan, Ann Arbor, [2]U.S. Army Research Lab., 2800 Powder Mill Road, Adelphi, MD.


## Motivations

- Samples from different sources can be interpreted as corresponding to different views of the same target.
- Learning in the multi-view setting is challenging:
  - Information **Fusion** ?
  - **Parsimony** *vs.* **Robustness** ?
  - **Unlabeled** samples ?
- In this work, we proposed a **stochastic consensus-based multi-view learning** framework [7] to handle these issues.

## Comparison to previous work

- Comparison of conventional fusion methods and the proposed consensus-based multi-view learning method (in **bold**)

| | fusion stage | parsi-mony | semi-sup. | noise tol | Bayes. | #. views |
|---|---|---|---|---|---|---|
| CCA | feature | x | √ | x | x | 2 |
| Bi-DAE | feature | x | √ | x | x | 2 |
| SVM-2K | feature | x | √ | x | x | 2 |
| Bayes-Fusion | decision | √ | x | √ | √ | ⩾ 2 |
| Boosting | decision | √ | √ | x | x | ⩾ 2 |
| Co-training | consens. | √ | √ | √ | x | 2 |
| Bayes Co-trn | consens. | x | √ | √ | √ | ⩾ 2 |
| MV-MED | consens. | √ | √ | x | √ | 2 |
| **CMV-MED** | consens. | √ | √ | √ | √ | ⩾ 2 |

## Our Contributions

1. We learn view-specific posterior distributions as features.
2. The proposed method maximizes the **stochastic agreement** among different models on **unlabeled** samples.
3. The proposed information-theoretical consensus measure is *robust* to noisy samples and outliers.

## Notations and Model

- Multiview observation model:

| | |
|---|---|
| $V$-view feature $\&$ target domain | $\mathcal{X}_1 \times \ldots \times \mathcal{X}_V \times \mathcal{Y}$; |
| Labeled i.i.d. samples | $(\mathbf{x}_n, y_n), n \in L$: labeled set. |
| Unlabeled i.i.d. samples | $\mathbf{x}_m, m \in U$: unlabeled set; |
| view-specific model | $\log p_i(y\|\theta^i(\mathbf{w}_i, \mathbf{x}^i)) \propto \frac{1}{2} y\theta^i\left(\mathbf{x}^i, \mathbf{w}_i\right) \in \mathcal{M}_i$, $\theta^i_m \equiv \theta^i(\mathbf{w}_i, \mathbf{x}^i_m)) \in \Theta_i, 1 \le i \le V$ |
| view-specific loss functional | $\mathcal{L}_i : \mathcal{Y} \times \mathcal{X}_i \times \mathcal{M}_i \to [0, \infty)$. e.g. $= \left[\xi_i - \log \frac{p_i(y_n\|\theta^i(\boldsymbol{w}^i, \boldsymbol{x}^i_n))}{p_i(y \neq y_n\|\theta^i(\boldsymbol{w}^i, \boldsymbol{x}^i_n))}\right]_+$ |

- Consensus part:

| | |
|---|---|
| Consensus-view model | $q(y\|\theta_m) \in \overline{\mathcal{M}} = \oplus_{i=1}^V \mathcal{M}_i, m \in U$. |
| Parameters of CV model | $\theta_m = \sigma(\theta^1_m, \ldots, \theta^V_m)$. |
| Pairwise consensus measure | $\mathbb{D}\left(\cdot \,\|\, \cdot\right) : \overline{\mathcal{M}} \times \mathcal{M}_i \to [0, \infty)$ e.g. $= \mathbb{KL}\left(q \,\|\, p_i\right)$, K-L divergence. |

- Prior distribution:

| | |
|---|---|
| Prior on view params and threshs | $p_0(\boldsymbol{\theta}^{[V]}, \boldsymbol{\xi}^{[V]}) = \prod_{j=1}^V p_0(\boldsymbol{\theta}^i, \boldsymbol{\xi}^i) = \prod_{j=1}^V p_0(\boldsymbol{\theta}^i) p_0(\boldsymbol{\xi}^i)$. |

The proposed ***Consensus-based Multi-view Maximum Entropy Discrimination*** (CMV-MED) [7] is to simultaneously learn the joint post-data distributions $q_i(\theta^i, \boldsymbol{\xi}^i)$, given the priors $p_0(\boldsymbol{\theta}^{[V]}, \boldsymbol{\xi}^{[V]})\ \forall i = 1, \ldots, V$, by solving the following optimization problem

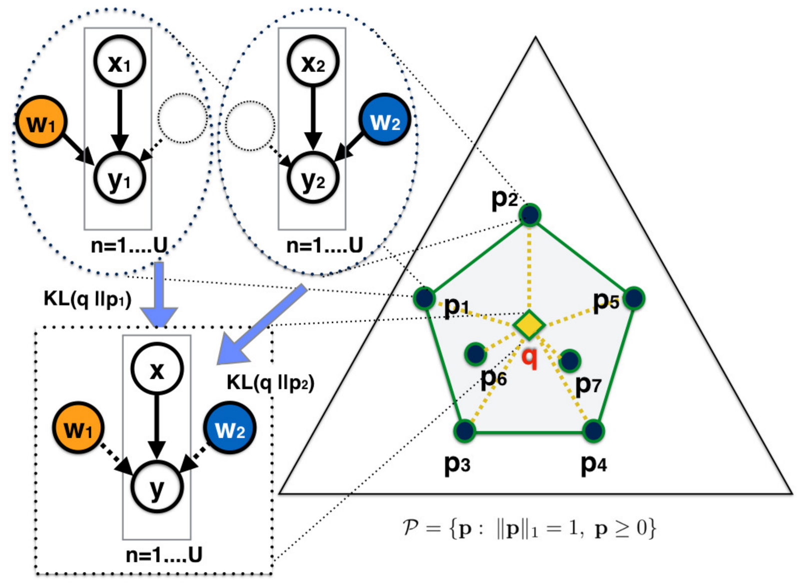
$$\min_{\substack{q_i(\theta^i, \boldsymbol{\xi}^i) \in \Delta,\\ i=1,\cdots,V,\\ q(y|\theta_m) \in \oplus_{i=1}^V \mathcal{M}_i, \rho > 0}} \sum_{i=1}^V \mathbb{KL}\left(q_i(\theta^i, \boldsymbol{\xi}^i) \,\|\, p_0(\theta^i) p_0(\boldsymbol{\xi}^i)\right) + \lambda\, \rho$$

$$\text{s.t.} \sum_{n \in L} \mathbb{E}_{q_i}\left[\mathcal{L}_i(y^i_n, \boldsymbol{x}^i_n, p_i(y|\theta^i))\right] \le 0, \quad 1 \le i \le V$$

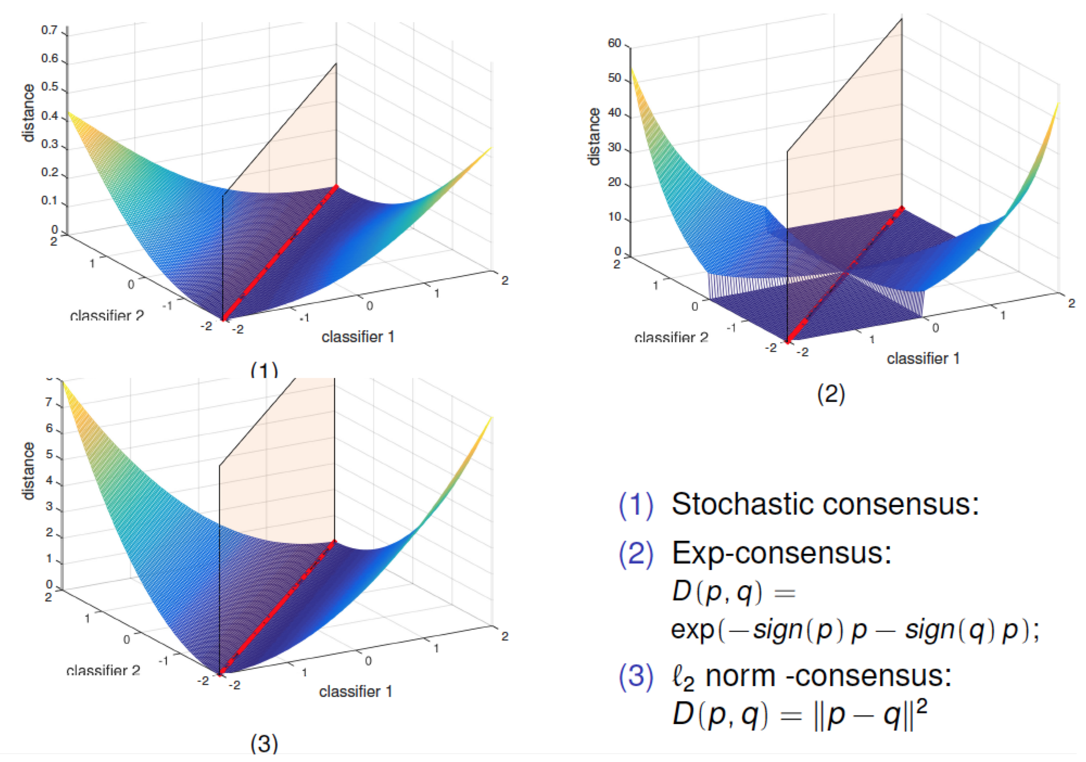
$$\sum_{m \in U} \sum_{i=1}^V \mathbb{KL}\left(q(y|\theta_m) \,\|\, p_i(y|\mathbb{E}_{q_i}\left[\theta^i|\, \boldsymbol{x}^i_m\right])\right) \le \rho. \quad (1)$$

[**Remarks**]:

1. The consensus view model $q(y|\theta_m)$ serves as a **relabeling** model and is **shared** among all views.
2. Given $q(\theta^i, \boldsymbol{\xi}^i)$, $q(y|\mathbf{x})$ is learned by information projection.




3. The constraint (1) induces a robust dissimilarity measure




## Variational EM Algorithm

For $t = 1, \ldots,$ until convergence

1. E-step: Given $\hat{\theta}^i_{t-1} = \mathbb{E}_{q_{t-1}}\left[\theta^i\right], 1 \le i \le V$, the consensus model is the log-average $\log q_t(y|\mathbf{x}_n) = \frac{1}{V}\sum_{i=1}^V \log p_{i,n}(y|\hat{\theta}^i_{t-1}) - \log Z(\mathbf{x}_n),\ n \in U,$
2. M-step: Given $q_t(y|\mathbf{x}_n)$,, solve for $q(\theta^i, \boldsymbol{\xi}^i)$ independently. In particular, each view corresponds to a maximum entropy learning problem with maximal margin error constraint $\mathcal{L}_i$. Under proper prior on soft-threshold $\xi$, it can be solved efficiently using SVM-like dual optimization. See e.g. [4].

## Experiments

We compare the proposed CMV-MED model with the SVM-2K [3], the MV-MED [6] as well as the conventional MED for each view on two real multi-view data sets: ARL-footstep[2] and WebKB4[1]

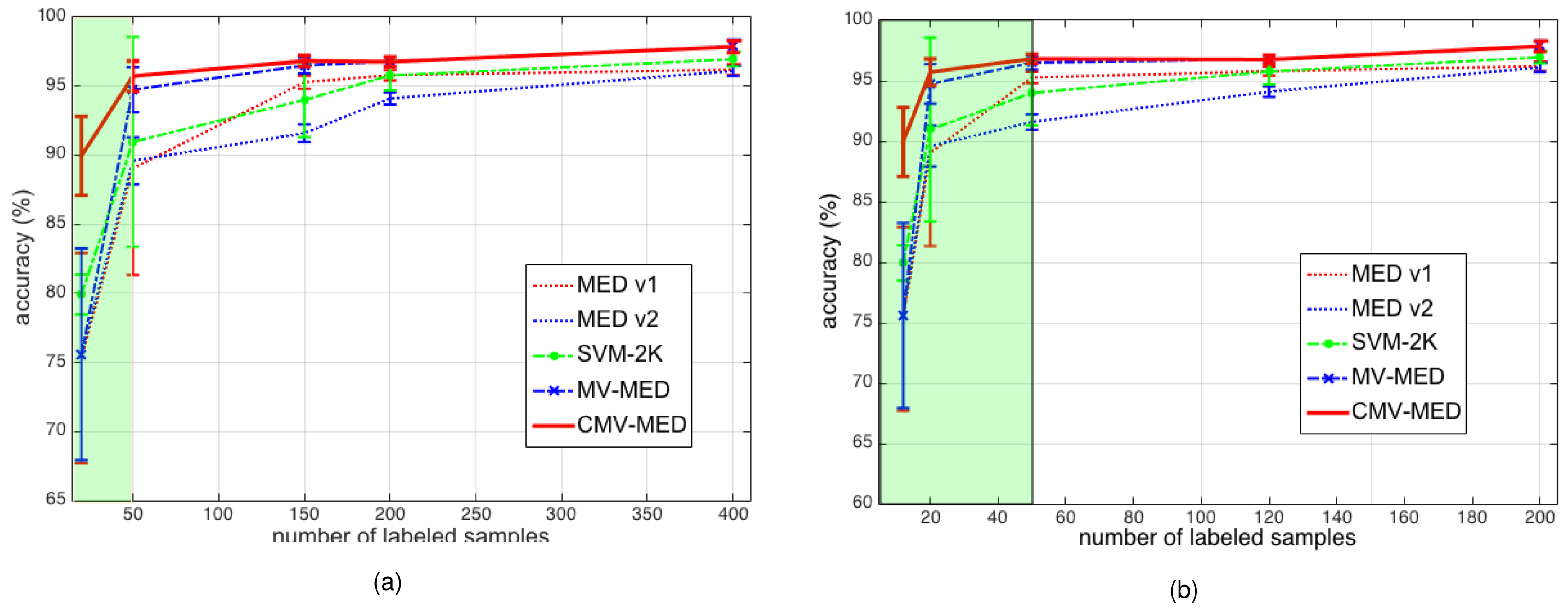



**Figure 1:** The classification accuracy vs. the size of labeled set for (a) **ARL-Footstep** data set [2], (b) **WebKB4** data set [1] The proposed CMV-MED outperforms MV-MED, SVM-2K and two single-view MEDs (view 1 and 2) and it has good stability when the number of labeled samples is small.


## Acknowledgements

This research was partially supported by US Army Research Office (ARO) grants W911NF-11-1-0391 and WA11NF-11-1-103A1.


## References


[1] Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Seán Slattery. Learning to construct knowledge bases from the world wide web. *Artificial intelligence*, 118(1):69–113, 2000.

[2] Thyagaraju Damarla, Asif Mehmood, and James Sabatier. Detection of people and animals using non-imaging sensors. *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–8, 2011.

[3] Jason Farquhar, David Hardoon, Hongying Meng, John S Shawe-taylor, and Sandor Szedmak. Two view learning: SVM-2K, theory and practice. In *Advances in neural information processing systems*, pages 355–362, 2005.

[4] Tommi Jaakkola, Marina Meila, and Tony Jebara. Maximum entropy discrimination. In *Advances in neural information processing systems*, 1999.

[5] Nicholas Kushmerick. Learning to remove internet advertisements. In *Proceedings of the third annual conference on Autonomous Agents*, pages 175–181. ACM, 1999.

[6] Shiliang Sun and Guoqing Chao. Multi-view maximum entropy discrimination. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 1706–1712. AAAI Press, 2013.

[7] Tianpei Xie, Nasser M Nasrabadi, and Alfred O Hero III. Semi-supervised multi-sensor classification via consensus-based multi-view maximum entropy discrimination. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 1936–1940. IEEE, 2015.