Identifying Rare Class in Absence of True Labels

Application to Monitoring Forest Fires from Satellite data

Varun Mithal

(University of Minnesota- Twin Cities)

NSF Expeditions Workshop August 4, 2015

Global Mapping of Forest Fires

Mapping fires is important for…

- Climate change studies

  e.g., linking the impact of a changing climate on the frequency of fires

- Carbon cycle studies

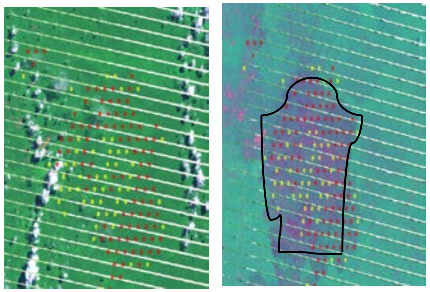
  e.g., quantifying how much $C0_2$ is emitted by fires (critical for UN-REDD)

- Land cover management

  e.g., identifying active deforestation fronts

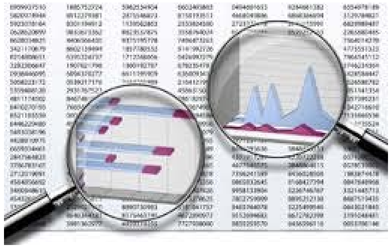
Aerial/Ground Surveys

- Accurate
- Expensive
- Globally infeasible



Manual inspection

- Human effort
- Difficult due to rare class
- Globally infeasible



Computational Techniques

- Automated
- Cost-effective
- Globally scalable
- Historical as well as near-real time

Predictive Modeling Approach

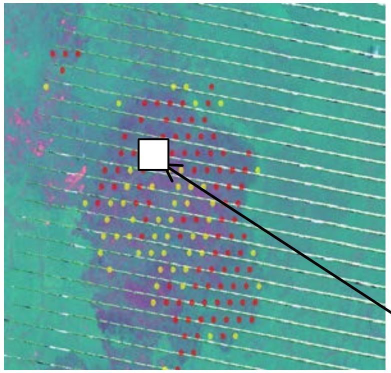
Given a feature vector $\boldsymbol{x} \in \mathbf{R}^d$
predict the class label $y \in \{0, 1\}$

Instance $\boldsymbol{x}_i \in \mathbf{R}^d$	**Label** $y_i \in \mathcal{Y} = \{0, 1\}$
$\boldsymbol{x}_1$	1
$\boldsymbol{x}_2$	0
$\boldsymbol{x}_3$	0
$\boldsymbol{x}_4$	1
.	.
$\boldsymbol{x}_N$	1

Forest Fire Mapping

Multispectral reflectance data

- 7 spectral bands
- 500 m spatial resolution
- 8-day composites



Forest fire mapping

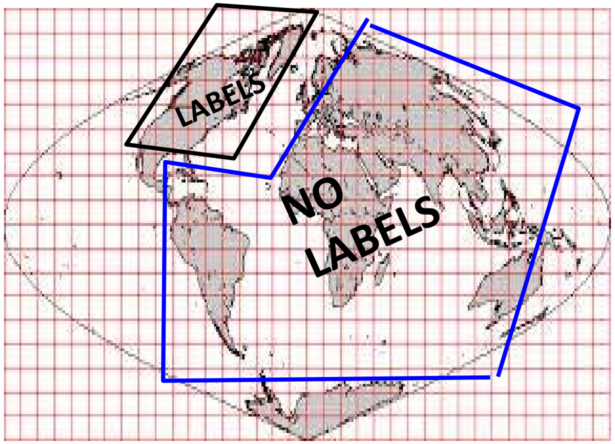
Predicts whether a given pixel is burned or not?

Challenges: Heterogeneity

Variations in the relationship between the explanatory and target variable

- Geographical heterogeneity
- Seasonal heterogeneity
- Land class heterogeneity

Train	Test	Precision	Recall	F-value
California	California	94	65	72
Georgia	California	53	53	53
Georgia	Georgia	87	53	66
California	Georgia	10	30	16



Global availability of labeled samples for burned area classification

Temporal heterogeneity:
Impossible to obtain training samples going back in time

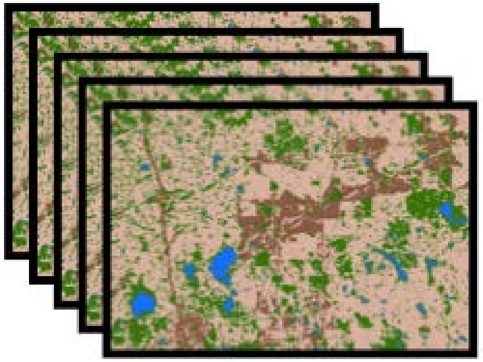
Challenges: Ultra skewed class distribution

Burned areas (California) in year 2008

Positives : 10^3 sq. km.

Negatives: 10^6 sq. km.

Prediction at every time step: $46 * 10^6$



- ➢ Requires extremely low FPR
- ➢ Overall accuracy is not very useful
- ➢ Need to jointly maximize precision and recall
  - Harmonic mean (F-measure)
  - Geometric mean

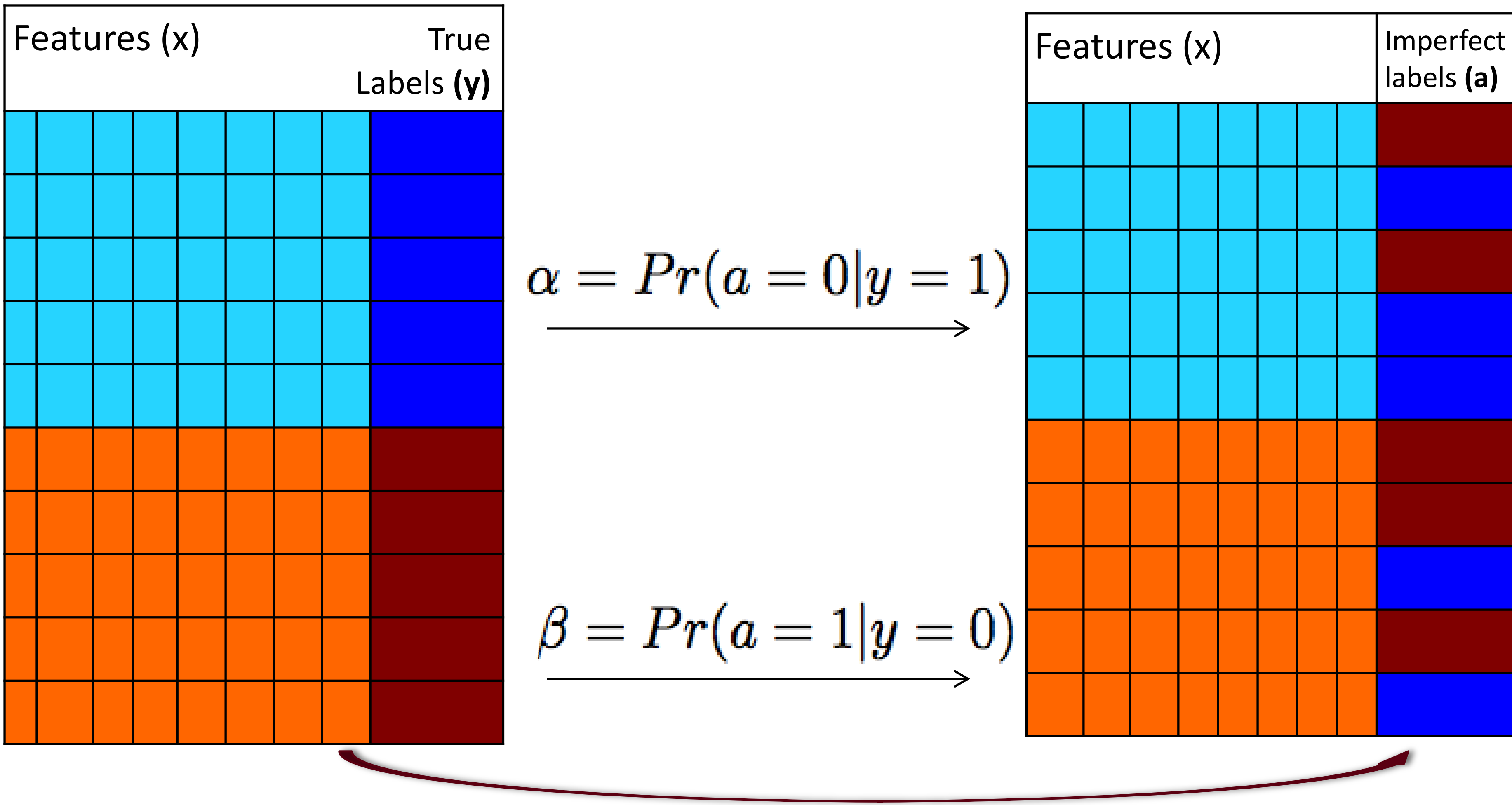
RAPT: RAre class Prediction in absence of ground Truth

- ***Step 1***: Learn classification models using imperfect (noisy) labels
- ***Step 2***: Combine predictions from classification model and the imperfect label
- ***Step 3***: Exploit guilt-by-association using spatial context

Learning with imperfect labels

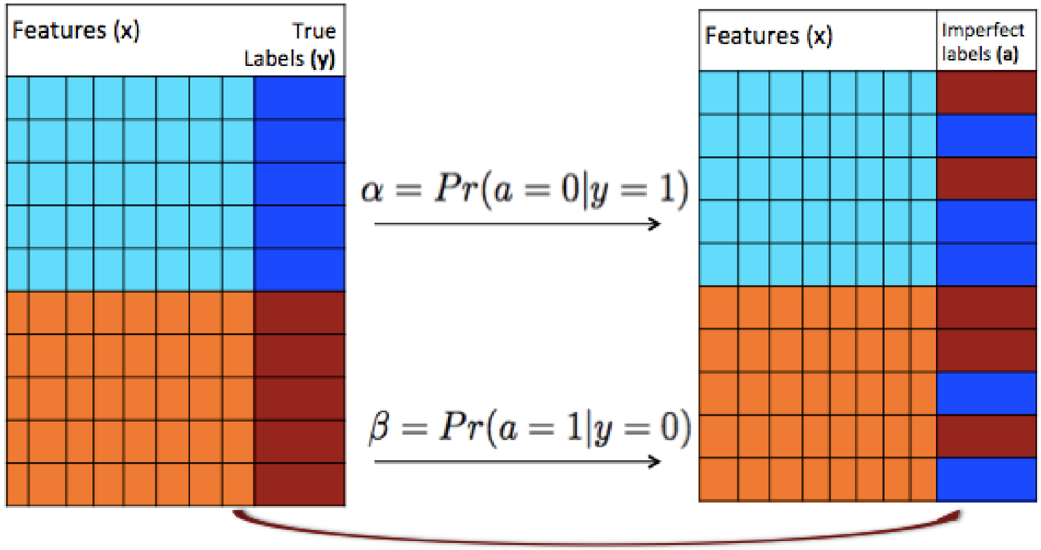
- **Supervised Learning**
  - **Expert-annotated Labels**
    - **Sufficient training samples**
      - *SVM*
      - *Decision tree*
      - *Logistic regression*
    - **Inadequate training samples**
      - *Semi-supervised*
      - *Active Learning*
      - *Multi-view*
      - *Multi-task*
  - **Imperfect Labels**
    - **Multiple annotators**
      - *Learning with crowds*
      - *Raykar et al.*
    - **Single annotator**
      - **Partial Supervision**
        - *Positive Unlabeled learning*
        - *Bing Liu et al.*
        - *Elkan et al.*
      - **Imperfect Supervision**
        - **Balanced**
          - *Natrajan et al.*
        - **Rare class**

Step 1: Train a classifier using imperfect labels



Use a set of features to derive imperfect labels **a**

Step 1: Train a classifier using imperfect labels

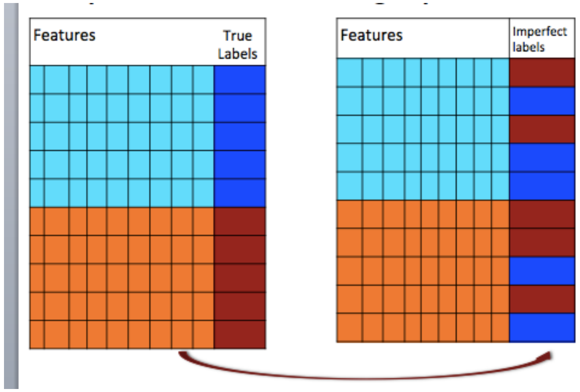


Assumptions

(1) $\alpha + \beta < 1$

(2) Imperfect label is conditionally independent of feature space given the true label

Learning with imperfect labels

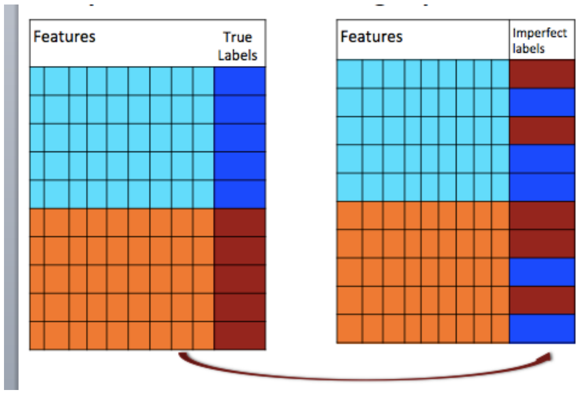


Assumptions

(1) $\alpha + \beta < 1$

(2) Imperfect label is conditionally independent of feature space given the true label

Learning with imperfect labels

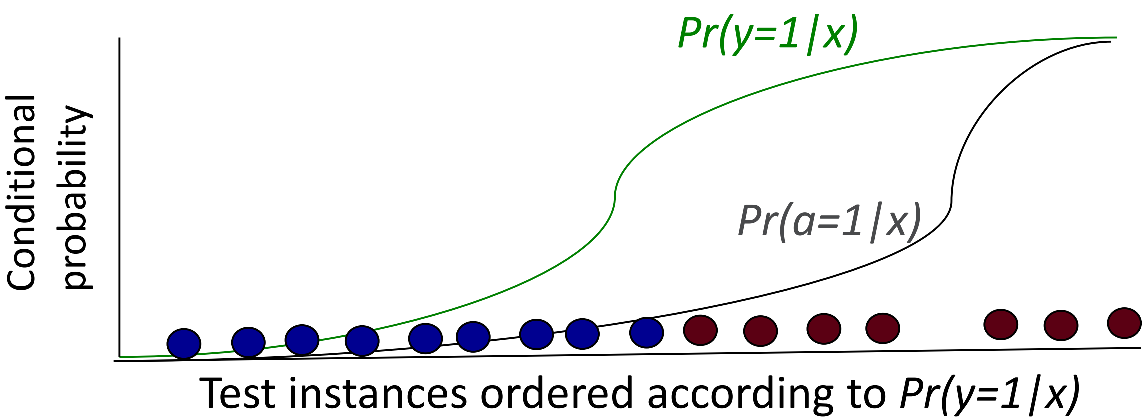


Assumptions

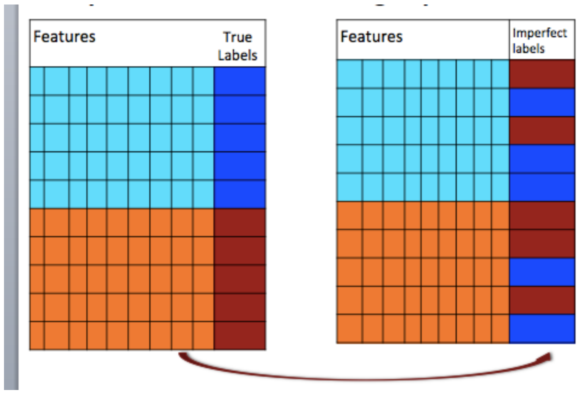
(1) $\alpha + \beta < 1$

(2) Imperfect label is conditionally independent of feature space given the true label

Ranking according to *Pr(a=1|x)* and *Pr(y=1|x)* is identical



Learning with imperfect labels

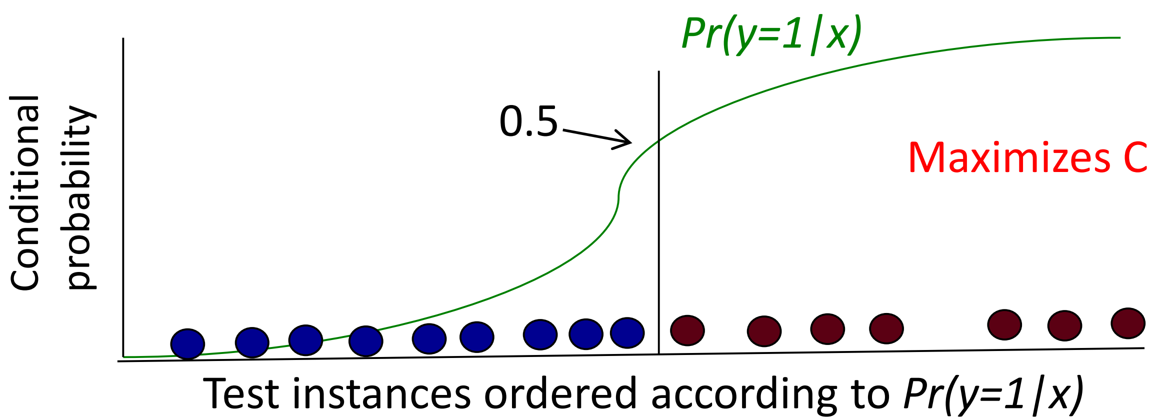


Assumptions

(1) $\alpha + \beta < 1$

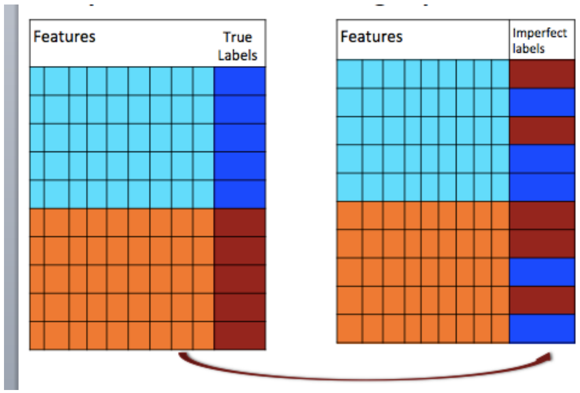
(2) Imperfect label is conditionally independent of feature space given the true label

Ranking according to *Pr(a=1|x)* and *Pr(y=1|x)* is identical



Maximizes Classification Accuracy

Learning with imperfect labels

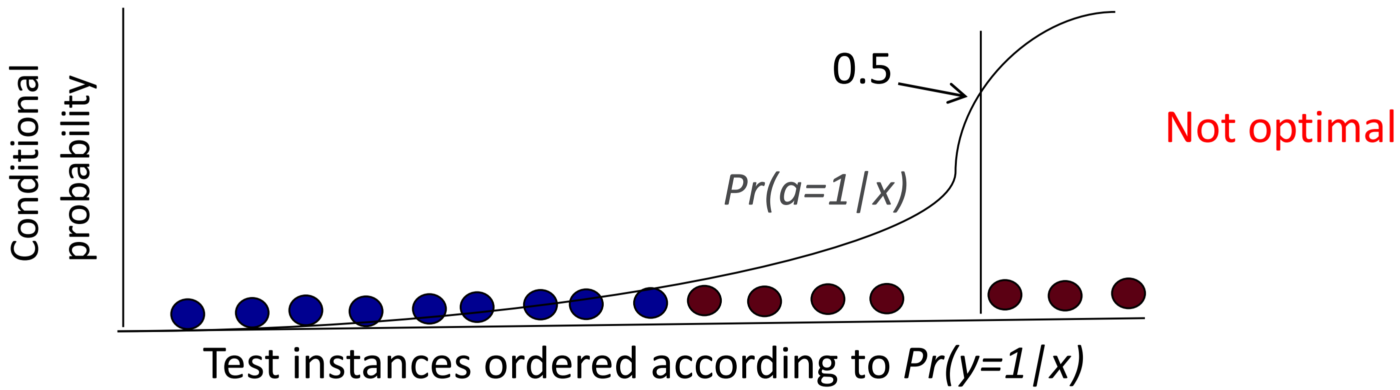


Assumptions

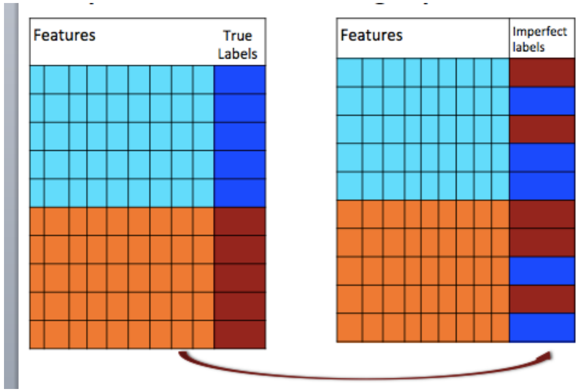
(1) $\alpha + \beta < 1$

(2) Imperfect label is conditionally independent of feature space given the true label

Ranking according to *Pr(a=1|x)* and *Pr(y=1|x)* is identical



Learning with imperfect labels

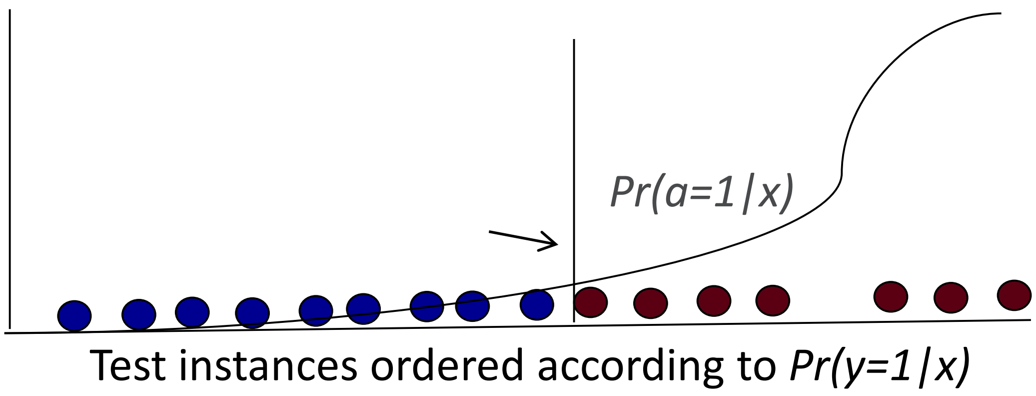


Assumptions

(1) $\alpha + \beta < 1$

(2) Imperfect label is conditionally independent of feature space given the true label

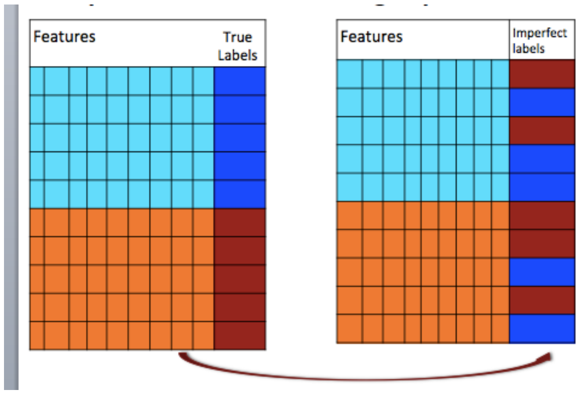
Ranking according to *Pr(a=1|x)* and *Pr(y=1|x)* is identical



Approach
Use labeled validation data set to select threshold.

Labeled data not available

Learning with imperfect labels

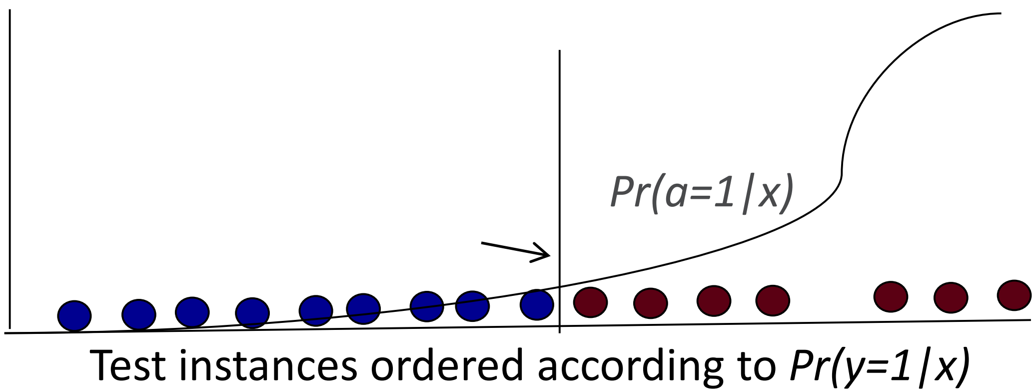


Assumptions

(1) $\alpha + \beta < 1$

(2) Imperfect label is conditionally independent of feature space given the true label

Ranking according to *Pr(a=1|x)* and *Pr(y=1|x)* is identical

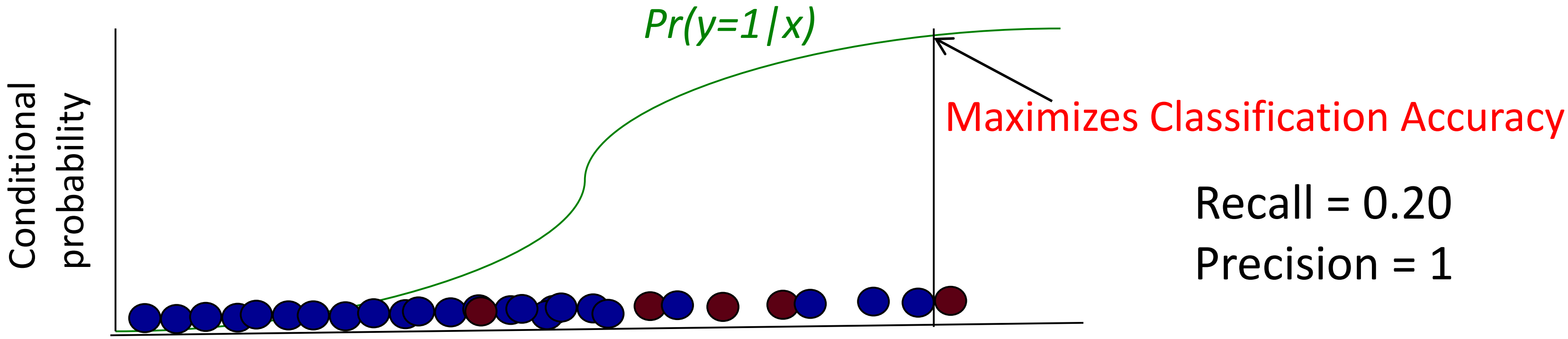


Approach
Select the threshold that maximizes classification accuracy by treating imperfect labels as target.

Our Contribution
We prove that for balanced datasets this approach is optimal.

*Identical prediction is possible using appropriate threshold on *Pr(a=1|x)*, for every threshold on *Pr(y=1|x).* Natarajan 2013

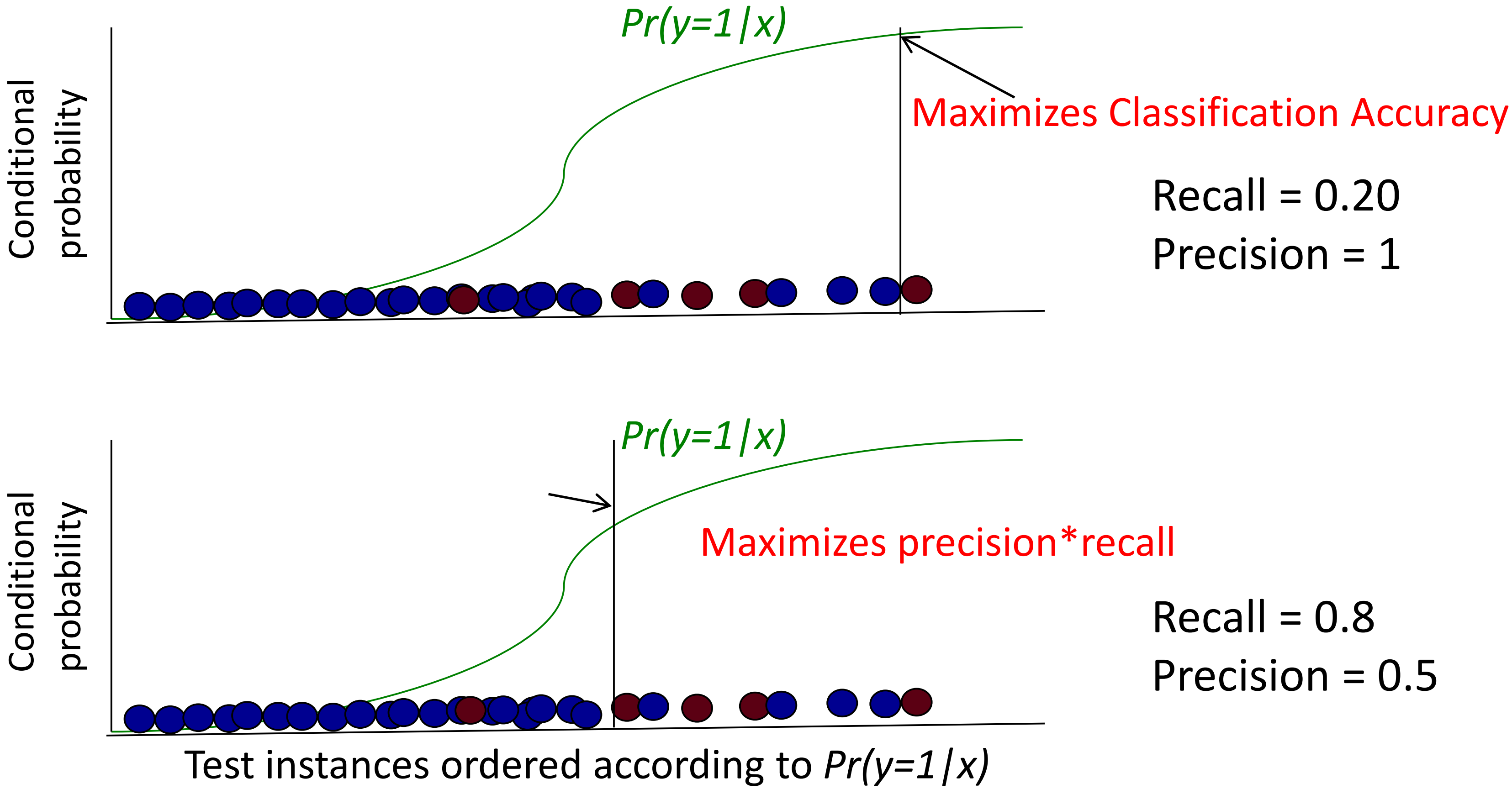
Rare class



Recall = 0.20

Precision = 1

Rare class



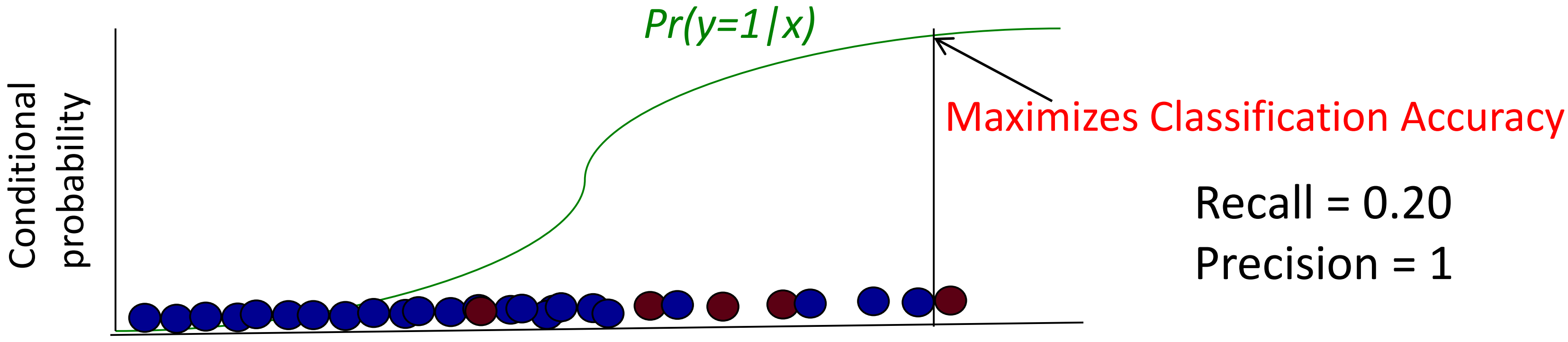
Recall = 0.20

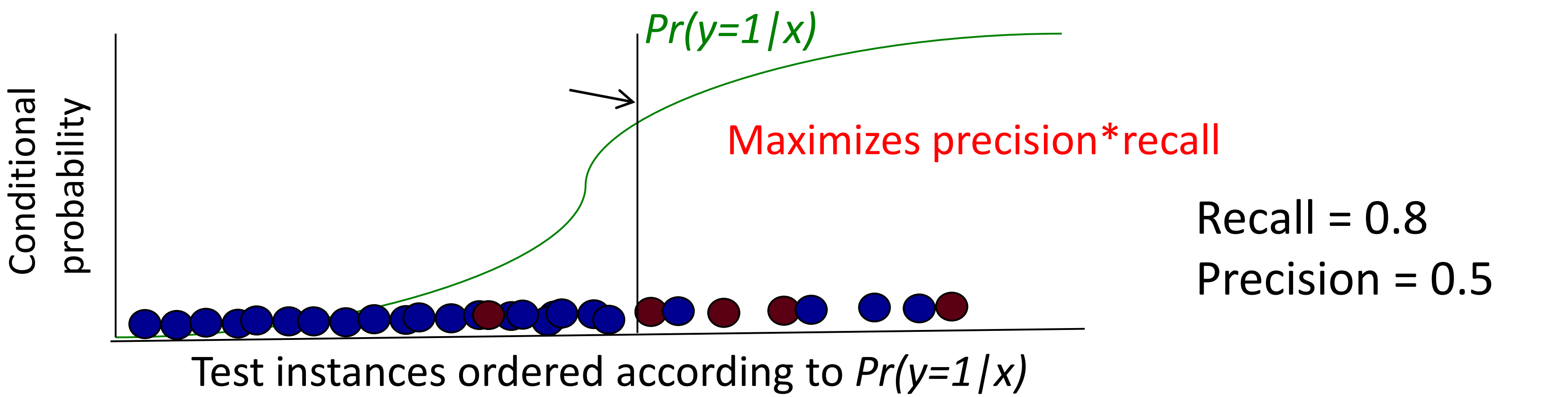
Precision = 1

Recall = 0.8

Precision = 0.5

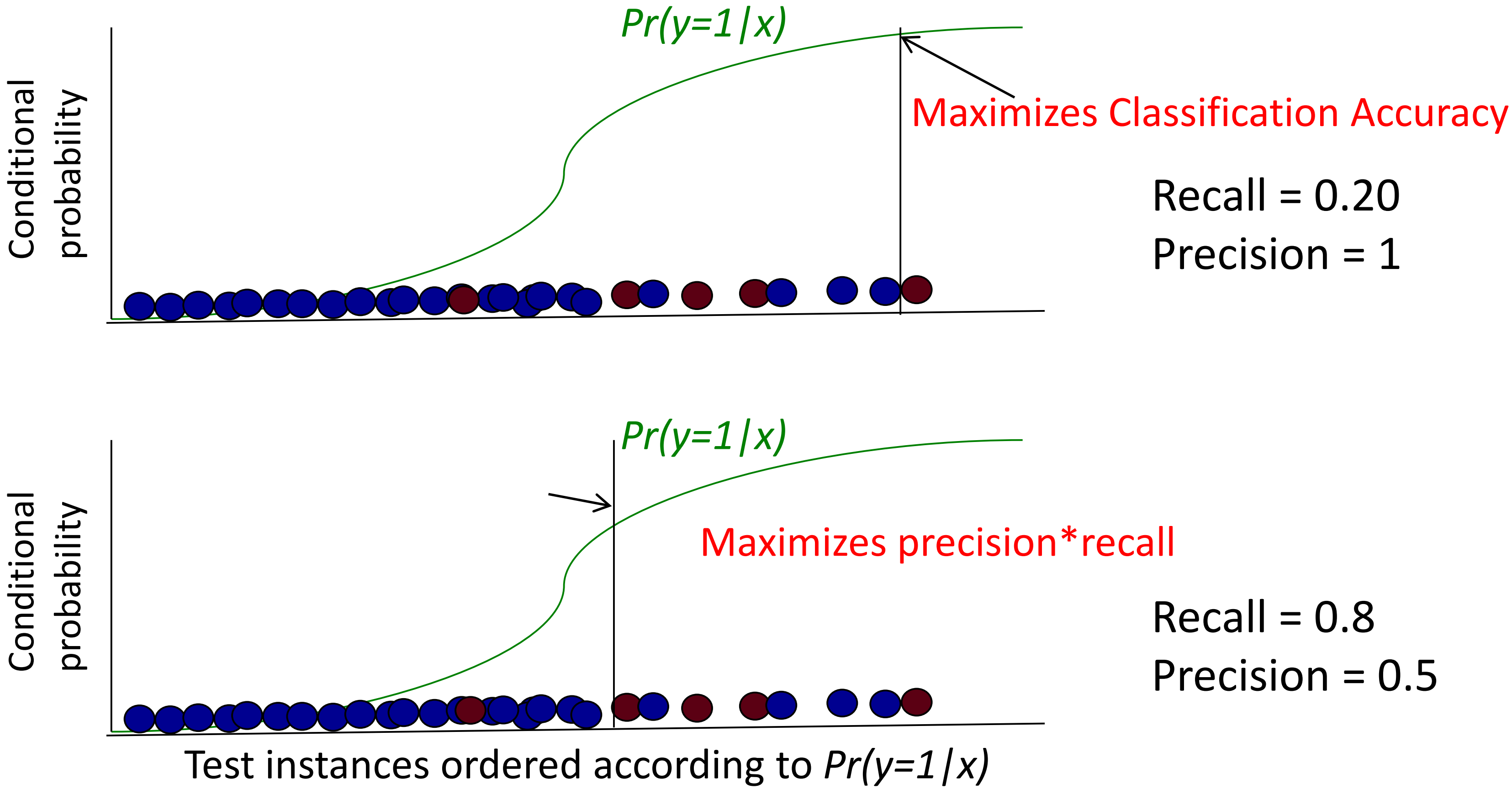
Rare class





Challenge: Accurately estimate precision and recall with imperfect labels

Rare class



Challenge: Accurately estimate precision and recall with imperfect labels

Our Contributions:

(1) A new method to estimate precision*recall using imperfect labels.

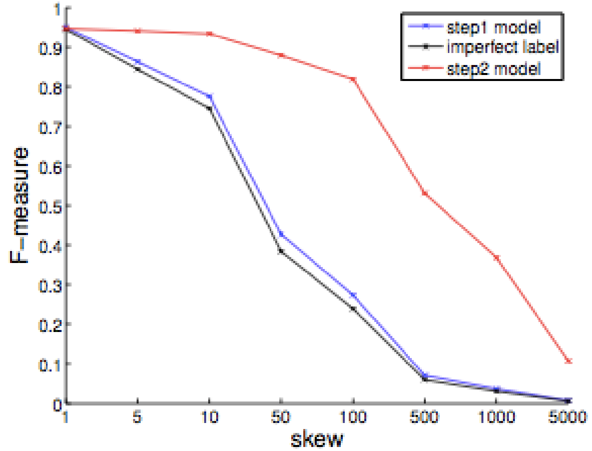
(2) We prove that the selected threshold maximizes the true precision*recall

Step 1: Train a classifier using weak labels

Step 2: Combine predictions of classifier with imperfect labels

- Instance is labeled positive only if it is flagged positive by both
- Considerably reduces the number of false positives
- Incorrectly prunes away some positives

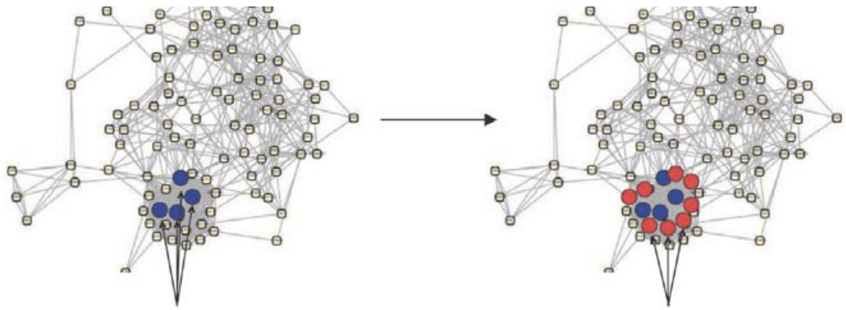
For rare class scenarios, the combination step drastically increases precision with relatively smaller loss of recall.



Step 1: Train a classifier using weak label

Step 2: Combine predictions

Step 3: Guilt-by-association



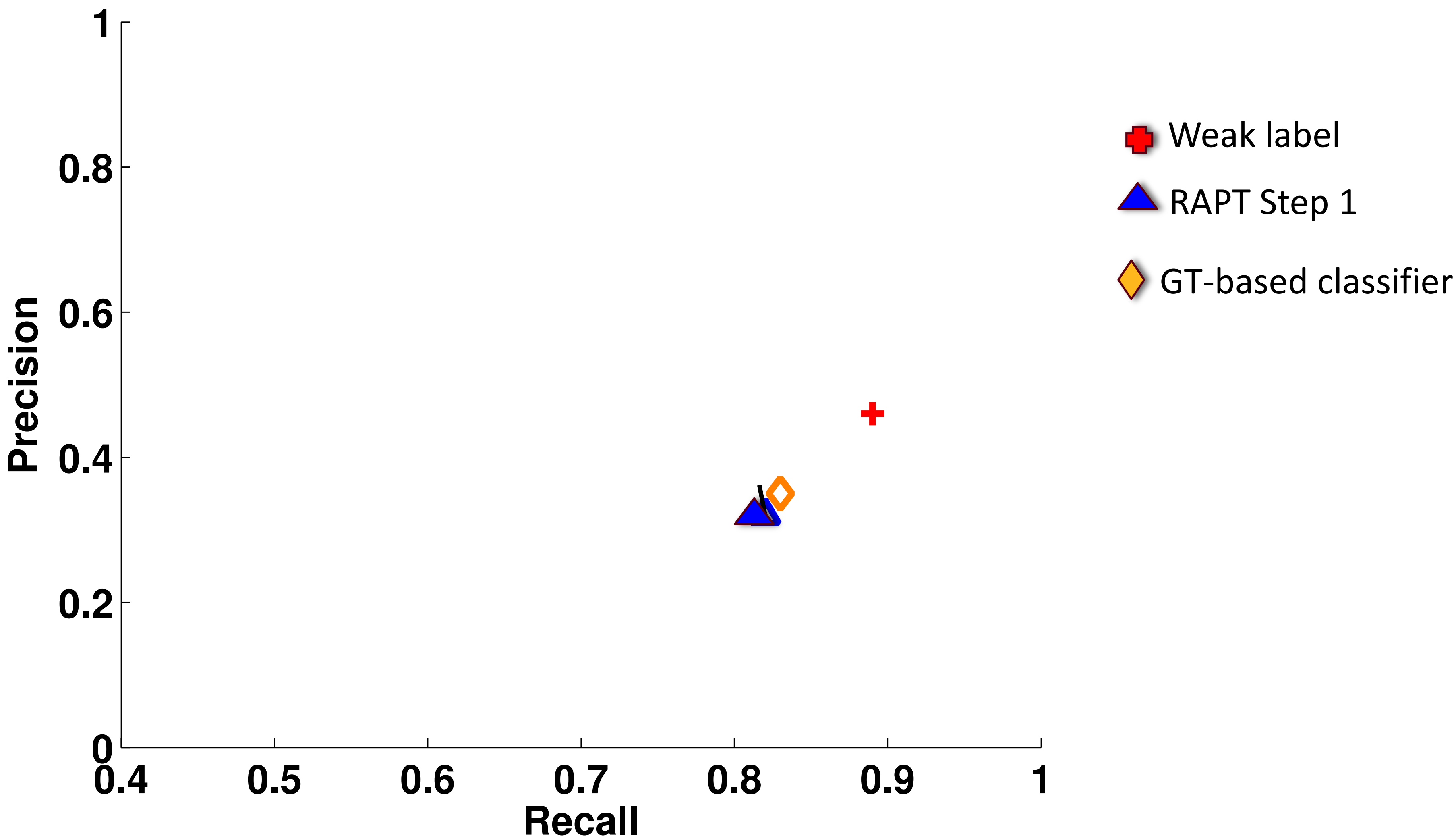
Observations:

- Combination step prunes away some positives
- Missed positives in the neighborhood of confident positives

Approach:

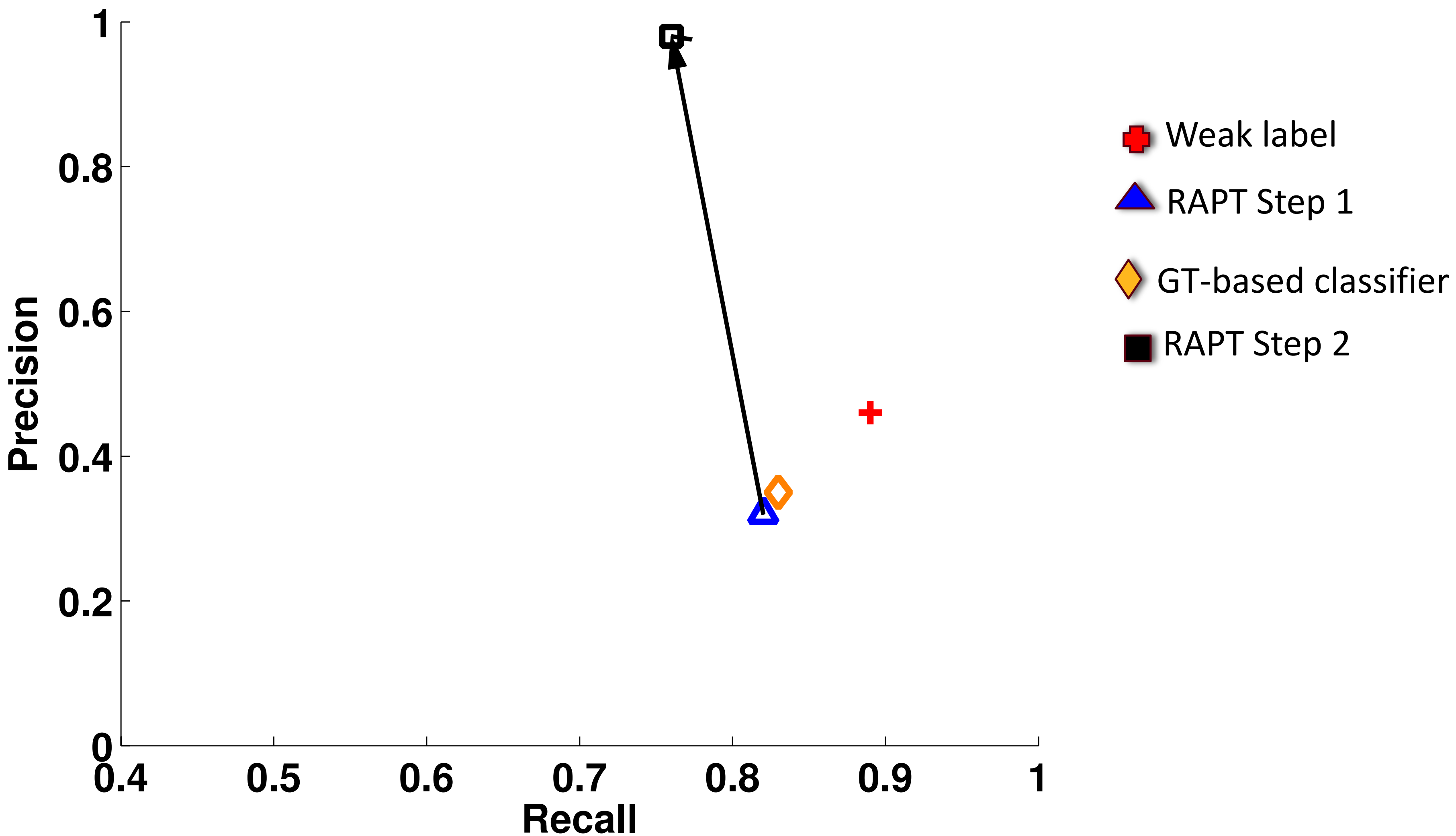
- A collective classification method to make use of labels of neighbors during final classification of each node

Results for Burned Area Mapping



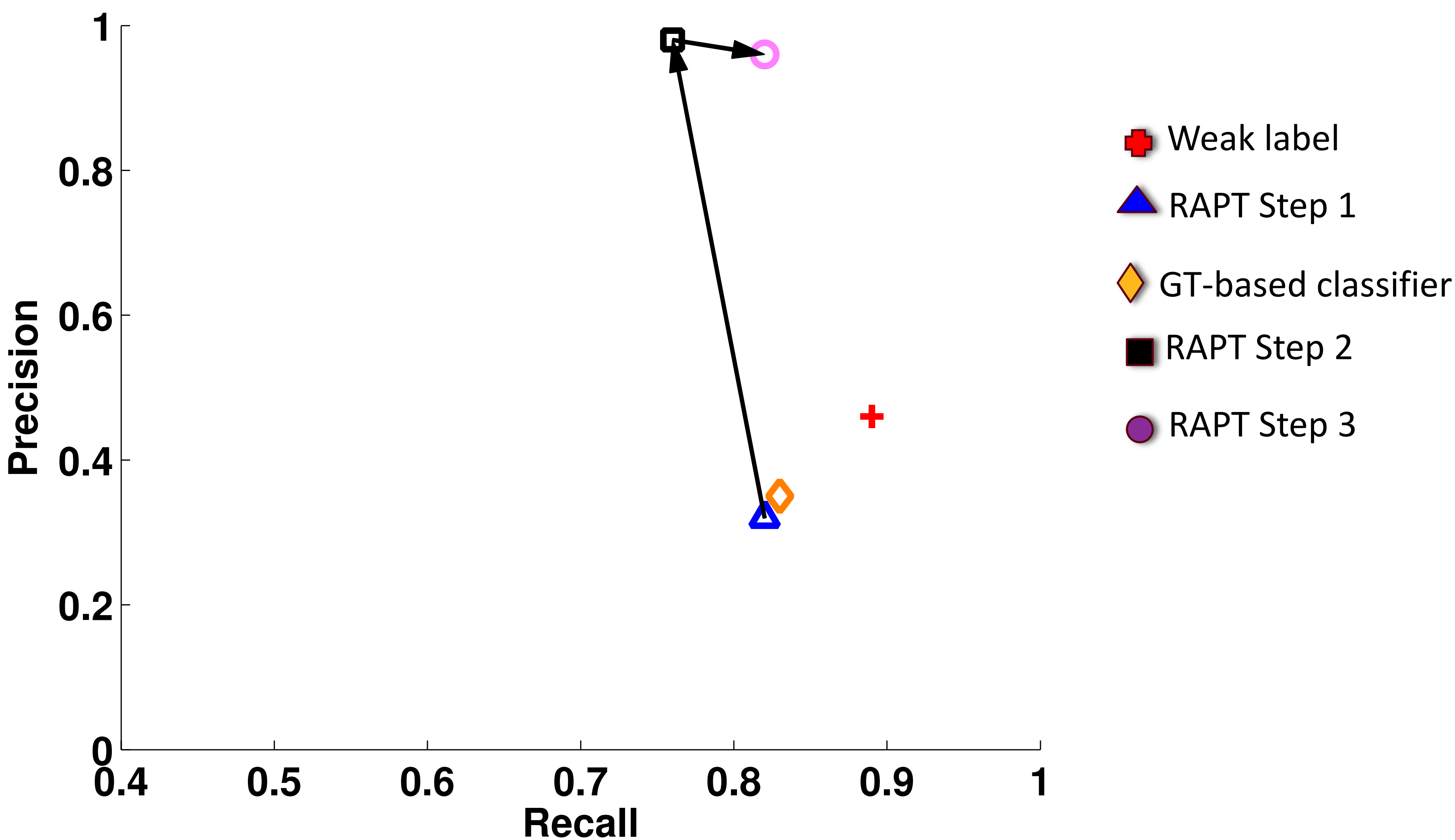
California State

Results for Burned Area Mapping



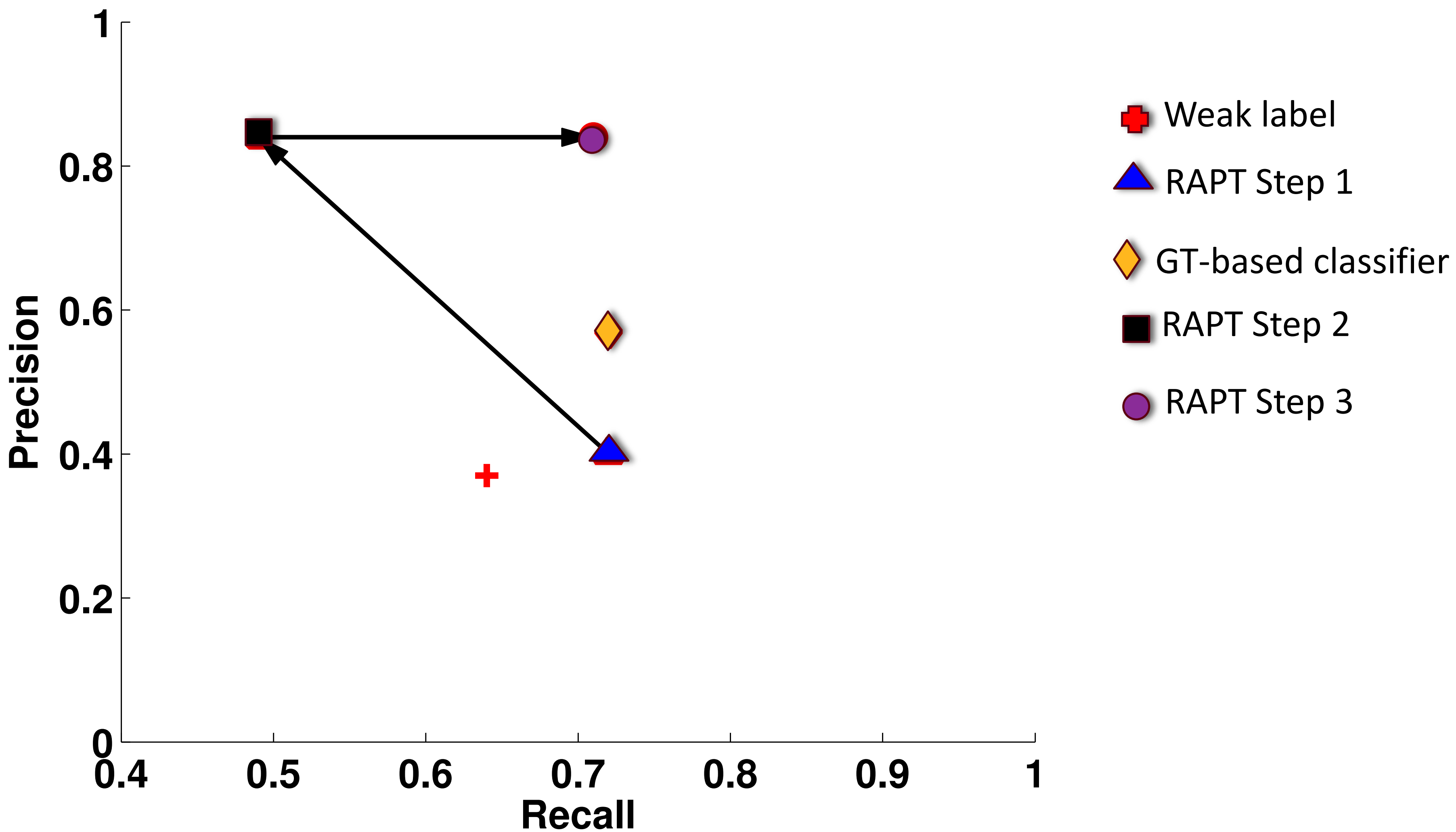
California State

Results for Burned Area Mapping



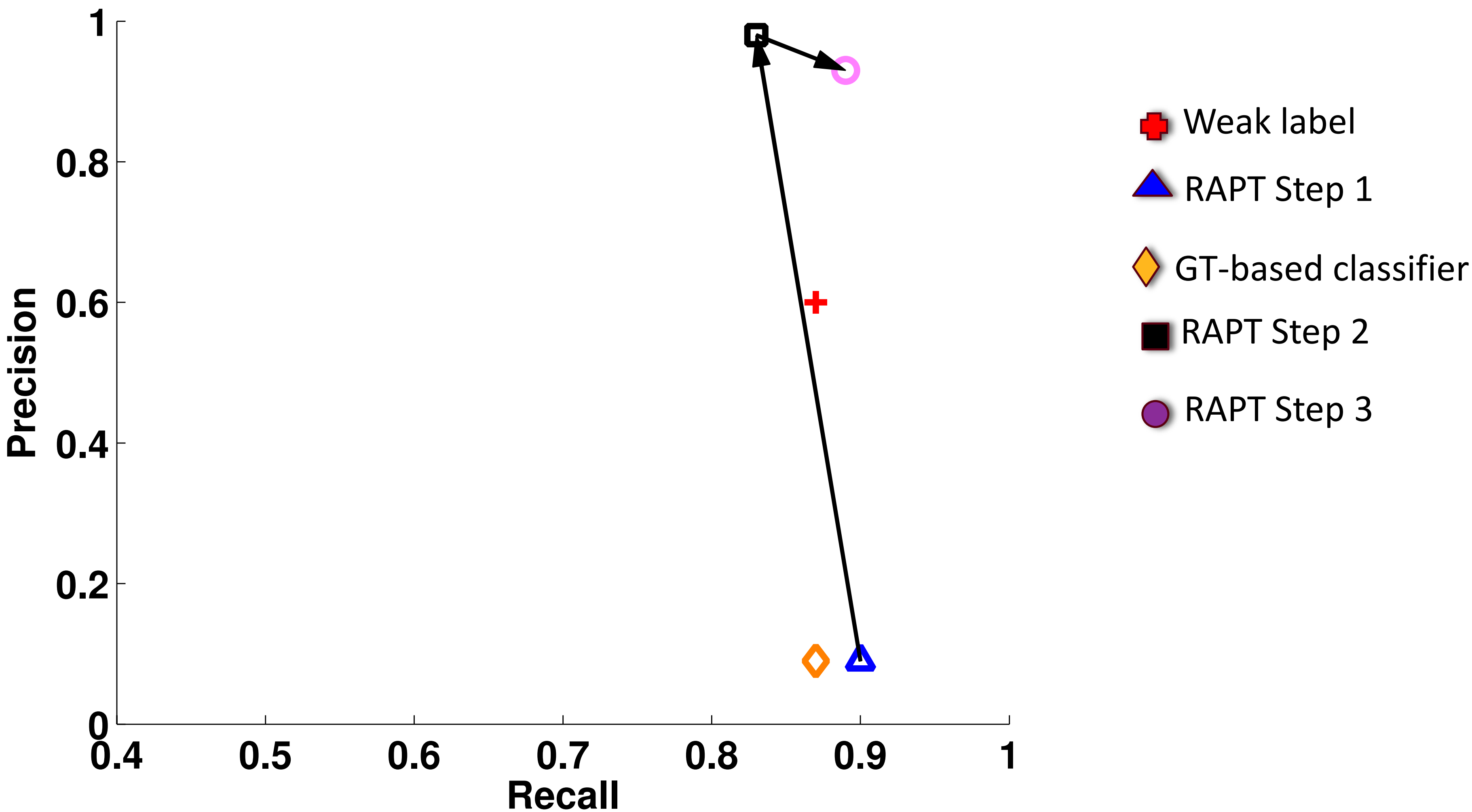
California State

Results for Burned Area Mapping



Georgia State

Results for Burned Area Mapping



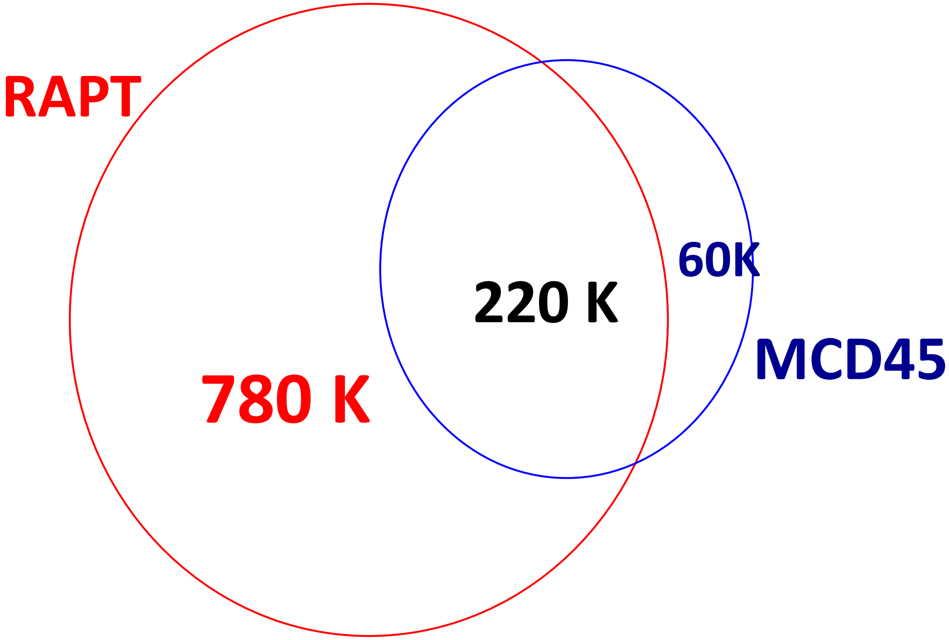
Montana State

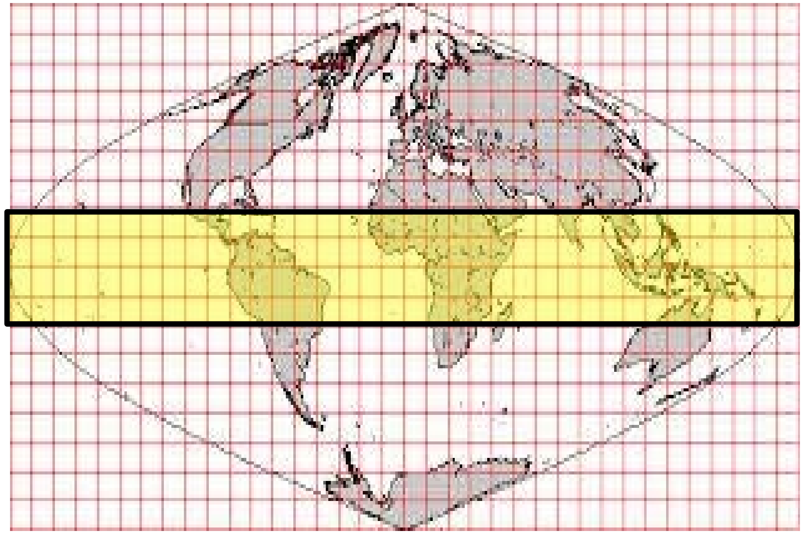
Global Monitoring of Fires in Tropical Forests

Fires in tropical forests during 2001-2014

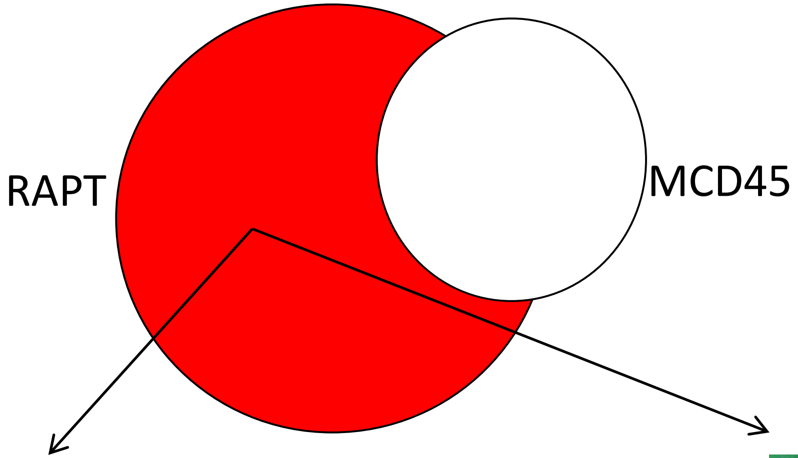
1 million sq. km. burned area found in tropical forests

- *more than three times the total area reported by state-of-art NASA products*.

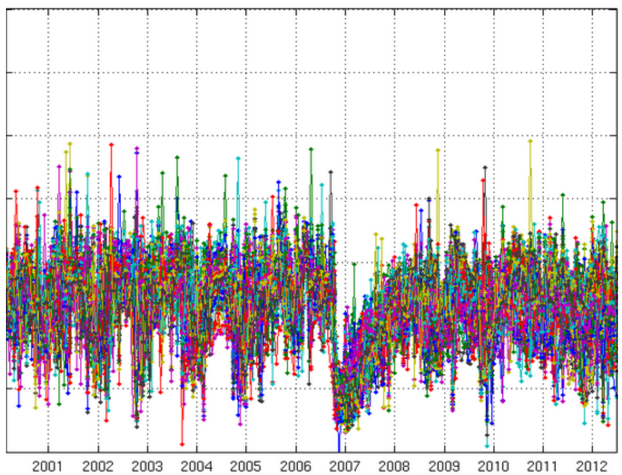




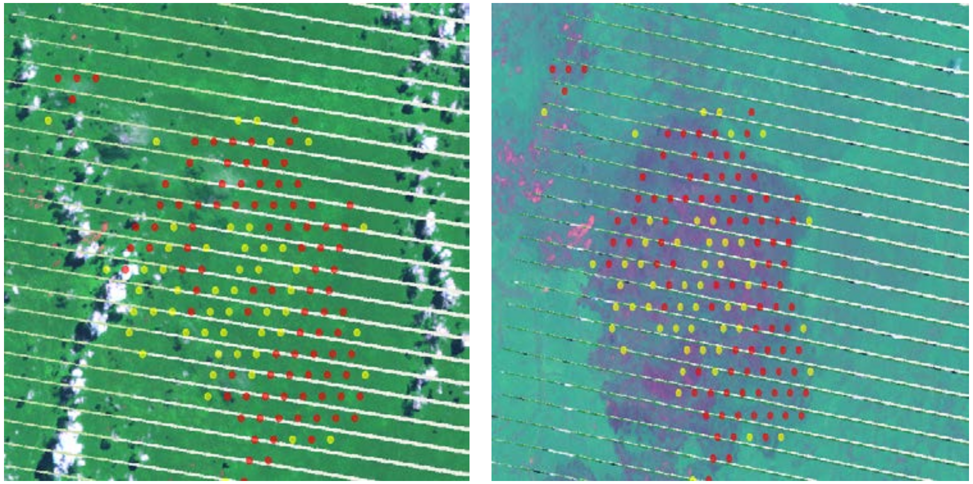
Validation: Multiple sources



Change in Vegetation series



Burn scar in Landsat composite



Before Fire Event **After Fire Event**

Validation confirmed that the additional burned areas detected using RAPT are actual burns missed by state-of-art products

Validation: Burn Index

A burn index tries to capture the degree of burn at a location and is computed as a function of spectral values before and after the event.

$$NBR = \frac{band2 - band7}{band2 + band7}$$

$$dNBR = NBR_{prefire} - NBR_{postfire}$$

A commonly used index is ***dNBR***

- Used for validation in previous studies, including MCD45

Validation: Burn Index

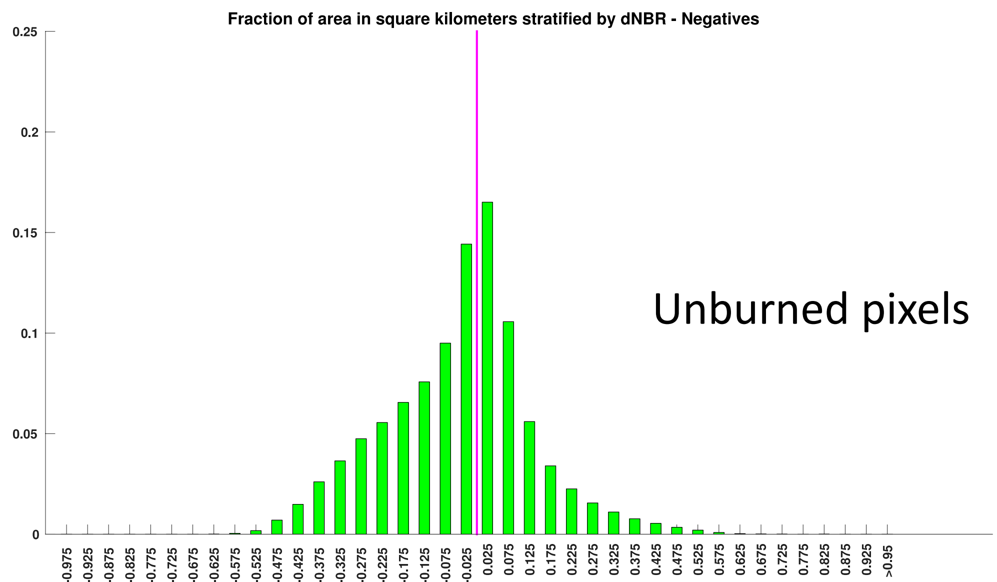
A burn index tries to capture the degree of burn at a location and is computed as a function of spectral values before and after the event.

$$NBR = \frac{band2 - band7}{band2 + band7}$$

$$dNBR = NBR_{prefire} - NBR_{postfire}$$

A commonly used index is ***dNBR***

- Used for validation in previous studies, including MCD45



Validation: Burn Index

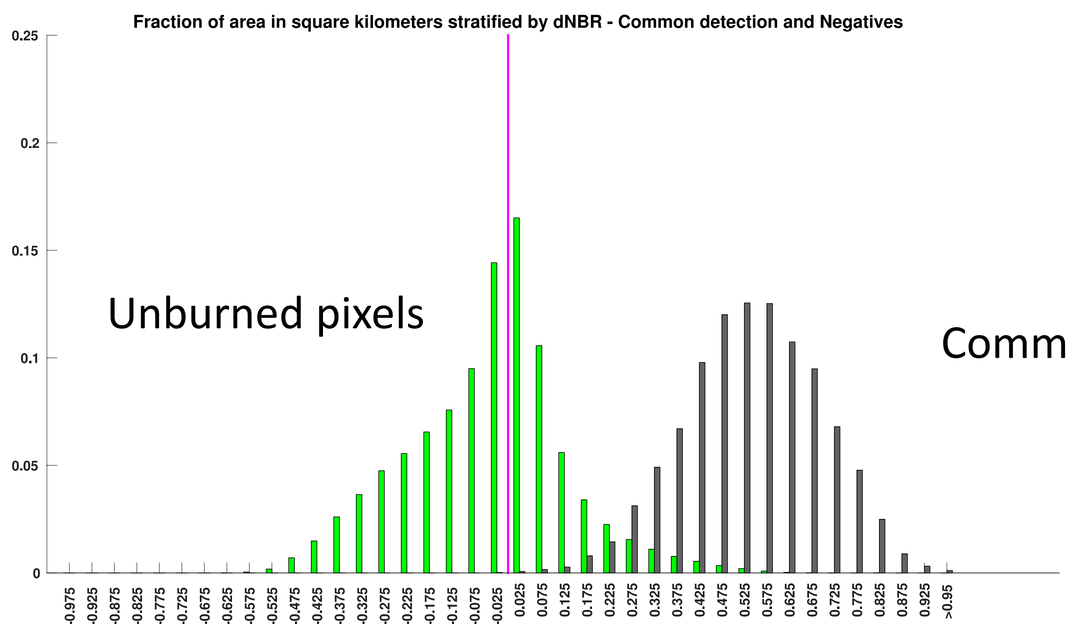
A burn index tries to capture the degree of burn at a location and is computed as a function of spectral values before and after the event.

$$NBR = \frac{band2 - band7}{band2 + band7}$$

$$dNBR = NBR_{prefire} - NBR_{postfire}$$

A commonly used index is ***dNBR***

- Used for validation in previous studies, including MCD45



Validation: Burn Index

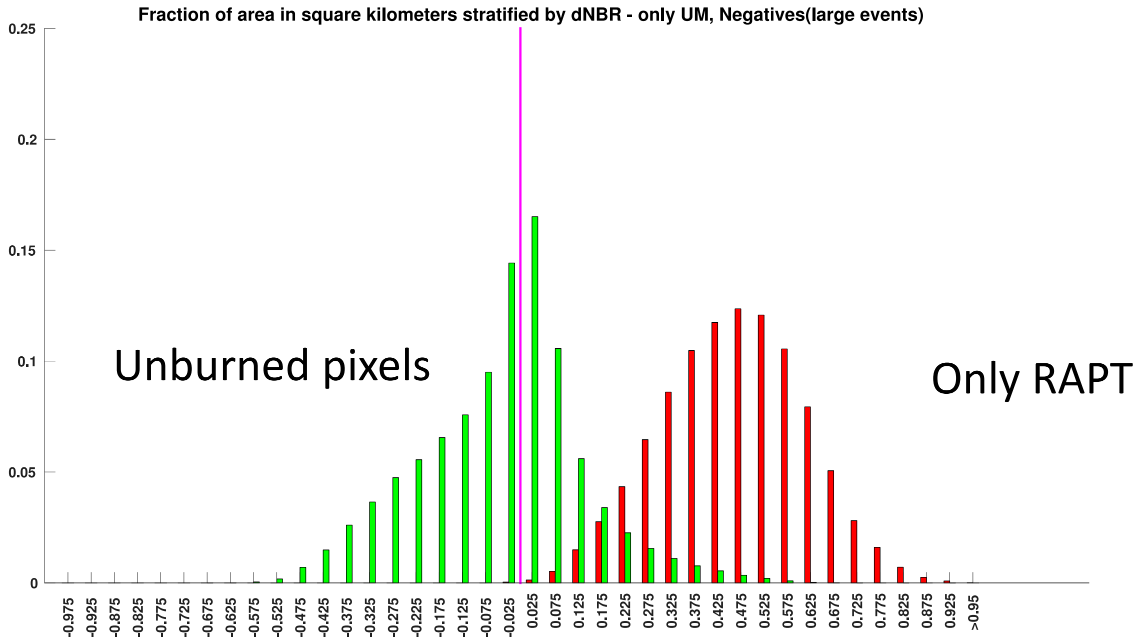
A burn index tries to capture the degree of burn at a location and is computed as a function of spectral values before and after the event.

$$NBR = \frac{band2 - band7}{band2 + band7}$$

$$dNBR = NBR_{prefire} - NBR_{postfire}$$

A commonly used index is ***dNBR***

- Used for validation in previous studies, including MCD45



Validation: Burn Index

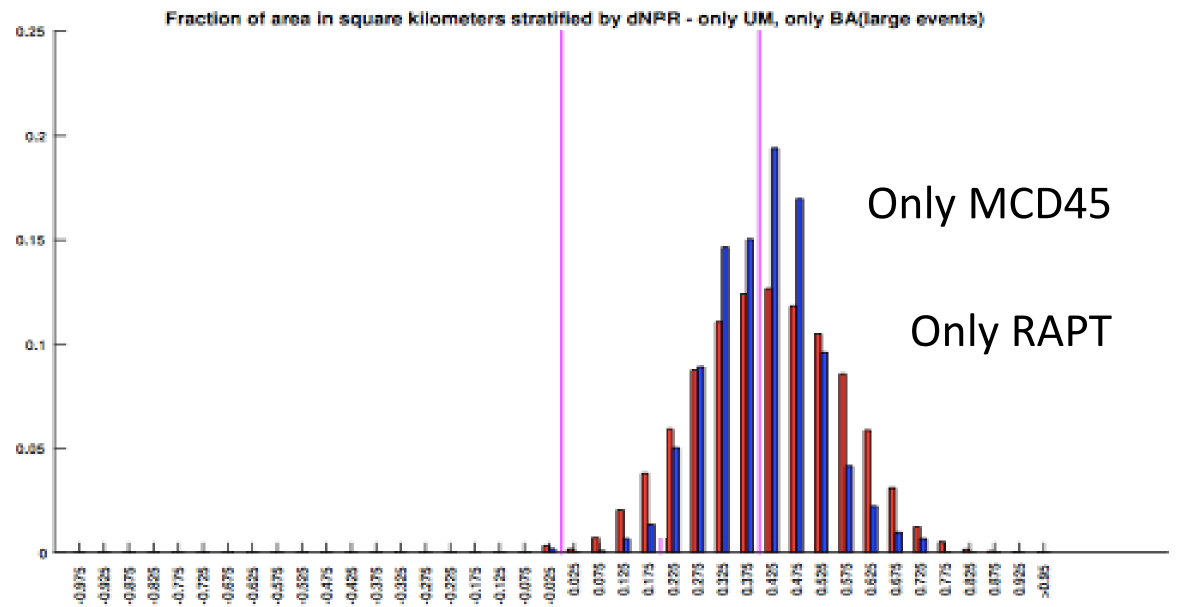
A burn index tries to capture the degree of burn at a location and is computed as a function of spectral values before and after the event.

$$NBR = \frac{band2 - band7}{band2 + band7}$$

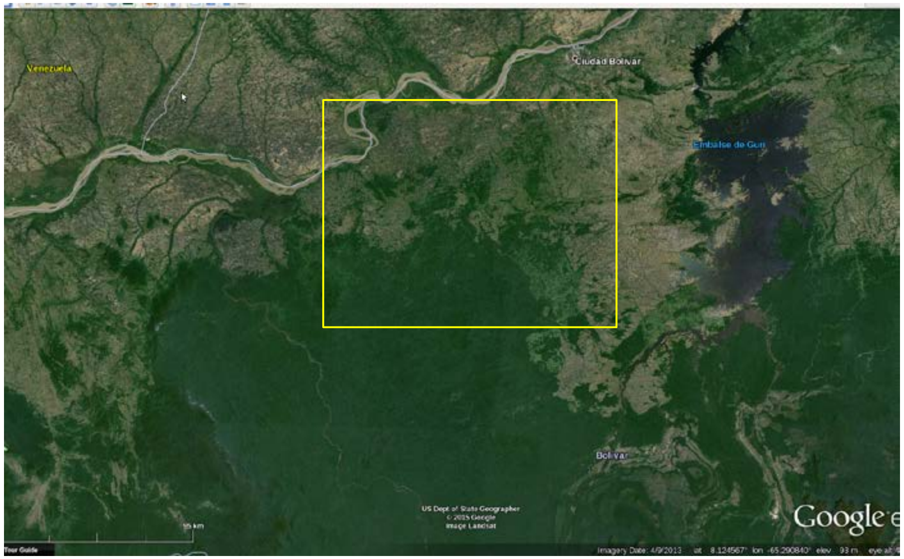
$$dNBR = NBR_{prefire} - NBR_{postfire}$$

A commonly used index is ***dNBR***

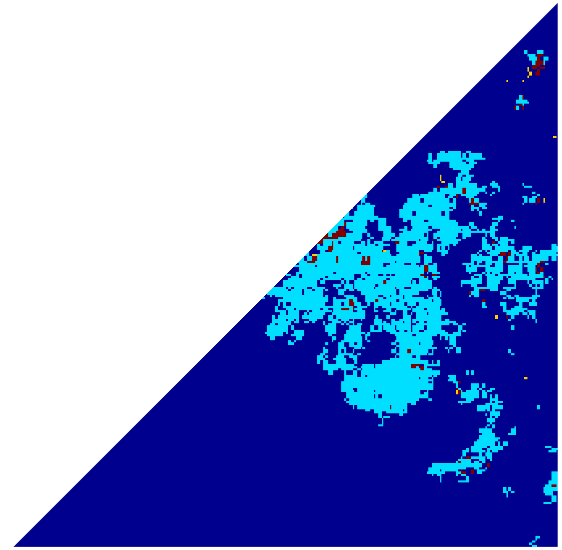
- Used for validation in previous studies, including MCD45



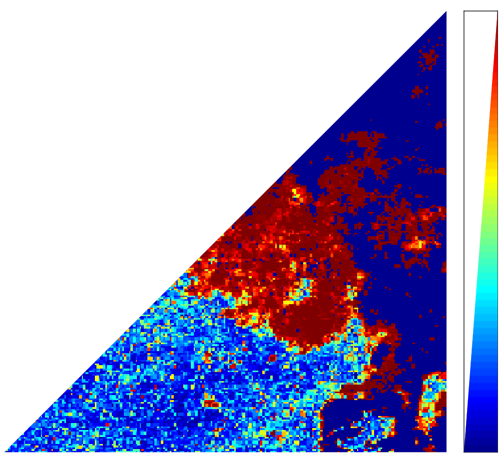
Dynamics of Fire Event



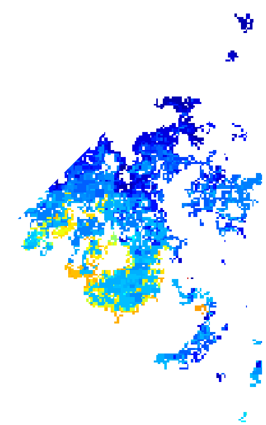
Region in North Brazil



Comparison with MCD45



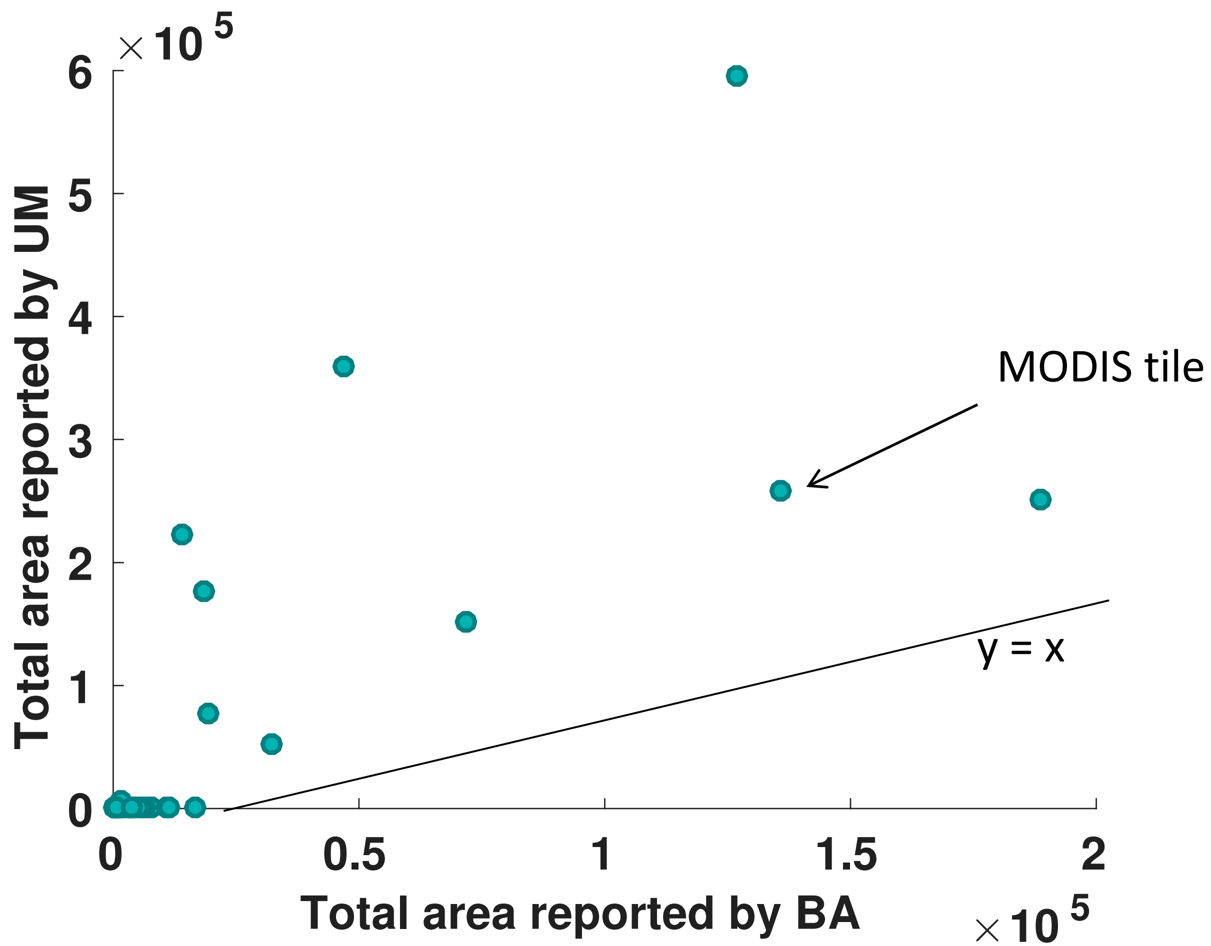
Probability of burn



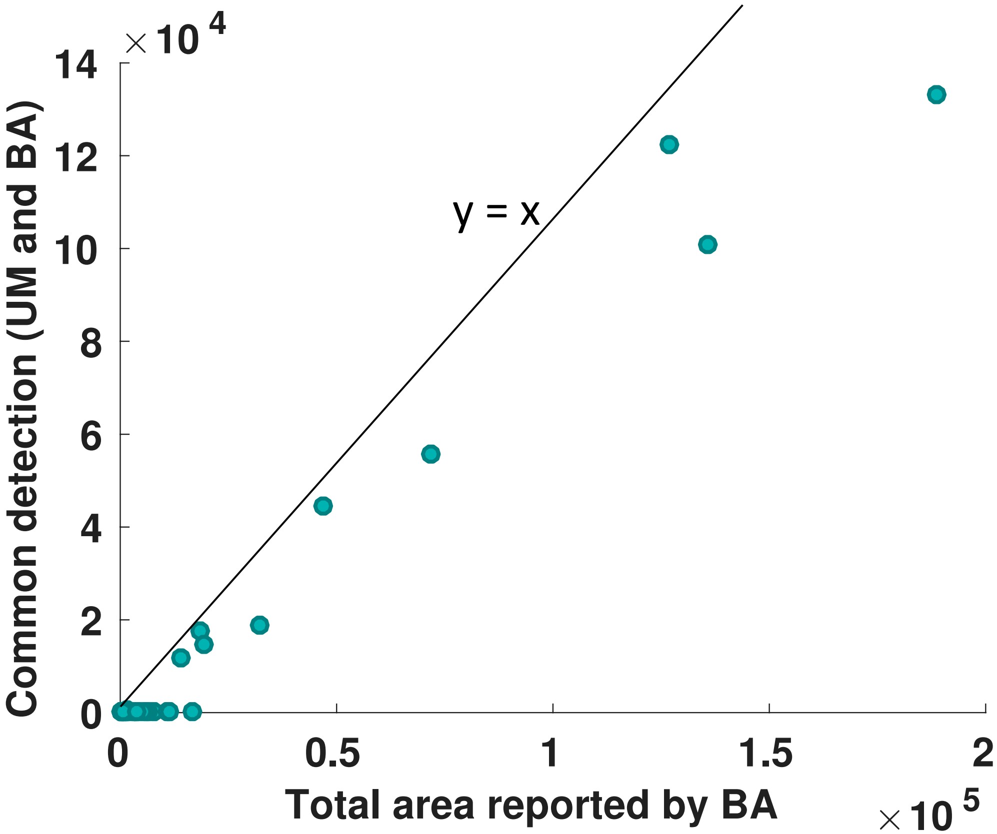
Time of burn

Questions?

Comparing with total burned areas reported by MCD45



What fraction of MCD45 do we recall?



Comparison of exclusive burned areas