Radiological Threat Detection by Canonical Correlation Analysis

Eric Lei, Peter Huggins, Artur Dubrawski
Auton Lab, Carnegie Mellon University, Pittsburgh, PA

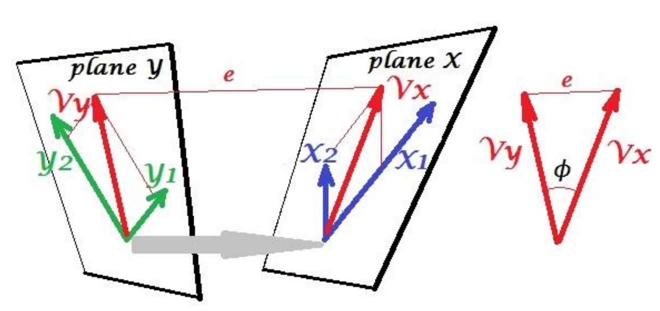
Exploiting Structured Correlations Between the Inside and Outside of an Energy Window

Introduction

- Given radiation data from noisy (urban) background environments, a popular method for threat detection is censored energy window (CEW).
- CEW does not fully exploit the data but instead aggregates in a way may lose information.
- It should be possible to develop a similar method that does not lose information.

Canonical Correlation Analysis

- CCA is a statistical method for finding structured correlations between sets *X* and *Y* of random variables.
- Finds linear combination of *X* and of *Y* such that the correlation between the combinations is maximized.



$$\max_{w_x, w_y} \operatorname{corr}(w_x^\top X, w_y^\top Y)$$

- The process can be repeated min{|*X*|, |*Y*|} times under the constraint that each combination of *X* is uncorrelated with every non-corresponding combination of *Y* and vice-versa.

Approach for Radiological Threat Detection

- Objective: Develop a classifier of photon count samples to detect a source signature.

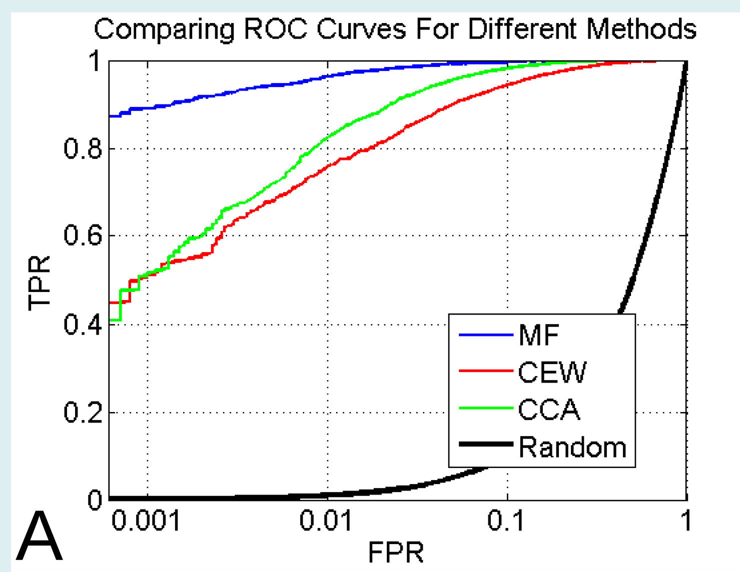
1. Compute energy window maximizing signal-to-noise ratio.
2. Apply CCA to find correlations between energy bins inside and outside the window.
3. Compute differences between corresponding linear combinations given by CCA:
$$w_x^\top X - w_y^\top Y$$
4. Use these differences as features to train a classifier.

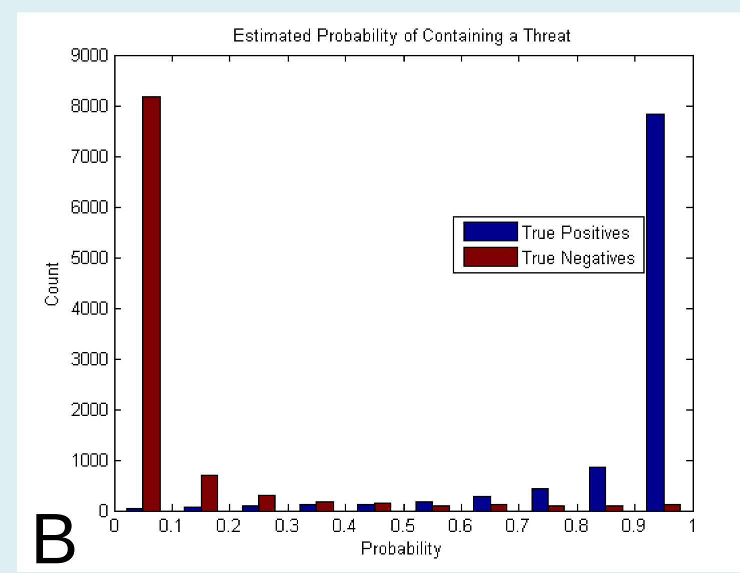
- Intuitively, the classifier learns the degree of correlation to expect with a source versus without.
- Our final method is a hybrid approach: Combine our CCA approach with CEW by adaptively switching between them based on ROC curves.

Experiments

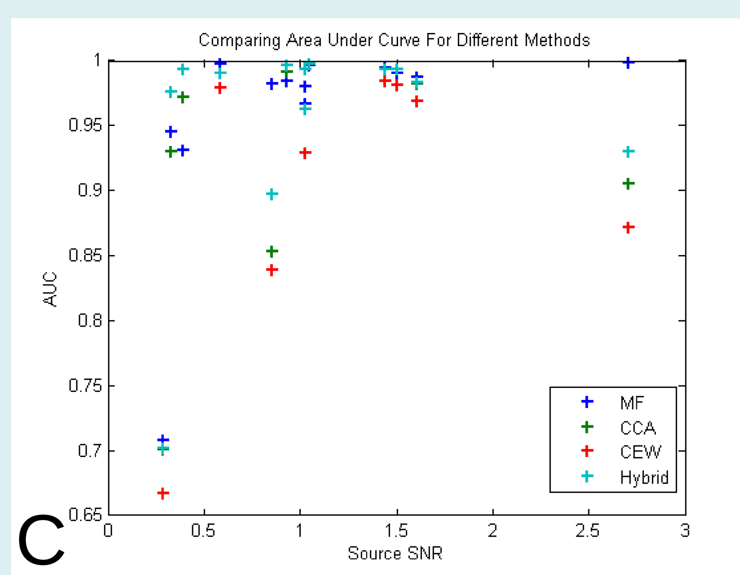
Comparing Methods for Classifying Threats

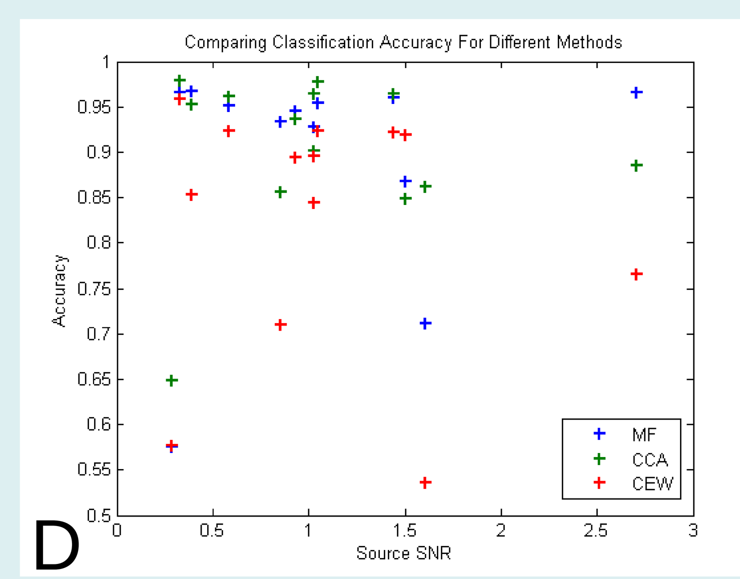
- Four methods: match filter (MF), CEW, random guessing, our CCA method, and a hybrid of CCA and CEW.
  - MF is source-aware—it uses more information than is usually known in practice, so it should perform the best.





- Figure A shows ROC curves on a particular source type.
  - CCA beats CEW except at low FPR.
  - MF beats both.
  - Hybrid CCA/CEW is the convex hull of CCA and CEW and so always beats both or matches the better of them.
- Figure B shows the CCA model's estimated probability of each sample being positive.

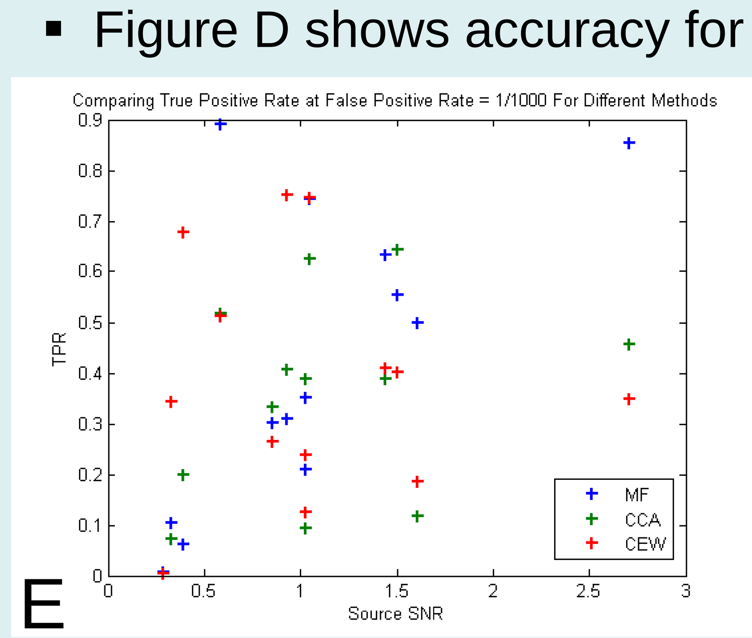


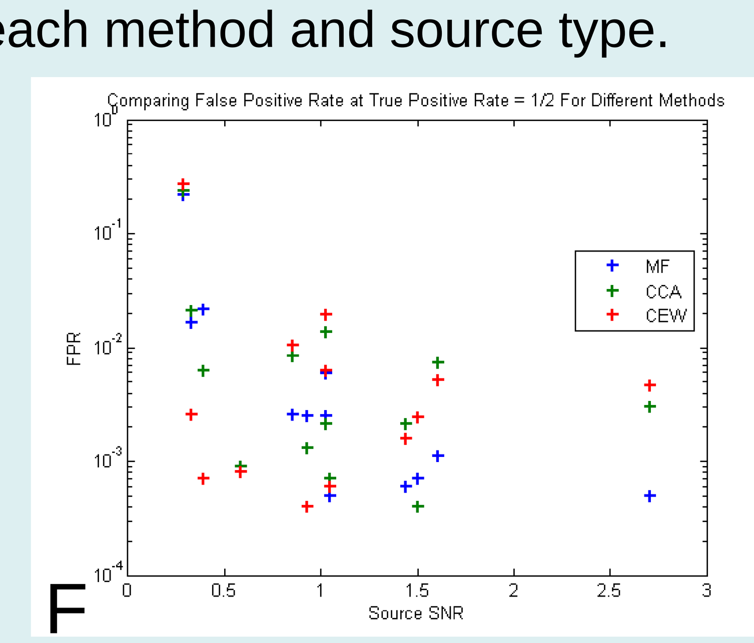


- Figure C shows AUC for each method and source type.
  - Comparing AUC leads to these *p*-values:

Hypothesis	*p*-value
CCA > CEW	9.0%
Hybrid > CEW	0.2%

- Figure D shows accuracy for each method and source type.





- Figure E shows true positive rate at a fixed false positive rate.
- Figure F shows false positive rate at a fixed true positive rate.
- Dataset: a collection of gamma-ray spectra measurements taken in Sacramento, California.
  - Synthetically injected to create positive samples.
  - Threat sources are shielded SNM sources with different configurations of fissile material and shielding.

Discussion

- CCA outperforms CEW on 9 of 13 tested source types.
- CCA sometimes even outperforms MF, which is unexpected.
  - MF uses exact source rates to detect the presence of a source.
- The hybrid method never loses to CEW and wins significantly on a paired test.
- CCA exploits the full distribution of counts in energy bins inside and outside the window.
  - This feature lets CCA leverage more information than CEW, making it more robust in theory.
  - In particular, CCA finds and leverages structured correlations between energy bins inside and outside the energy window.

Finding the Energy Window

- Task: Given a source type, find a set of energy channels with approximately maximum signal-to-noise ratio.

1. Compute S/B for each channel where
   - S is the channel source rate.
   - B is the channel's mean background rate.
2. Sort the channels by S/B in descending order.
3. For each possible integer k, add together the S and B for the top k channels, yielding S' and B' respectively.
4. Output the k channels maximizing $S'/\sqrt{B'}$.

Conclusion

- CCA is a method for finding correlations between two sets.
- CCA can be used to be competitive threat detection algorithms.
  - Sometimes outperforms even MF, which is unexpectedly positive.
- Exploits distribution of counts in energy window by applying CCA to the inside and outside of an energy window.
  - CEW simply aggregates the counts.
- In some cases, it loses to CEW, likely due to overfitting.
  - Remedied by hybrid CCA-CEW
- Needs more extensive tests on different source types and datasets.

References

Härdle, Wolfgang; Simar, Léopold (2007). "Canonical Correlation Analysis". *Applied Multivariate Statistical Analysis*. pp. 321–330

Acknowledgements: This work has been partially supported by DNDO under competitively awarded grant 2010-DN-077-ARI040-02, by DTRA under award HDTRA1-13-1-0026, and by the NSF under award 1320347. This support does not constitute an express or implied endorsement on the part of the Government.