Clustering of Multiple Instance Data

Andrew Karem
Multimedia Research Lab

Advisor: Dr. Hichem Frigui

Outline

- **Background**
  - **Multiple Instance Learning Framework (MIL)**
  - **Target Concepts and Diverse Density**
- **Contributions**
  1. Optimizing Multi-concept MDD Metrics
  - **Crisp Clustering of Multiple Instance Data (CCMI)**
  - **Fuzzy Clustering of Multiple Instance Data (FCMI)**
  - **Possibilistic Clustering of Multiple Instance Data (PCMI)**

  2. Negative Target Concepts
  3. Embedded Feature Space Classification using Positive and Negative TCs.
- **Results and Analysis**
- **Conclusions**

Conventional Classification

Objective: Train a classifier to answer the question "Is this object a mushroom?"

Single Instance Learning

Training Data

Feature Extraction

1 to 1

$x_1=[x_{11}, x_{12}, .., x_{1m}]$

L=1

1 to 1

$x_2=[x_{21}, x_{22}, .., x_{2m}]$

L=0

1 to 1

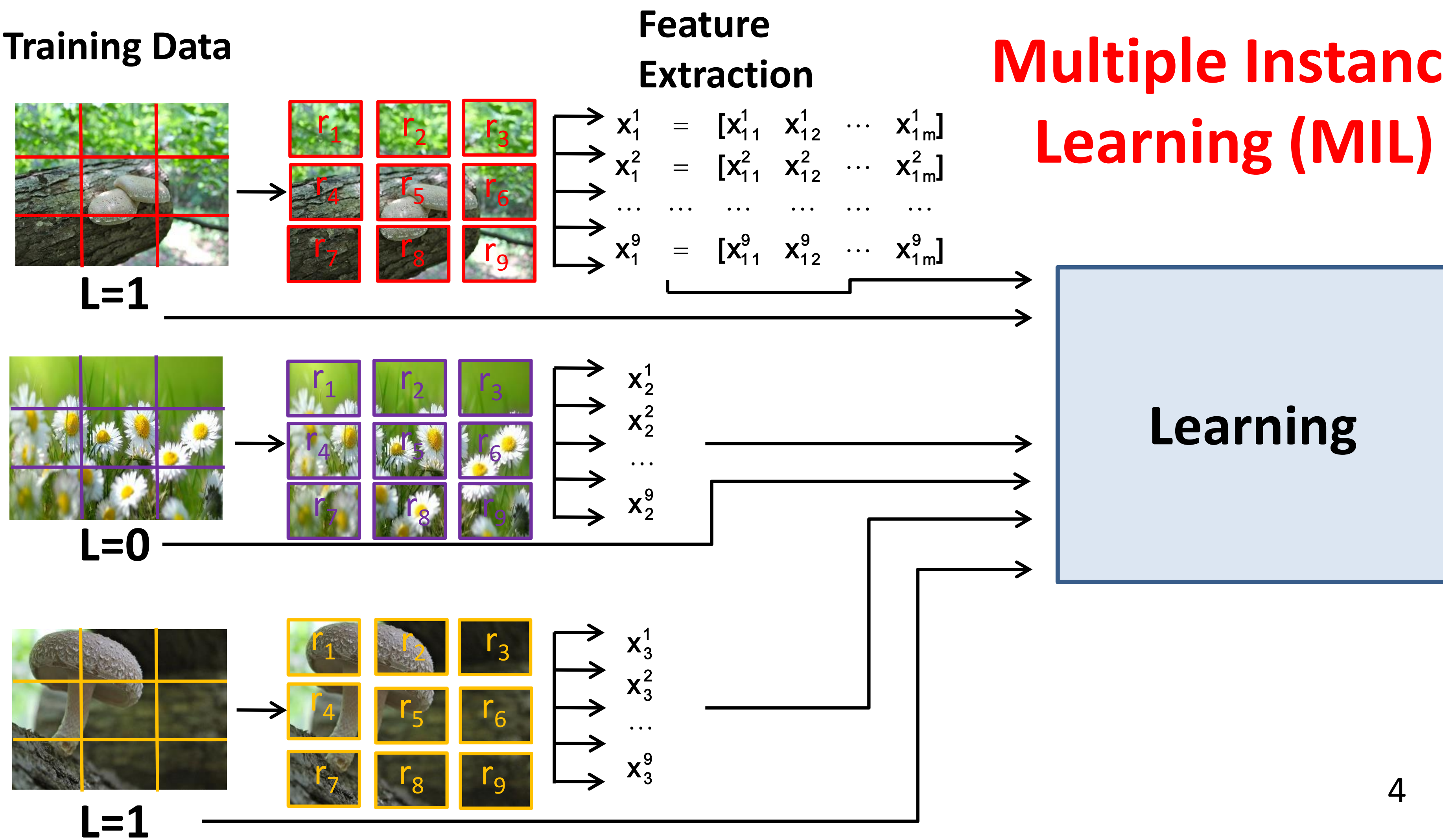
$x_3=[x_{31}, x_{32}, .., x_{3m}]$

L=1

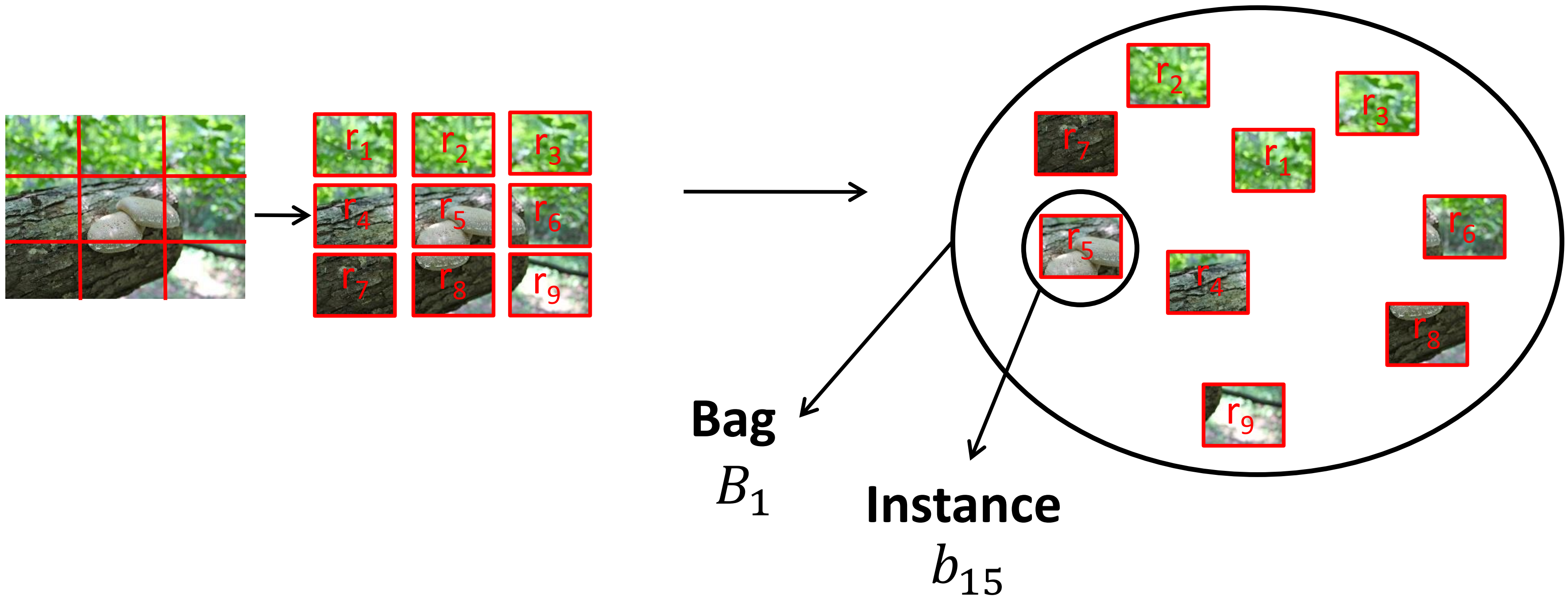
Learning

Train classifier to answer question
"Does this image contain mushrooms?"

Multiple Instance Learning (MIL)



MIL Terminology



$\boldsymbol{B_1} = \{\boldsymbol{b_{11}}, \boldsymbol{b_{12}}, .., \boldsymbol{b_{19}}\}$

Label(B_1) is known.

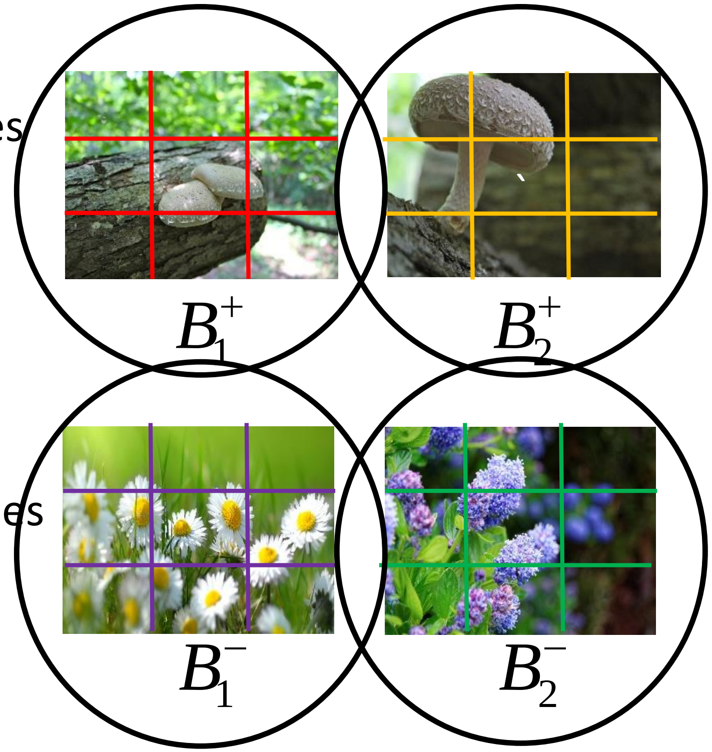
If positive, Label(b_{11}), Label(b_{12}), etc. are unknown.

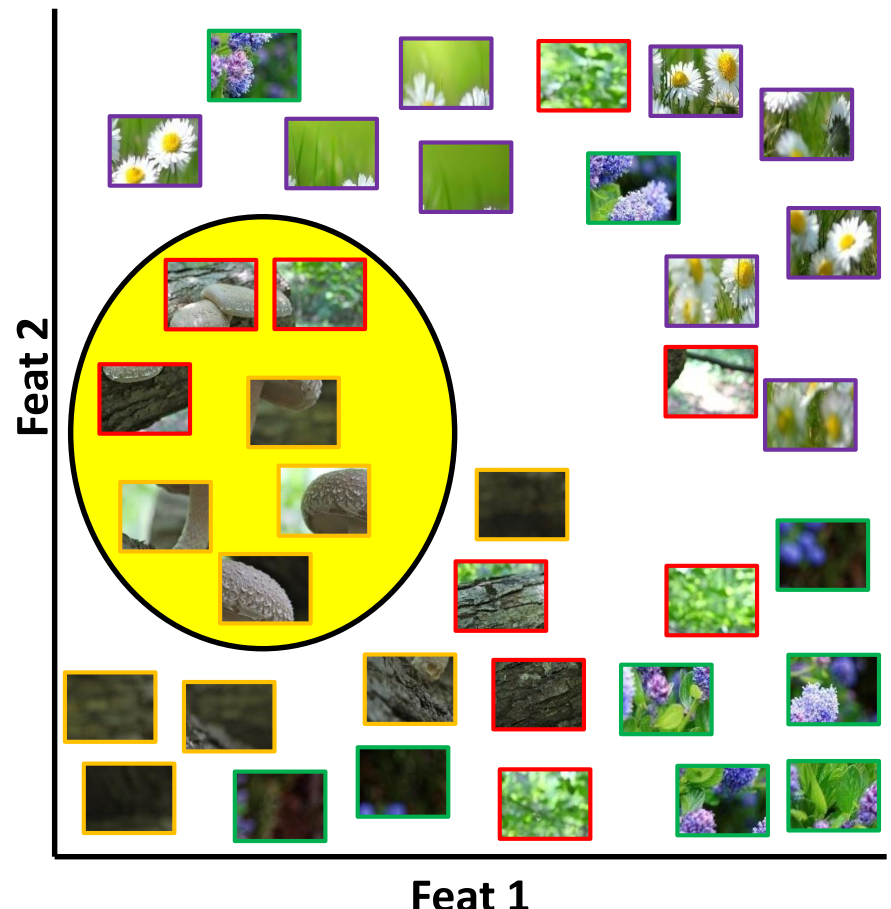
If negative, Label(b_{11}), Label(b_{12}), etc. are negative.

Instance Feature Space And Target Concepts

Positive Samples (Bags)

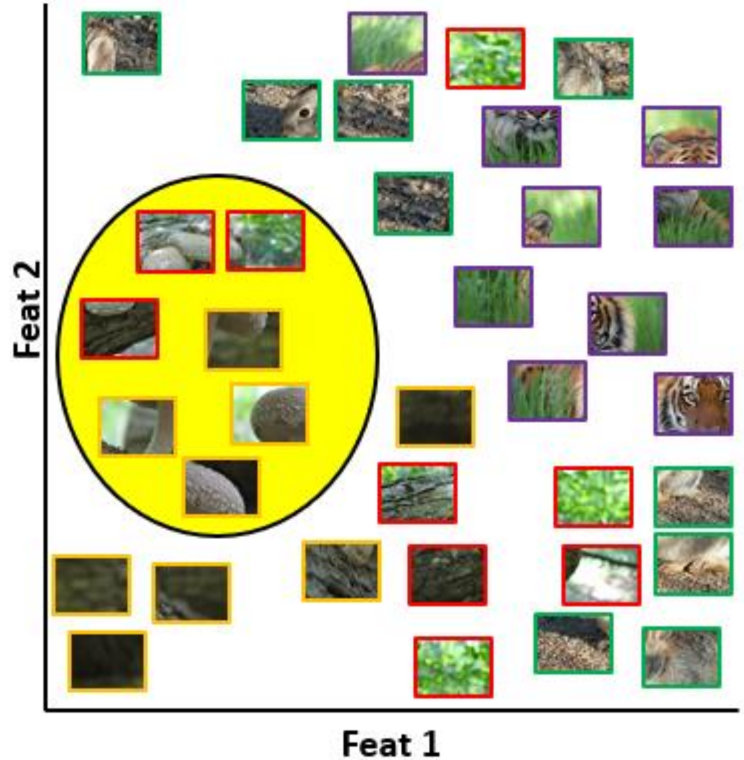
Negative Samples (Bags)





Target Concept t

Diverse Density (DD) Algorithm*

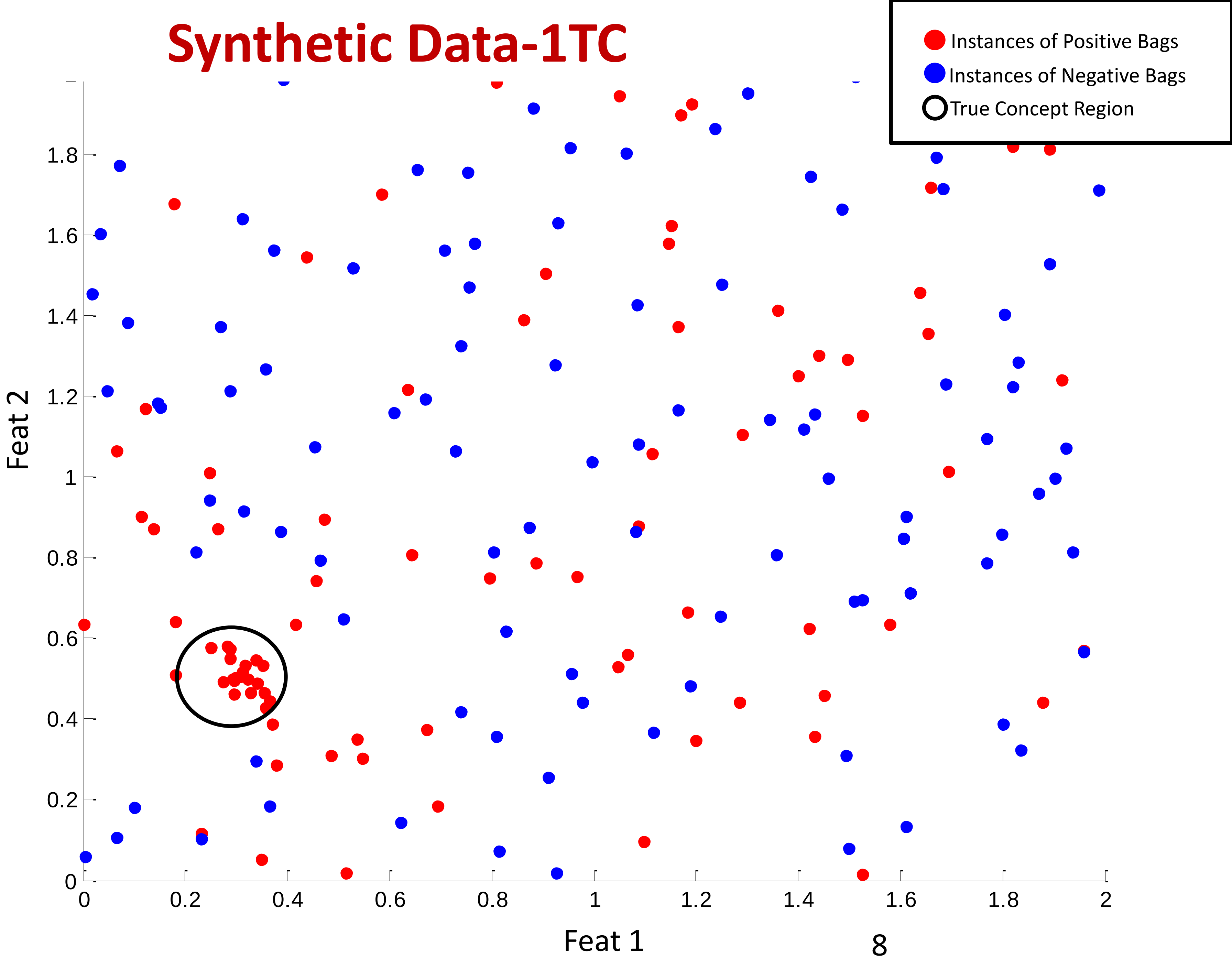


Idea: Locate target concept (e.g. mushroom concept)

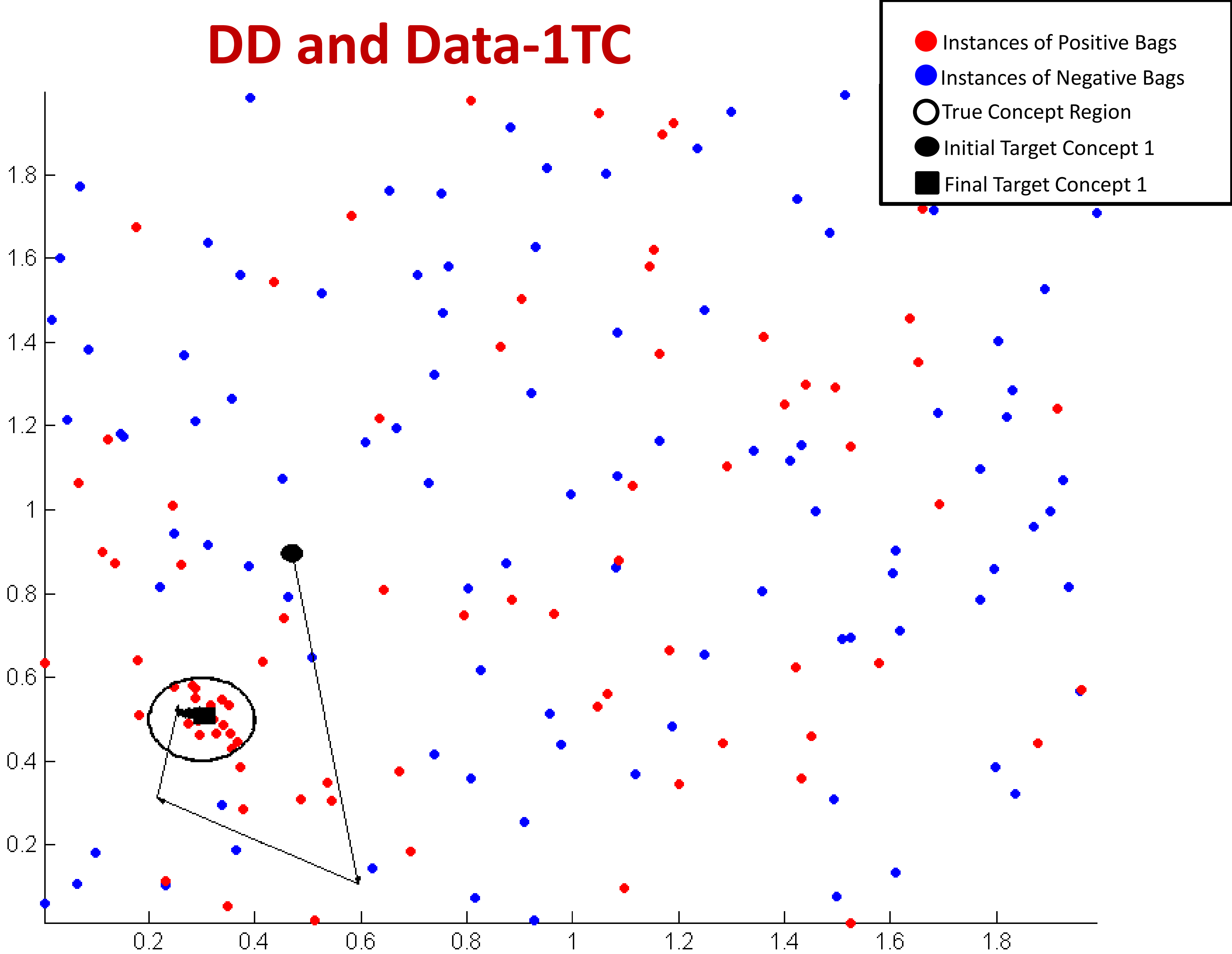
$$DD(t;B)=\left(\prod_{n=1}^{N^+}\Pr(t\mid B_n^+)\right)\left(\prod_{n=1}^{N^-}\Pr(t\mid B_n^-)\right)$$

*Maron, Oded, and Tomás Lozano-Pérez. "A framework for multiple-instance learning." *Advances in neural information processing systems* (1998): 570-576.

Synthetic Data-1TC



DD and Data-1TC



Not all mushrooms are created equal.

Limitation of DD

- For classes with high intra-class variation, one target concept is not sufficient.
- In this case, we require multiple target concepts to correctly model the data.
- Author of DD recommends running algorithm with multiple initializations.
- In our experiments this doesn’t work.

Proposed Work:
Fuzzy Multiple Concept Diverse Density (FMDD)

- **Assume we have K target concepts $t_1, t_2, \ldots, t_K$.**
- **Any bag B_n can belong to any concept t_k with membership u_{kn}**

$$T = \{t_1, t_2, \ldots, t_K\}$$

$$B = \{B_1, B_2, \ldots, B_N\}$$

$$U = \begin{bmatrix} u_{11} & u_{12} & \ldots & u_{1N} \\ u_{21} & u_{22} & \ldots & u_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ u_{K1} & u_{K2} & \ldots & u_{KN} \end{bmatrix}$$

$$FMDD(T,U;B) = \left(\prod_{n=1}^{N^+} \prod_{k=1}^{K} (\Pr(t_k \mid B_n^+))^{u_{kn}^m} \right) \left(\prod_{n=1}^{N^-} \prod_{k=1}^{K} (\Pr(t_k \mid B_n^-))^{u_{kn}^m} \right)$$

constraints $\sum_{k=1}^{K} u_{kn} = 1$ **and** $u_{kn} \in [0,1]$

Proposed Work: Optimizing the FMDD

Objective: Find $t_1,\ldots, t_K$ that maximize

$$\boxed{FMDD(T,U;B) = \prod_{n=1}^{N}\prod_{k=1}^{K}(\Pr(t_k|B_n))^{u_{kn}^m}}$$

subject to $\sum_{k=1}^{K} u_{kn} = 1$ and $u_{kn} \in [0,1]$

Equivalent to minimizing negative log-likelihood:

$$J(T,U;B) = \sum_{n=1}^{N}\sum_{k=1}^{K} u_{kn}^m\{-log(Pr(t_k|B_n))\}$$

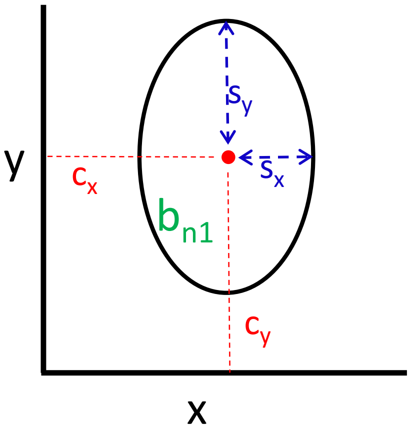
Bag and Instance Probabilities

$$\Pr(t_k \mid B_n) = \begin{cases} 1 - \prod_{i=1}^{I} \left[1 - \Pr(b_{ni} \in t_k)\right] & if \;\; B_n \in B^+ \\ \prod_{i=1}^{I} \left[1 - \Pr(b_{ni} \in t_k)\right] & if \;\; B_n \in B^- \end{cases}$$

NOISY-OR

$$\Pr(b_{ni} \in t_k) = e^{-\left(\sum_{l=1}^{L} s_{kl}^{\;2} (b_{nil} - c_{kl})^2\right)}$$

Point-and-Scale



ℓ=1:L are the individual instance features.

Fuzzy Clustering of Multiple Instance (FCMI) Algorithm

1) Initialize $c_1 \ldots c_K$ and scales $s_1 \ldots s_K$ for $t_1 \ldots t_K$

2) Repeat

2a) $\boxed{\text{Compute } U.}$

2b) For each target concept t_k, k=1..K

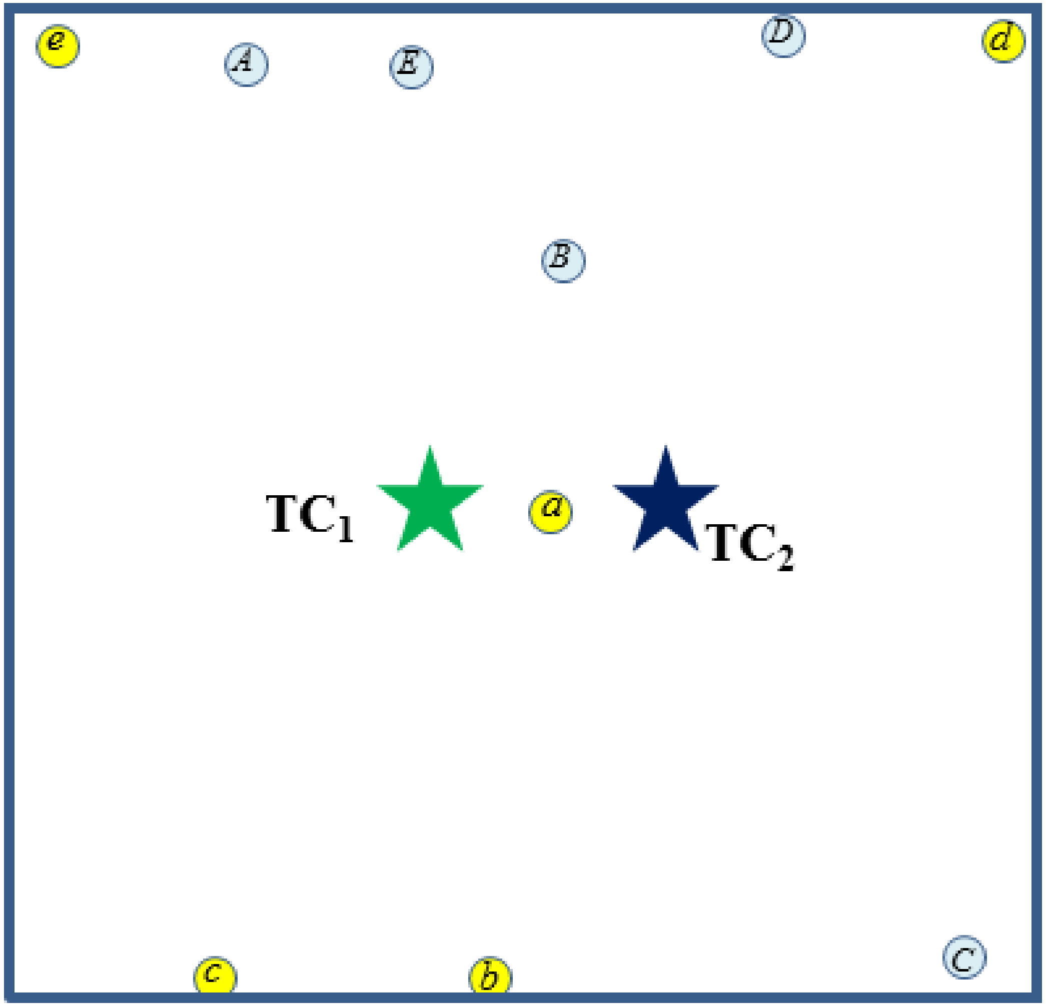
Repeat

Use gradient descent to optimize $\boxed{c_k \text{ and } s_k.}$

$$\frac{\partial \Pr(b_{ni} \in t_k)}{\partial c_{kl}} = 2(b_{nil} - c_{kl}) s_{kl}^2 e^{-(\sum_{l=1}^{L} s_{kl}(b_{nil}-c_{kl})^2)}$$

$$\frac{\partial \Pr(b_{ni} \in t_k)}{\partial s_{kl}} = 2 s_{kl} (b_{nil} - c_{kl})^2 e^{-(\sum_{l=1}^{L} s_{kl}(b_{nil}-c_{kl})^2)}$$

- Membership ambiguities with the Fuzzy Membership



- B1={A,B,C,D,E}
- B2={a,b,c,d,e}
- $u_{11} \approx u_{21} \approx 0.5$
- $u_{12} \approx u_{22} \approx 0.5$

Proposed Work: Possibilistic Multiple Concept DD (PMDD)

- We relax the membership constraint that $$\sum_{k=1}^{K} u_{kn} = 1$$

$$PMDD(\mathcal{T},\mathcal{U}) = \frac{\prod_{n=1}^{N}\prod_{k=1}^{K} Pr(t_k|B_n)^{u_{kn}^m}}{\prod_{n=1}^{N}\prod_{k=1}^{K} e^{\eta_k (1-u_{kn})^m}}$$

Possibilistic Clustering of Multiple Instance (PCMI) Algorithm

1) Initialize $c_1 \ldots c_K$ and scales $s_1 \ldots s_K$.

2) Repeat

   2a) $\boxed{\text{Compute U.}}$

   2b) For each target concept t_k, k=1..K

      Repeat

         Use gradient descent to optimize c_k and s_k.

      Until c_k does not change significantly.

$$u_{qn} = \begin{cases} \dfrac{1}{1-\left\{\frac{\log(Pr(t_q|B_n))}{\eta_q}\right\}^{\frac{1}{m-1}}} & if\ label(B_n)=1 \\ 1 & if\ label(B_n)=0 \end{cases}$$

Identifying Optimal Number of TC

1. Run PCMI with over-specified number of TC.

2. Merge compatible TCs.

$$\frac{\sum_{n=1}^{N^+} |u_{kn} - u_{k'n}|}{\sum_{n=1}^{N^+} |u_{kn}| + \sum_{n=1}^{N^+} |u_{k'n}|} < \theta_M$$

3. Eliminate TCs with limited positive bag response.

$$\frac{1}{N^+}\sum_{n=1}^{N^+} Pr(t_k|B_n^+) < \theta_{QN}$$

4. Eliminate TCs dominated by negative bag response.

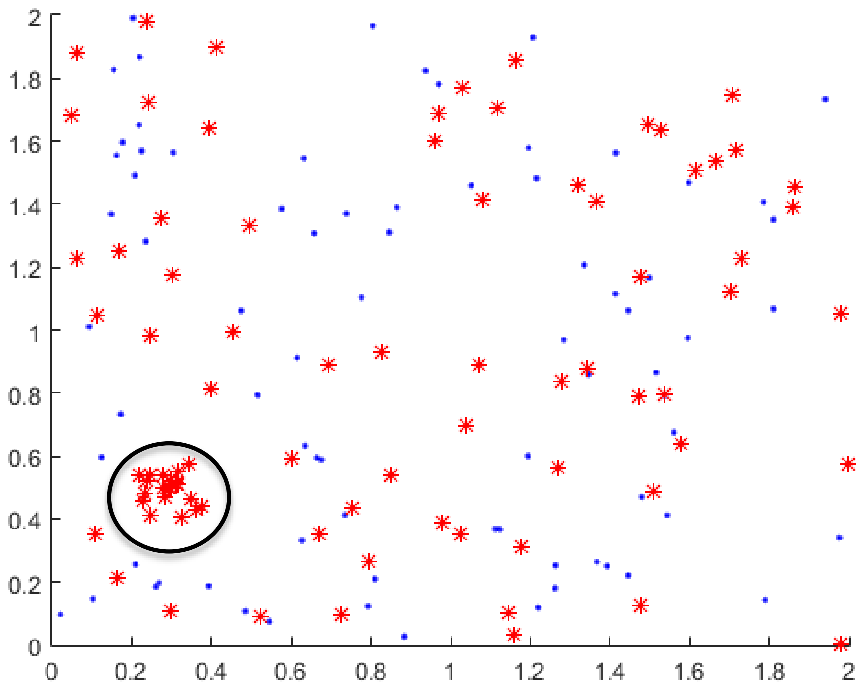
$$\frac{\frac{1}{N^+}\sum_{n=1}^{N^+} Pr(t_k|B_n^+)}{1 - \frac{1}{N^-}\sum_{n=1}^{N^-} Pr(t_k|B_n^-)} < \theta_{QL}$$

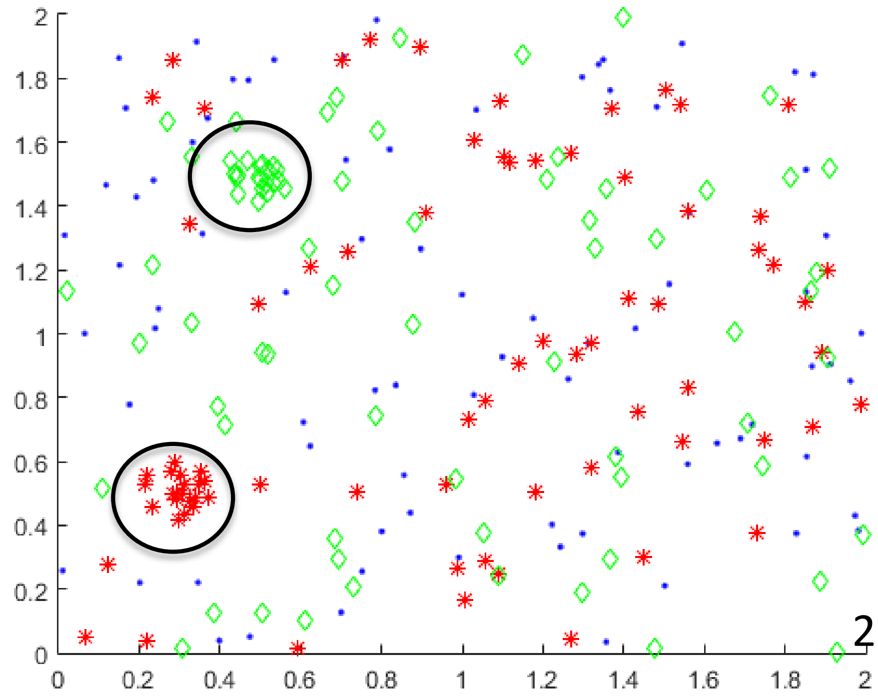
Results and Analysis: Illustrative Datasets

Illustrative Data Analysis

- Compare performance of FCMI and PCMI to DD.

- Data-1TC: Single target concept, 20 positive bags, 20 negative bags
- Data-2TC: Two target concepts, 20 positive bags each, 20 negative bags

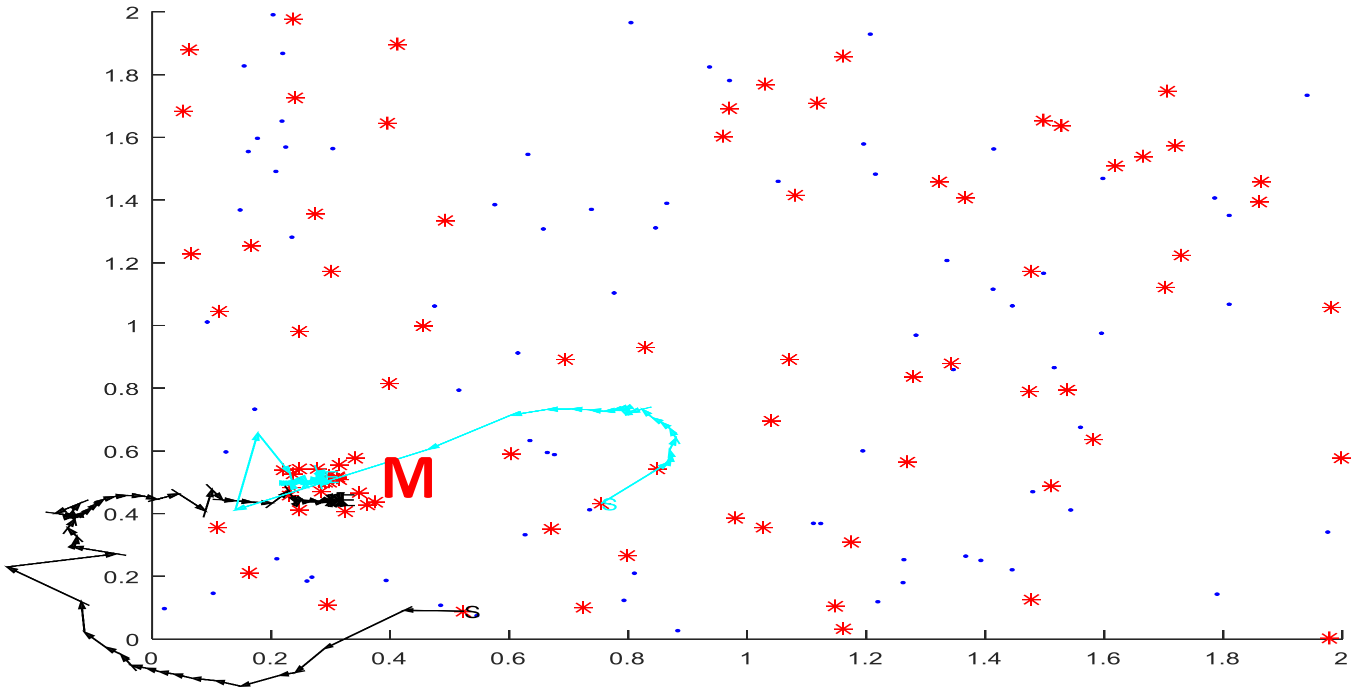




Experimental Results: Data-1TC

	DD	FCMI (K=1)	FCMI (K=2)*
Correct TrueC	74%	74%	86%
TrueC Failure	26%	26%	14%

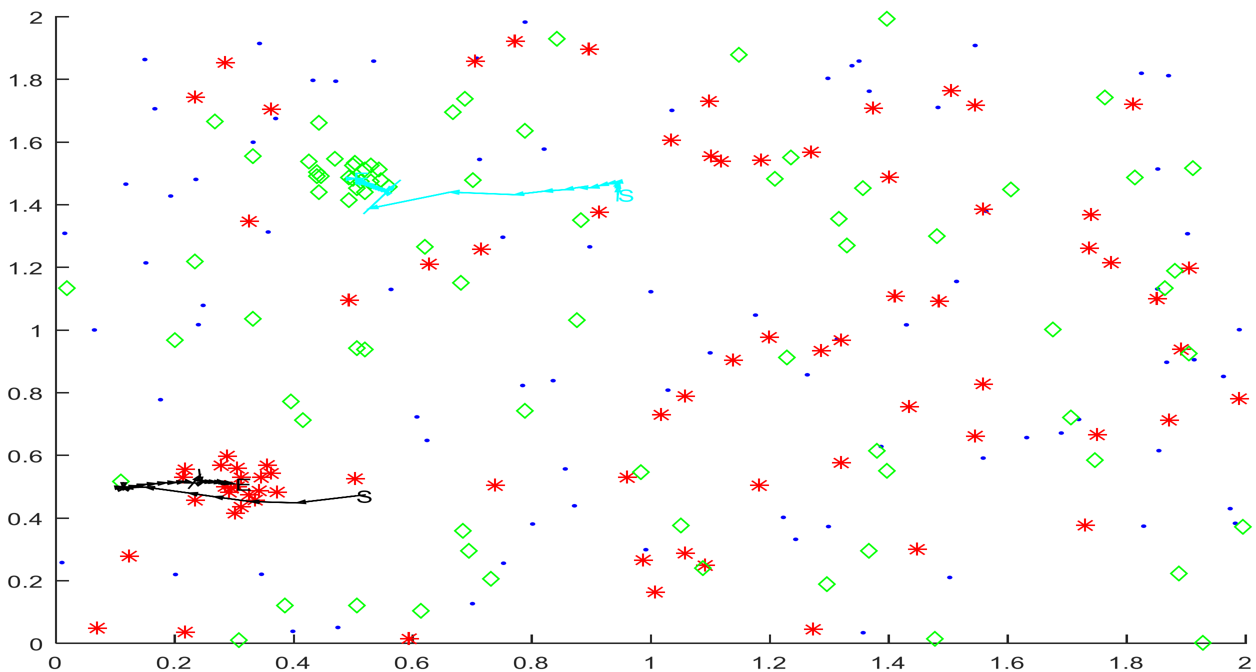
***After PCMI, Merging, and Weak TC Elimination**



Experimental Results: Data-2TC

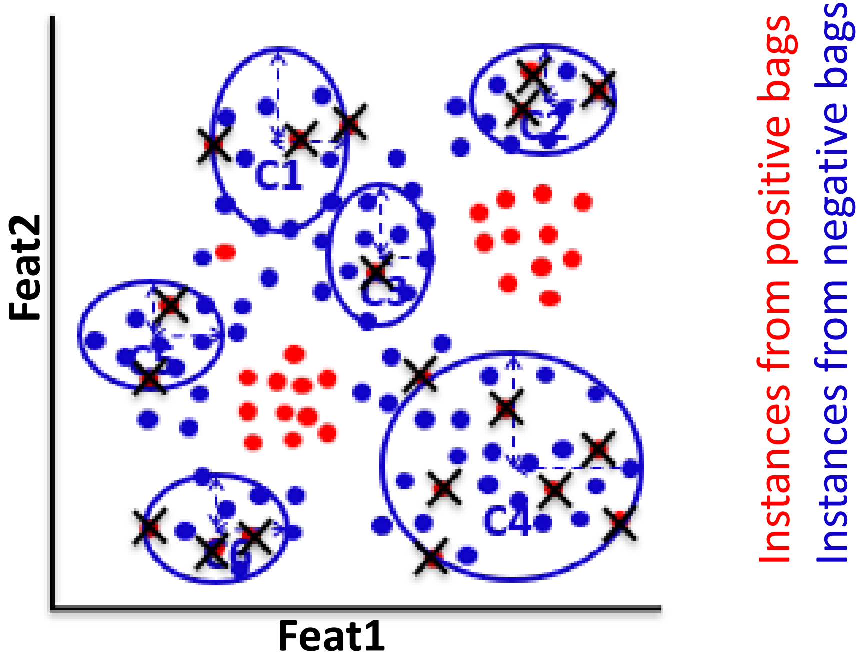
	DD	FCMI (K=2)	FCMI (K=4)*	FCMI (K=2)**
2 Correct TrueC	N/A	46%	88%	100%
1 Correct TrueC	0%	20%	12%	0%
TrueC Midpoint	62%	22%	0%	0%
TrueC Failure	38%	12%	0%	0%

***After PCMI, Merging, and Weak TC Elimination**
***KDE-based Initialization**



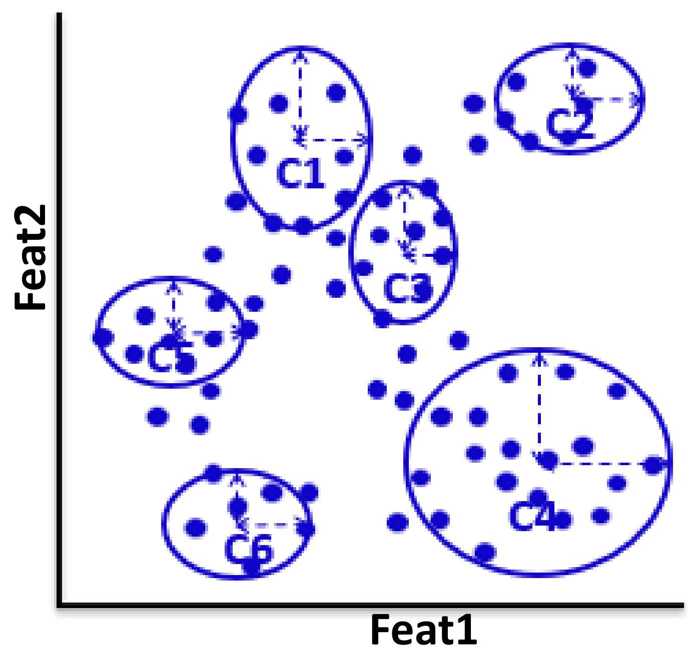
KDE Initialization and Negative TCs

$$KDE(x) = \exp\left(\sum_{j=1}^{d} -\beta||x - C_d\,||\right)$$



Mark any instance x from positive bags for which KDE(x)>θ_{KDE}.

Cluster all negative instances.



Cluster remaining "likely positive" instances.

Results and Analysis: Sensitivity Datasets

How does the FCMI respond to changes in data generation parameters?

- We generated 1100 datasets corresponding to 11 parameters and 10 distinct values per parameter.

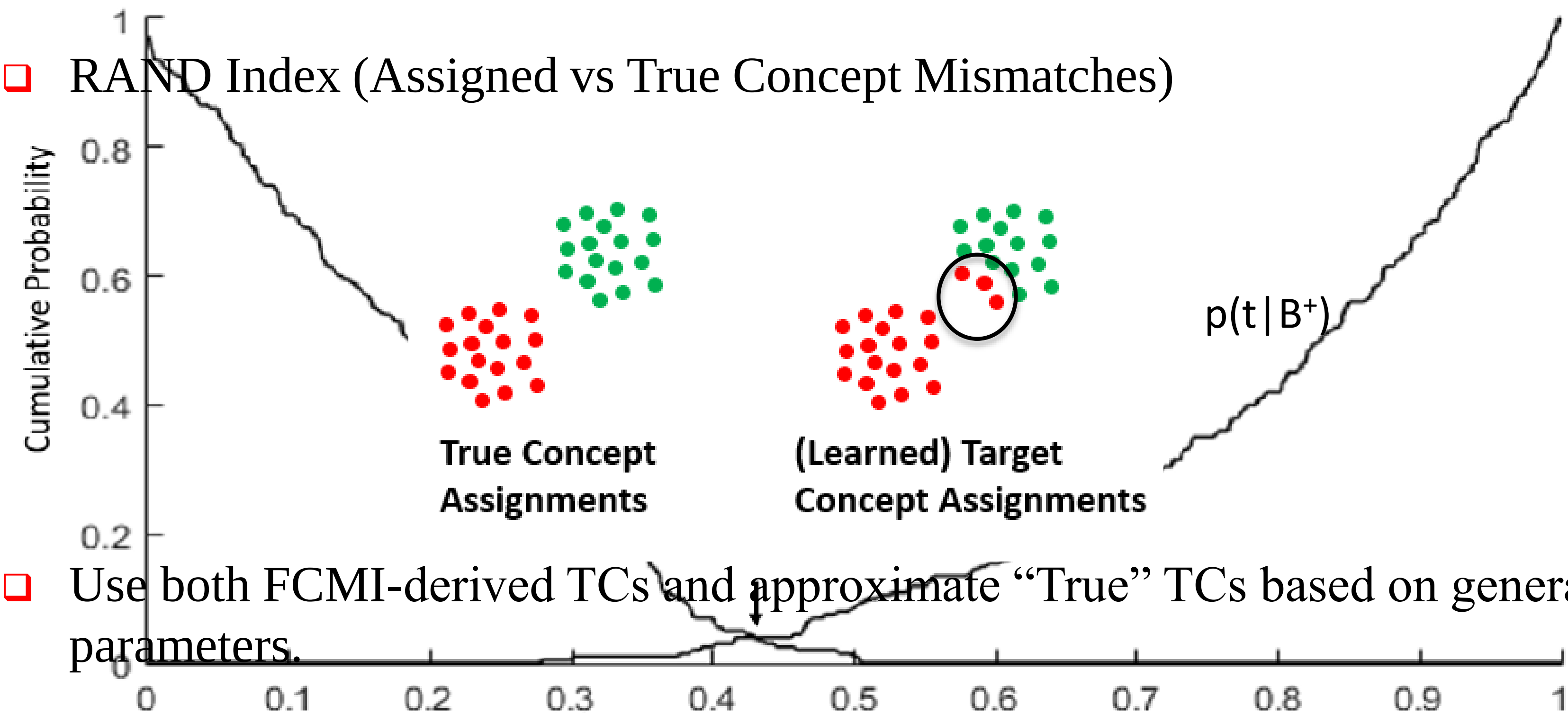
1. Bag Quantity Parameters: Changes to the ratio of positive and negative bags
2. Positive Instance Distribution Parameters: Changes to the shape of the true concept distribution.
3. Data Dimensionality Parameters: Changes to the number of features, number of true concepts, and instances per bag.

Three Performance Measures

- Mean Centroid Error: $\sum_{k=1}^{K} ||c_{k,TRUE} - c_{k,TC}||$

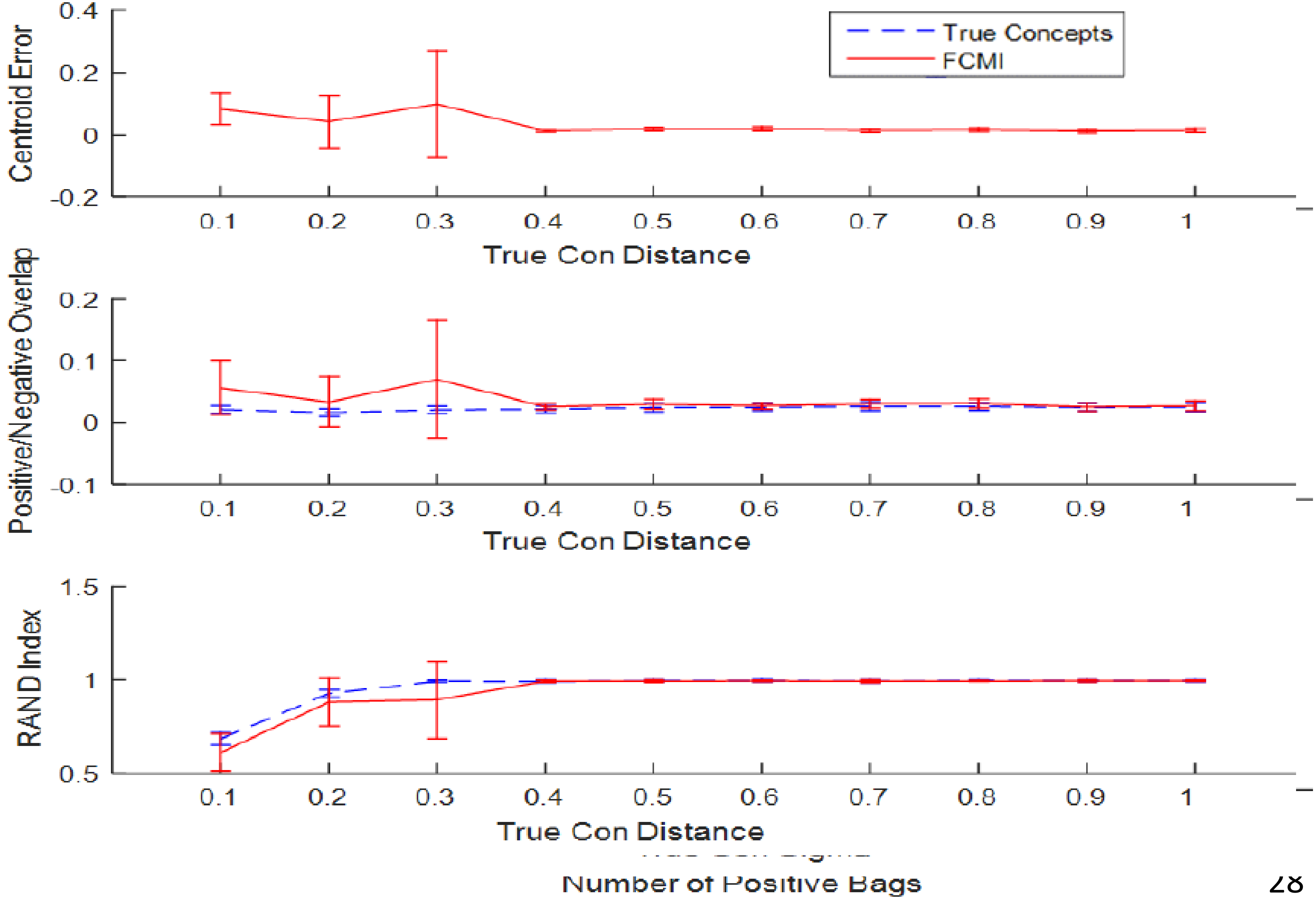
- Positive/Negative Bag Probability Overlap: $MaxBP(B_n, \mathcal{T}) = \max_{k=1}^{K} Pr(t_k|B_n)$

- RAND Index (Assigned vs True Concept Mismatches)



- Use both FCMI-derived TCs and approximate "True" TCs based on generative parameters.

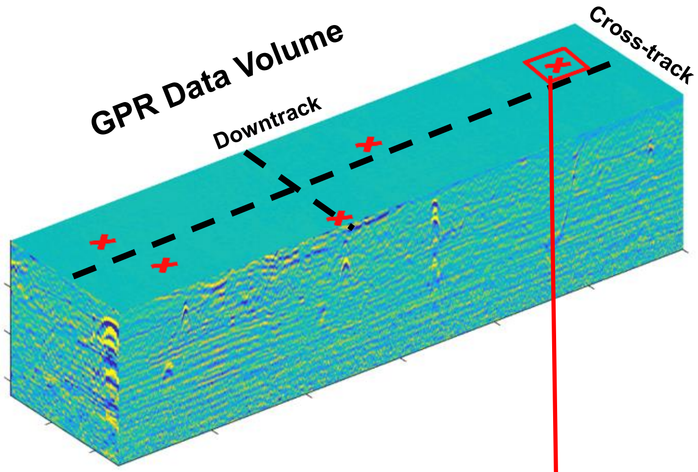
Sensitivity Results

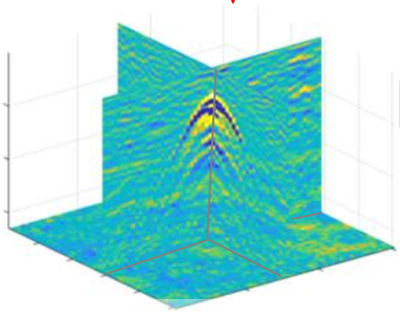


UNIVERSITY OF LOUISVILLE

Application to Buried Explosive Object Detection (BEO)

BEO GPR Collection System





Prescreener alarm data volume

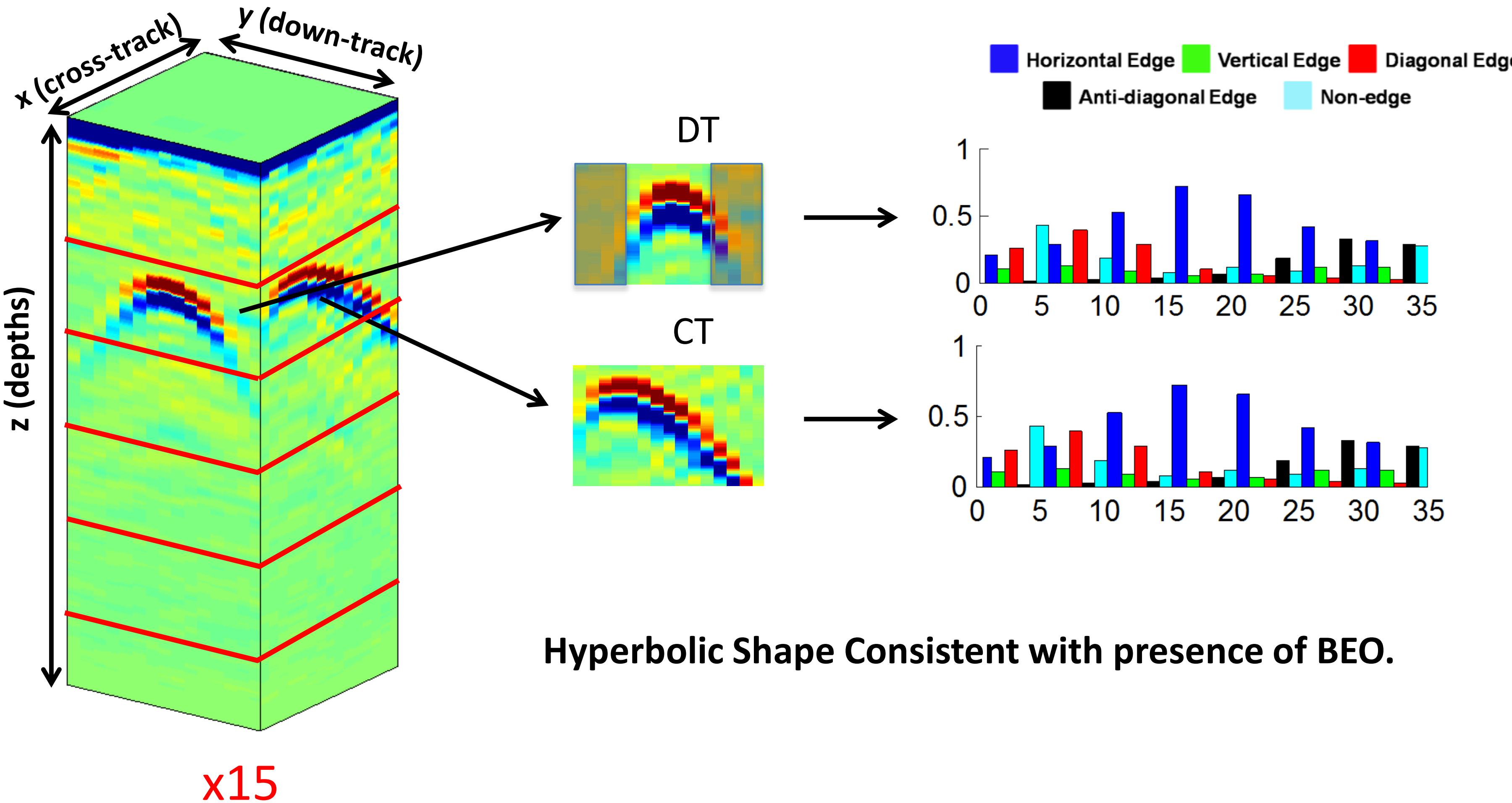
Feature Extraction

Discrimination Algorithm

Threat? (probability of threat)

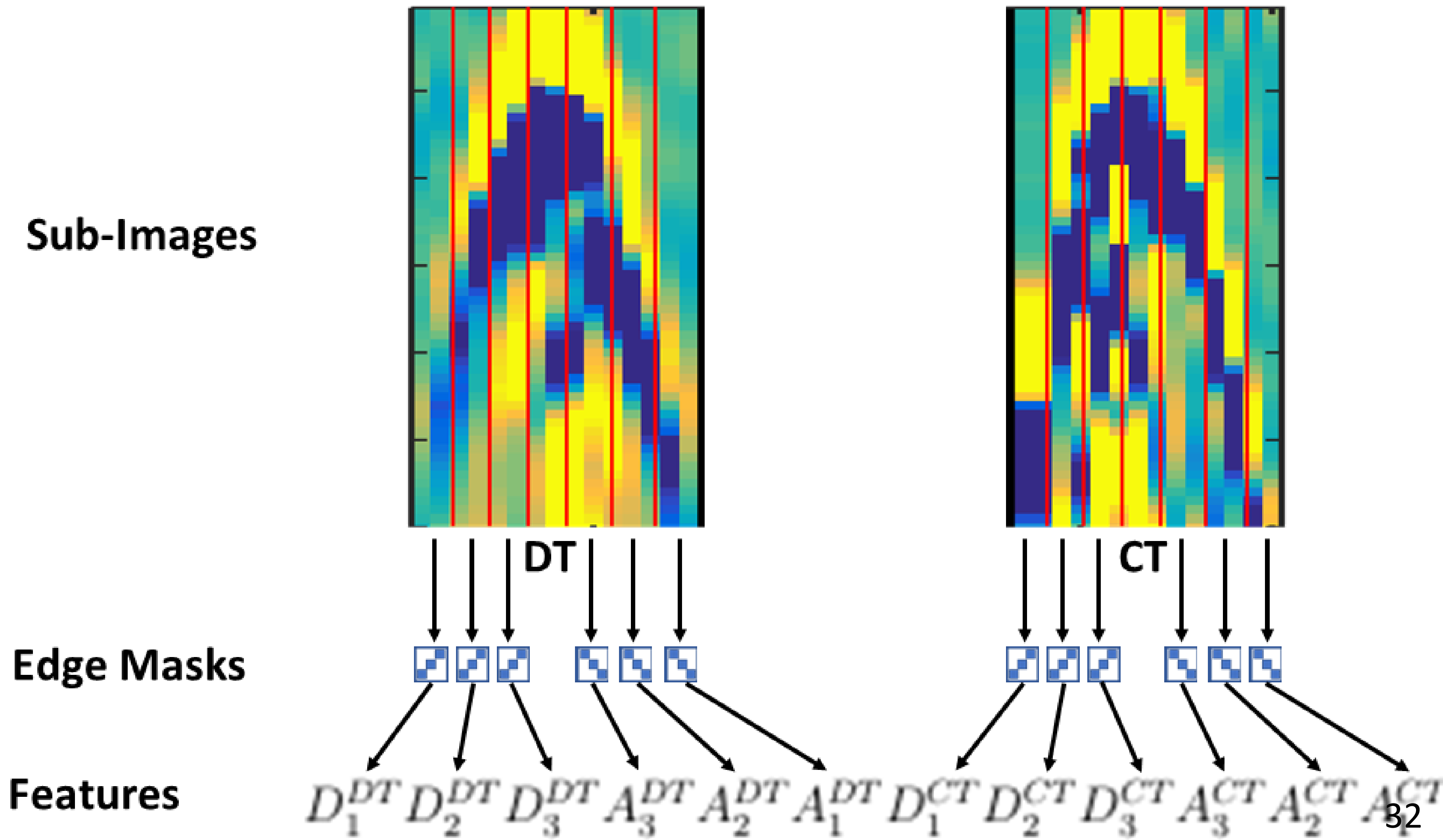
BEO Data: The EHD Feature

Sample BEO Alarm GPR Data



Hyperbolic Shape Consistent with presence of BEO.

Most Discriminative Features



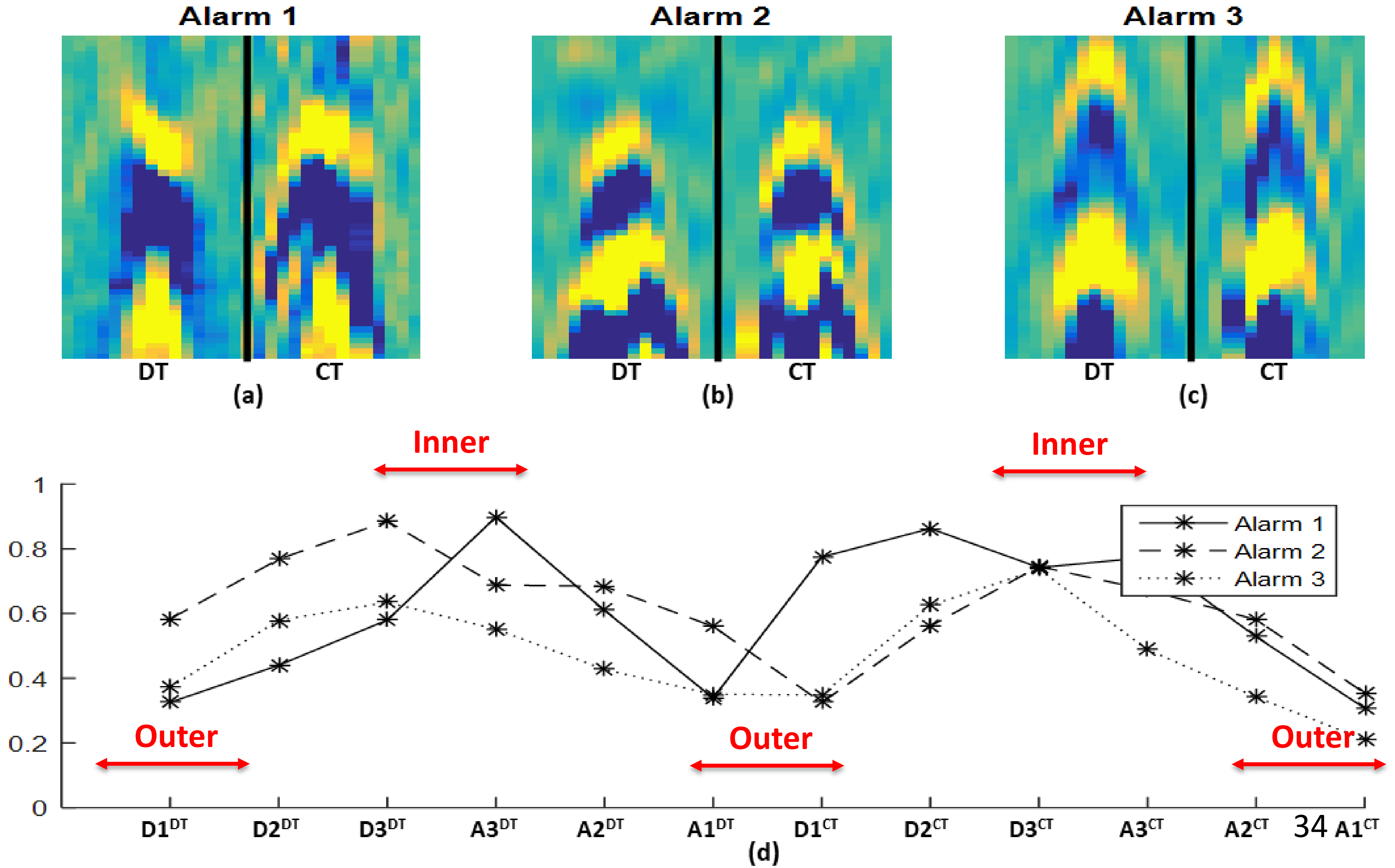
BEO Clustering Experiment

- Data consist of 500 Diverse BEOs and 500 FA.
- FCMI run with only the most discriminative EHD features with 10 TC.

1. 1st goal was to locate clusters with distinctive BEO types and characteristics.
2. 2nd goal was to investigate viability of using Embedded Feature Space-based classification with the TCs.

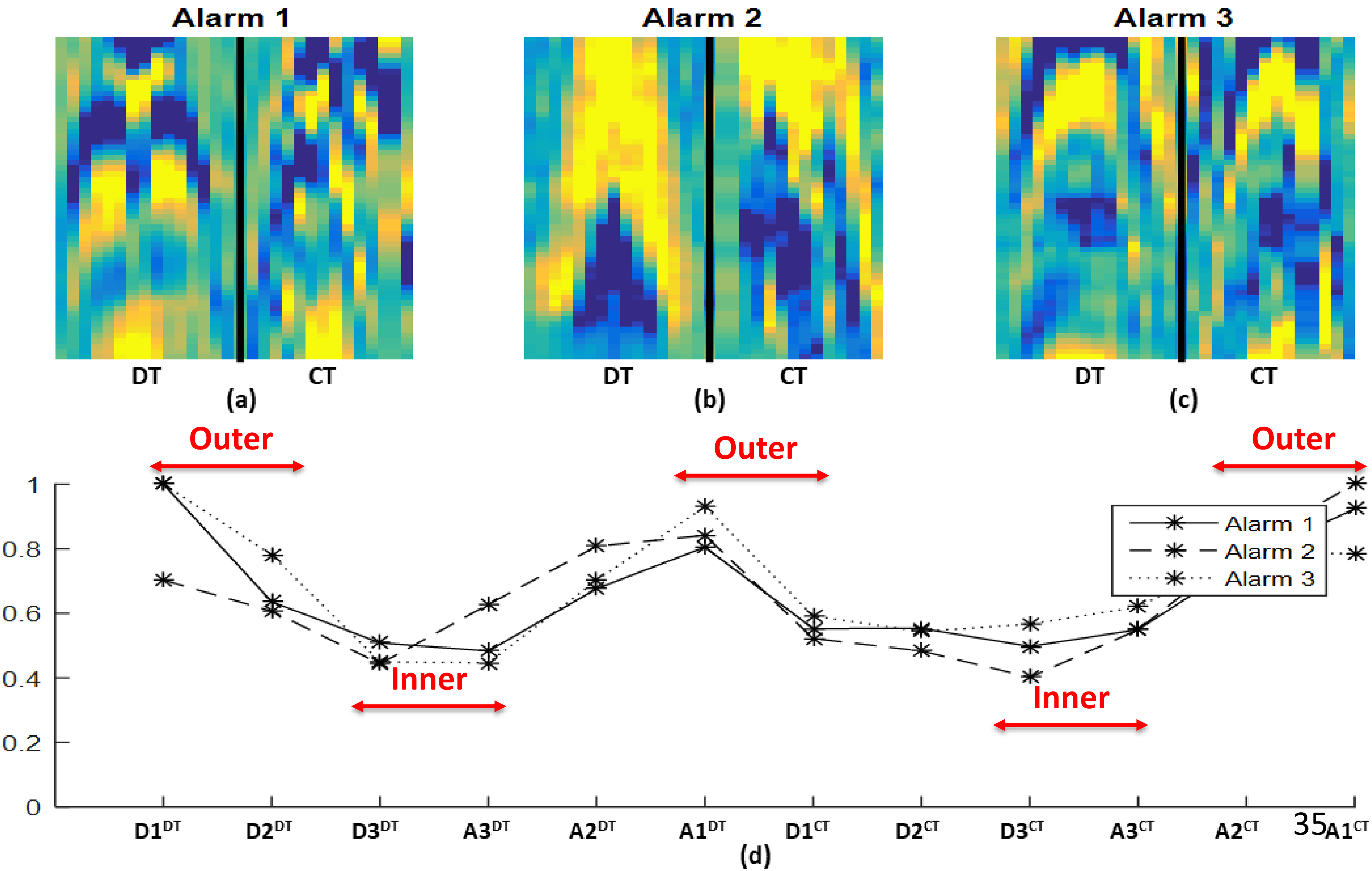
Distinctive BEO Clusters

Target Concept 3



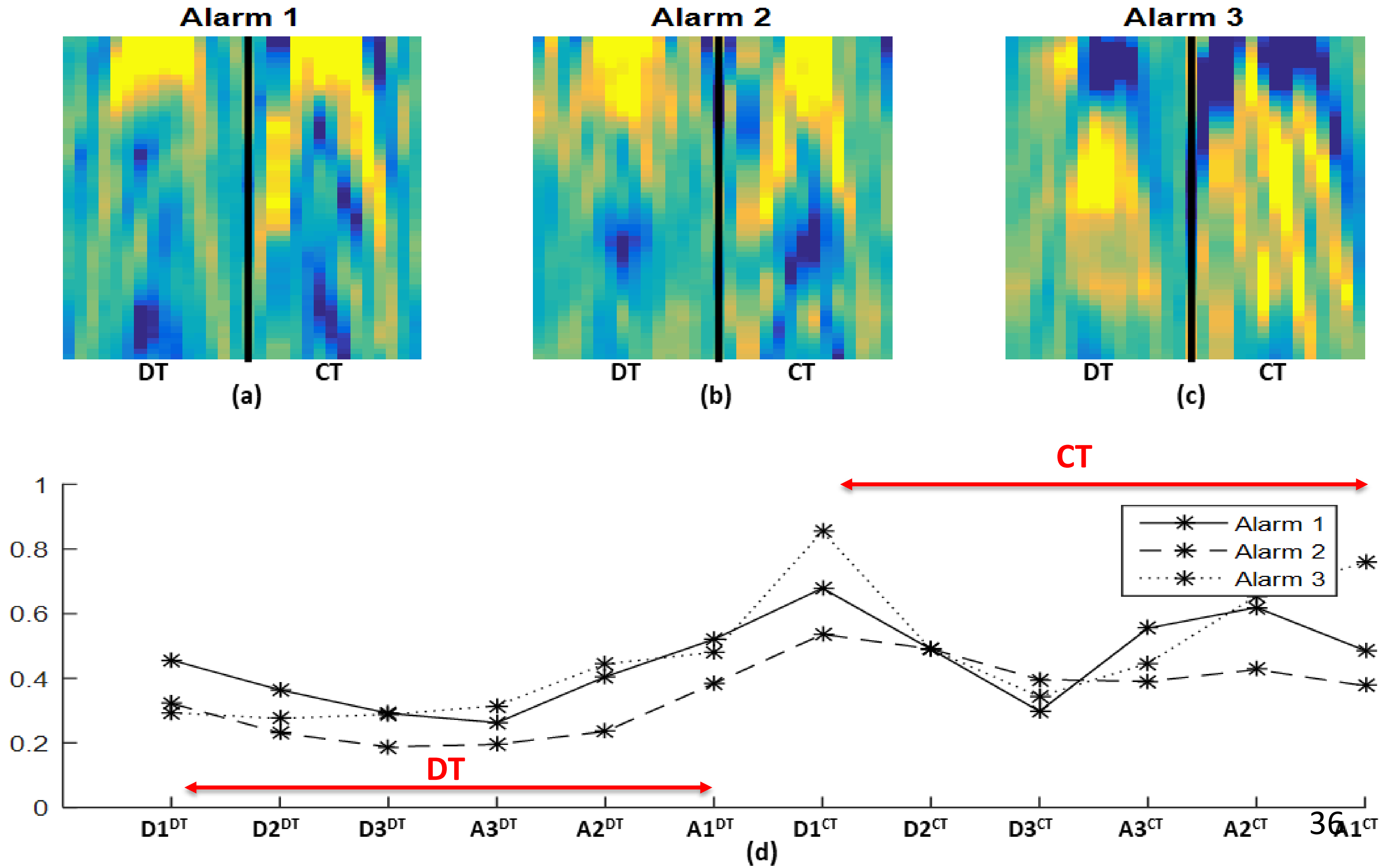
Distinctive BEO Clusters

Target Concept 4

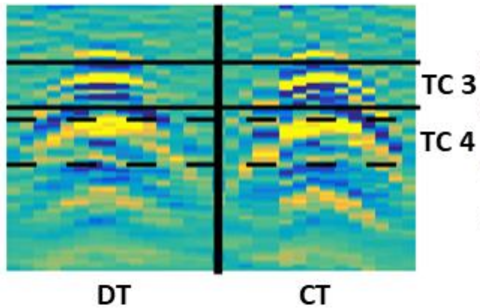


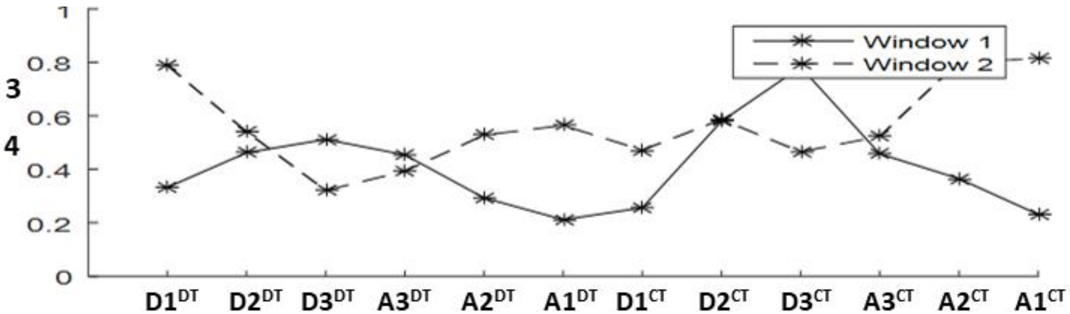
Distinctive BEO Clusters

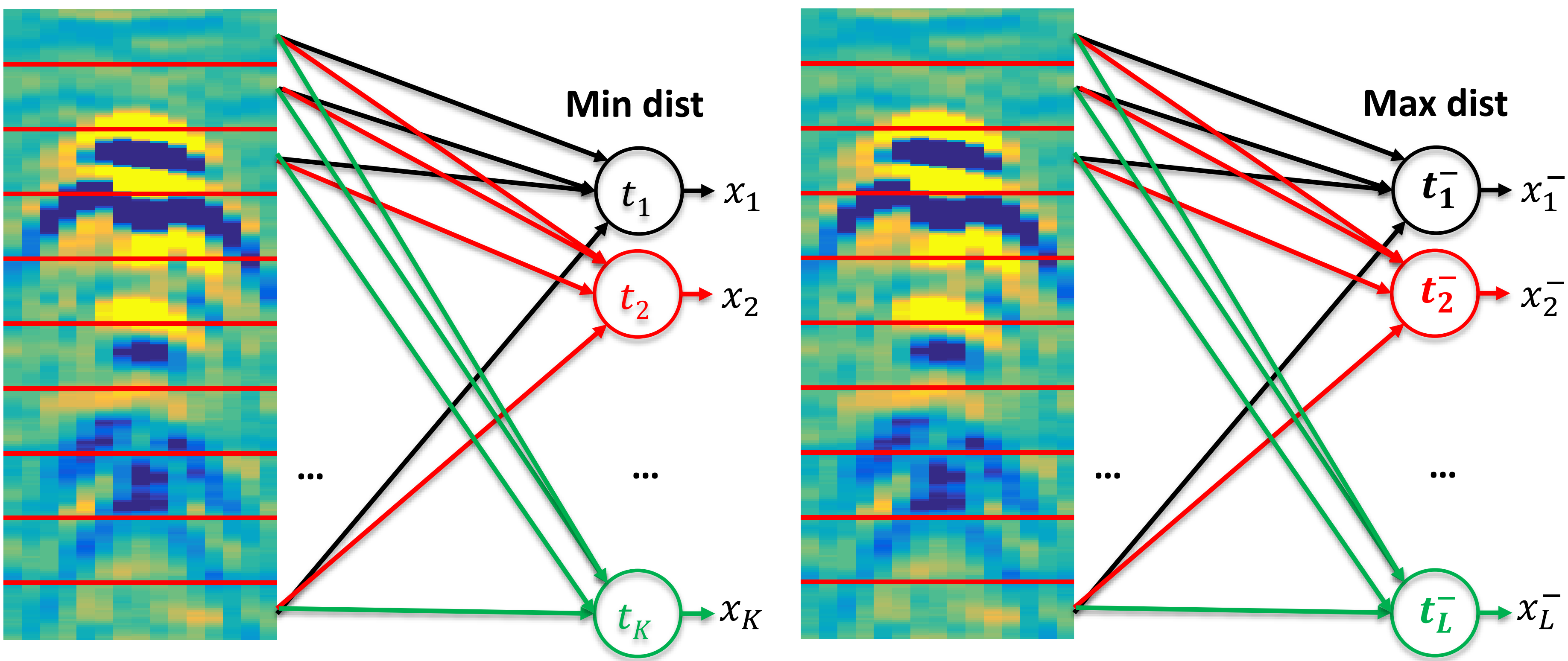
Target Concept 5



Embedding Features with TCs







$$x_1, x_2, \ldots, x_K, x_1^-, x_2^-, \ldots, x_L^-$$

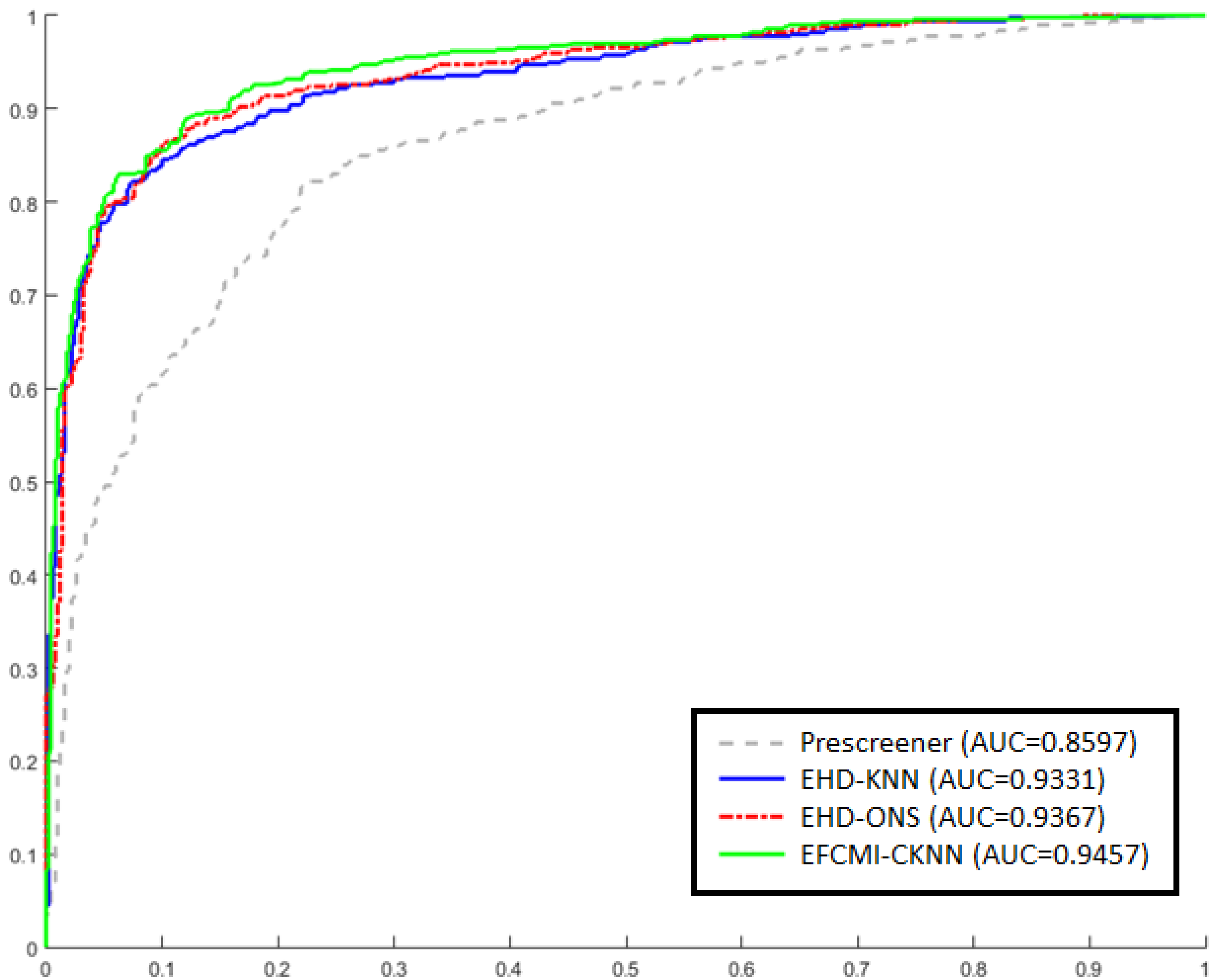
Benchmark Results (AUC)

	MUSK 1	MUSK 2	ELEPHANT	FOX	TIGER
EFCMI-CKNN	.9279	.9078	.9074	.6654	.8281
EFCMI-SVM	.9338	.9341	**.9203**	.6761	.8172
EFCMI-ONS	.9480	.9263	.8808	**.6955**	.7962
DD-SVM[1]	**.9626**	**.9683**	.8877	.6564	.7635
MILES[2]	.9428	.9560	.9018	**.6955**	**.9172**

[1] Chen, Y., & Wang, J. Z. (2004). Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, *5*(Aug), 913-939.

[2] Chen, Y., Bi, J., & Wang, J. Z. (2006). MILES: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(12), 1931-1947.

BEO Results



EHD-KNN:

Frigui, H., & Gader, P. (2009). Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic $ k $-nearest neighbor classifier. *IEEE Transactions on Fuzzy Systems*, *17*(1), 185-199.

Conclusions

- Proposed CMDD, FMDD, and PMDD metrics adapted from principles of Clustering Theory and the DD metric, and proposed the CCMI, FCMI, and PCMI algorithms to optimize CMDD, FMDD, and PMDD metrics.
- Used two synthetic data experiments to compare our algorithms to DD and test robustness to parameter changes.
- Demonstrated FCMI could locate diverse clusters of BEOs using a set of discriminative features.
- Proposed classifier construction based on learned positive and negative TCs, and Embedded Feature Space.
- Applied the EFCMI-based classifiers to Benchmark data and BEO data and demonstrated them to be competitive with existing approaches.

Publications

1. Malof, J. M., Reichman, D., **Karem, A.**, Frigui, H., Ho, D. K., Wilson, J. N., ... & Collins, L. M. (2018). A large-scale multi-institutional evaluation of advanced discrimination algorithms for buried threat detection in ground penetrating radar. *IEEE Transactions on Geoscience and Remote Sensing*
2. **Karem, A.**, Trabelsi, M., Moalla, M., & Frigui, H. (2018, April). Comparison of several single and multiple instance learning methods for detecting buried explosive objects using GPR data. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIII* (Vol. 10628, p. 106280G). International Society for Optics and Photonics.
3. **Karem, A.**, & Frigui, H. (2016, December). Multiple Instance Learning with multiple positive and negative target concepts. In *2016 23rd International Conference on Pattern*
4. **Karem, A.**, Khalifa, A. B., & Frigui, H. (2016, May). A fisher vector representation of GPR data for detecting buried objects. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXI* (Vol. 9823, p. 98231C). International Society for Optics and Photonics.
5. **Karem, A.**, & Frigui, H. (2015, August). Fuzzy clustering of multiple instance data. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-7). IEEE.
6. Patrick Corbett, Yetu Yachim, Andrea Ascuena, and **Andrew Karem** (2013). Understanding nextGen students' information search habits: A usability perspective. In R. McClure & J. Purdy (Eds.), The New Digital Scholar: Exploring and Enriching the Research and Writing Practices of NextGen Students. American Association of Information and Science Technology.
7. **Karem, A.**, & Frigui, H. (2011, July). A multiple instance learning approach for landmine detection using ground penetrating radar. In *2011 IEEE International Geoscience and Remote Sensing Symposium* (pp. 878-881). IEEE.
8. **Karem, A.**, Fadeev, A., Frigui, H., & Gader, P. (2010, April). Comparison of different classification algorithms for landmine detection using GPR. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XV* (Vol. 7664, p. 76642K). International Society for Optics and Photonics.
9. Frigui, H., Fadeev, A., **Karem, A.**, & Gader, P. (2009, May). Adaptive edge histogram descriptor for landmine detection using GPR. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIV* (Vol. 7303, p. 730321). International Society for Optics and Photonics.