ILLINOIS INSTITUTE OF TECHNOLOGY

Active Learning with Rationales

Mustafa Bilgic

In collaboration with my PhD students

Bio

- Education
  - BS, University of Texas at Austin, 2000
    - Recommender systems and explanations
    - w/ Raymond Mooney
  - PhD, University of Maryland at College Park, 2010
    - Active learning and statistical relational learning
    - w/ Lise Getoor
- Currently
  - Associate Professor of Computer Science
  - Director of the Machine Learning Laboratory
  - Director of the Masters in Artificial Intelligence Program

Research Interests

- Machine learning
- Probabilistic graphical models
- Recommender systems
- Active machine learning

Recent Projects

- Active learning
- Active inference
- Learning with rationales
- Filter bubbles in news recommender systems
- Deep learning for biological image analysis
- Active evaluation
- Human-like classification

Recent Projects

- **Active learning**
- Active inference
- **Learning with rationales**
- Filter bubbles in news recommender systems
- Deep learning for biological image analysis
- **Active evaluation**
- **Human-like classification**

Machine Learning Background

Machine Learning

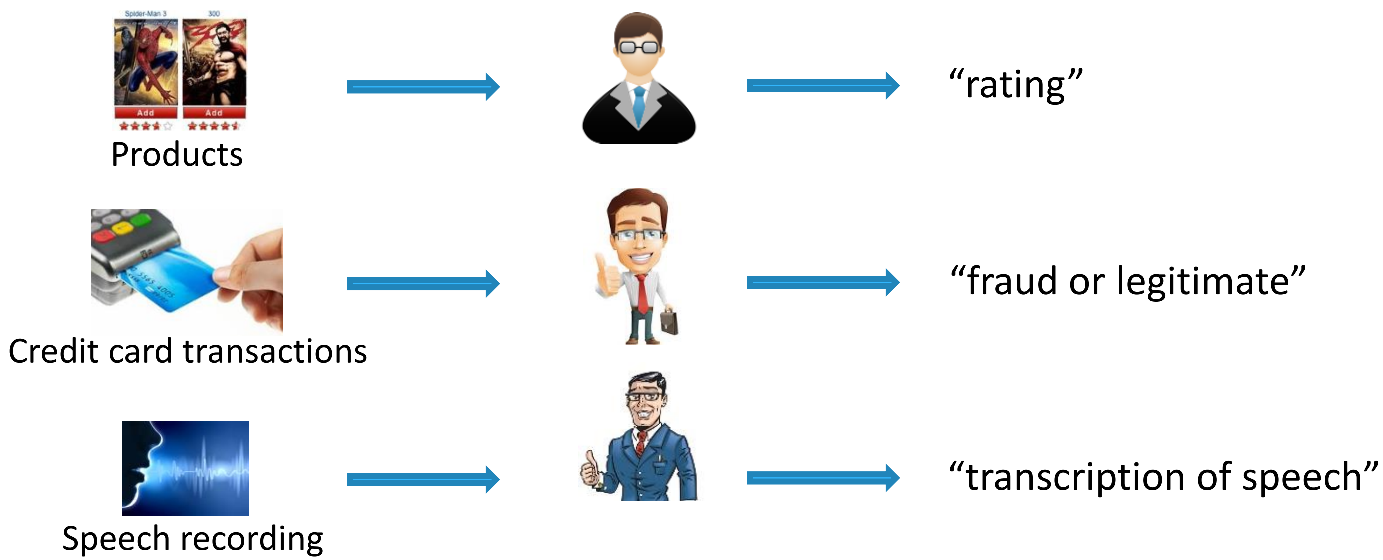
- Supervised learning
- Unsupervised learning
- Reinforcement learning

Machine Learning

- **Supervised learning**
- Unsupervised learning
- Reinforcement learning

Supervised learning

- Train a predictive model on instances of data
- The model is a function that maps data to a target ***f*:*X*→*Y***
- Humans provide the ***supervision*** on instances



The X in *f:X→Y*

1. Feature-based representation
   - Each instance is a vector of features
     - A patient: symptoms, laboratory test results, …
     - A loan application: income, credit score, …
     - A document: a bag of words
     - An image: scale-invariant feature transform (SIFT)
     - Gene sequence: n-grams, …
   - Vector-based classifiers
     - Naïve Bayes, logistic regression, decision trees, support vector machines, neural networks, …

The X in *f:X→Y*

2. Similarity-based representation
    - Pair-wise similarity among the instances
        - How similar are these images, documents, gene sequences, …?
    - Similarity-based classifiers
        - Nearest neighbor, support vector machines

The X in *f:X→Y*

3. Image, text, sequence, "raw" data
   - Let the classifier learn the "features"
   - Neural networks with several hidden layers
     - a.k.a. deep learning
   - Examples
     - Convolutional neural networks for image analysis
     - Long Short-Term Memory networks for text analysis

The Y in *f:X*→*Y*

- The target variable
    - Patients: the diagnosis
    - Loan application: the decision
    - Document: the category
    - Image: the person
- Often, it is hard to obtain, because it might require
    - Expertise
    - Manual labor
    - Laboratory tests

Bias-Variance Trade-off

- The more assumptions a model makes, the less data it needs
  - Naïve Bayes typically requires less data than logistic regression
- The fewer the assumptions a model makes, the more data it needs
  - Deep learning with millions of parameters
  - GPT-3 has 175 billion parameters

1. Active Learning

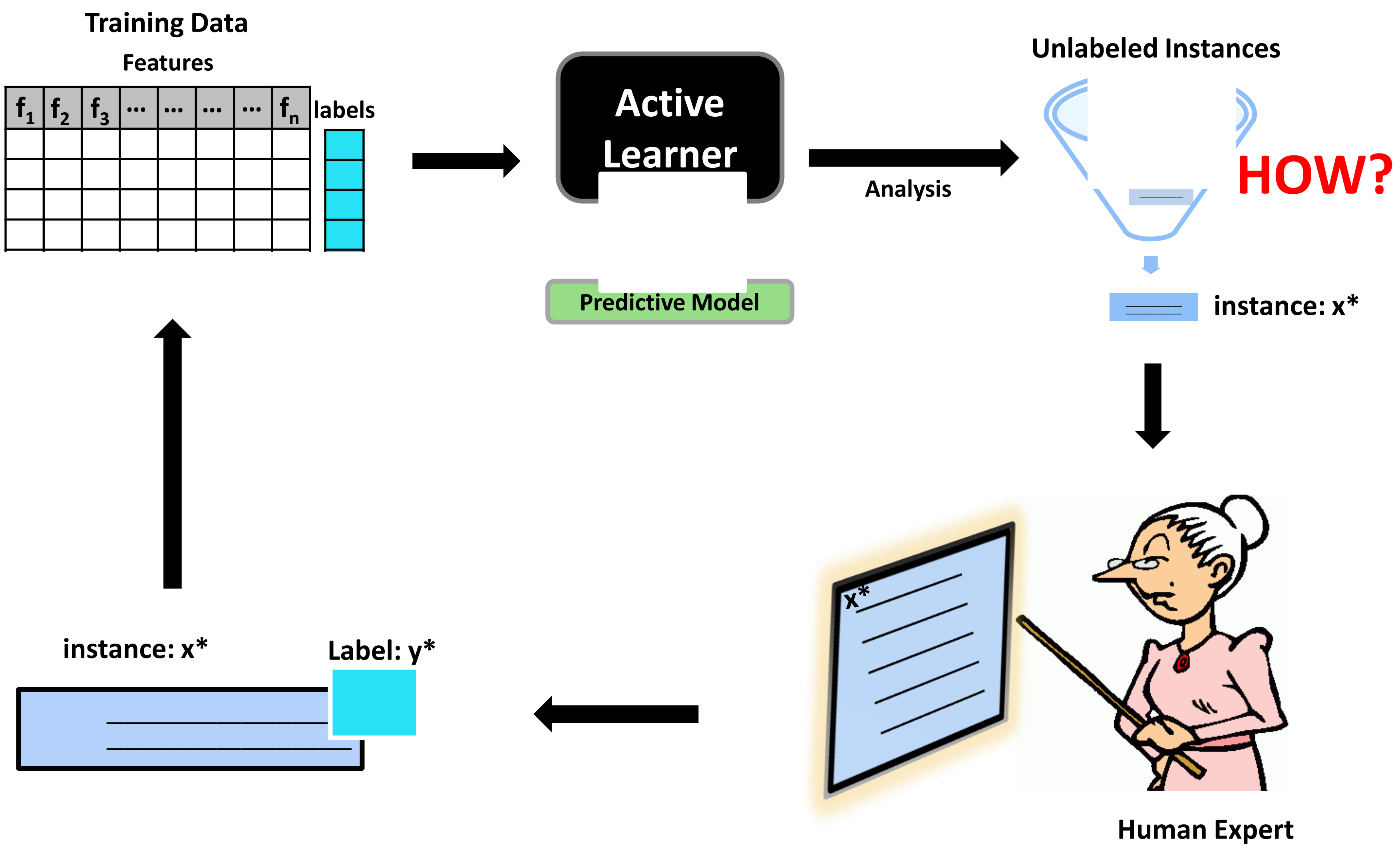
Active Learning

- The *X* is plenty; the *Y* is scarce
- X; Y
  - Images; annotations
  - Speech; transcription
  - Text; translation
  - Review; sentiment
  - News; category
  - …

How to choose few, but useful instances for labeling?

Active Learning (AL)

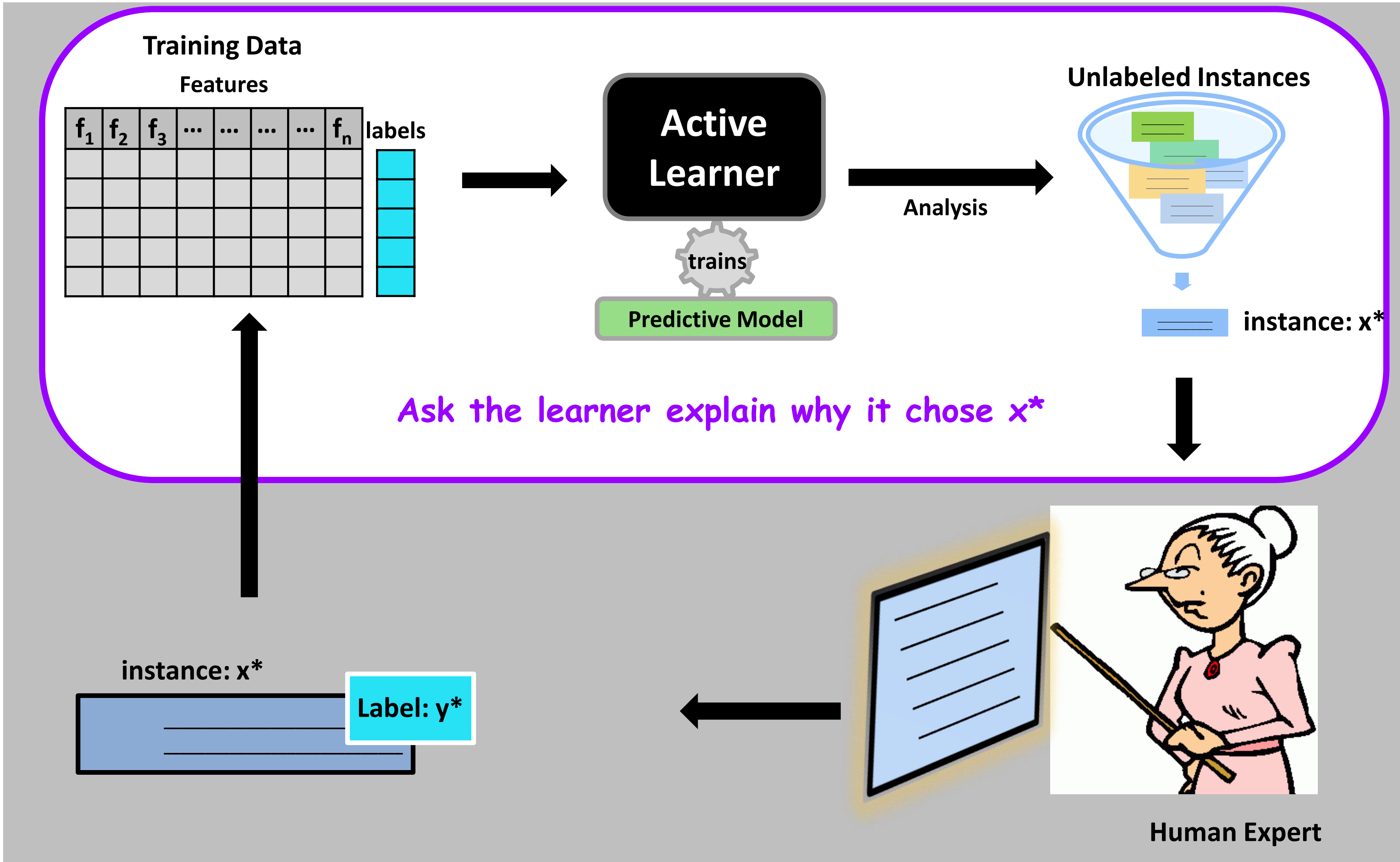
Active learning algorithm



Active learning strategies

- Common utility-based active learning algorithms:
  - Query-by-Committee [Seung, Opper & Sompolinsky, COLT'92]
  - Uncertainty Sampling [Lewis & Gale, SIGIR'94]
  - Variance Reduction [Cohn, Ghahramani & Jordan, JAIR'96]
  - Bias Reduction [Cohn, NIPS'97]
  - Expected Error Reduction [Roy & McCallum, ICML'01]
  - And many more…

Ask the learner "why"



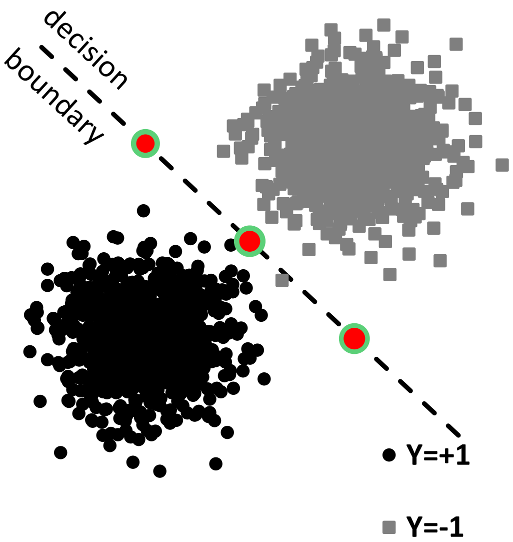
Uncertainty sampling

[Lewis & Gale, SIGIR'94]

- Queries instances about which the classifier is most uncertain how to label

- E.g., entropy as an uncertainty measure

$$x^* = \underset{x^{(i)} \in \mathcal{U}}{\operatorname{argmax}} - \sum_{y \in Y} P_\theta(y|x^{(i)}) \log \left(P_\theta(y|x^{(i)})\right)$$



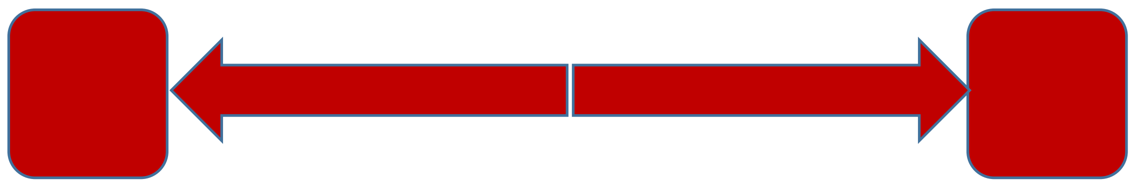
Ask the learner why it is uncertain about x*

Evidence-based framework

We discovered two reasons for model's uncertainty on instances

Conflicting-evidence uncertainty:

$E_{-1}(X)$ $E_{+1}(X)$



Insufficient-evidence uncertainty:

$E_{-1}(X)$ $E_{+1}(X)$

Traditional uncertainty sampling:

Does not consider the reasons for uncertainty, as long as $E_{-1}(X) \approx E_{+1}(X)$

Datasets & measures

Eight datasets:

Dataset	Domain	Size	Minority class %
Spambase	Email. classification	4,601	39.4%
Ibn Sina	Handwriting recognition	20,722	37.8%
Calif. Housing	Social	20,640	29%
Nova	Text processing	19466	28.4%
Sick	Medical	3,772	6.1%
Zebra	Embryology	61,488	4.6%
LetterO	Letter recog.	20,000	4%
Hiva	Chemo-inform.	42,678	3.5%

Medium-imbalanced

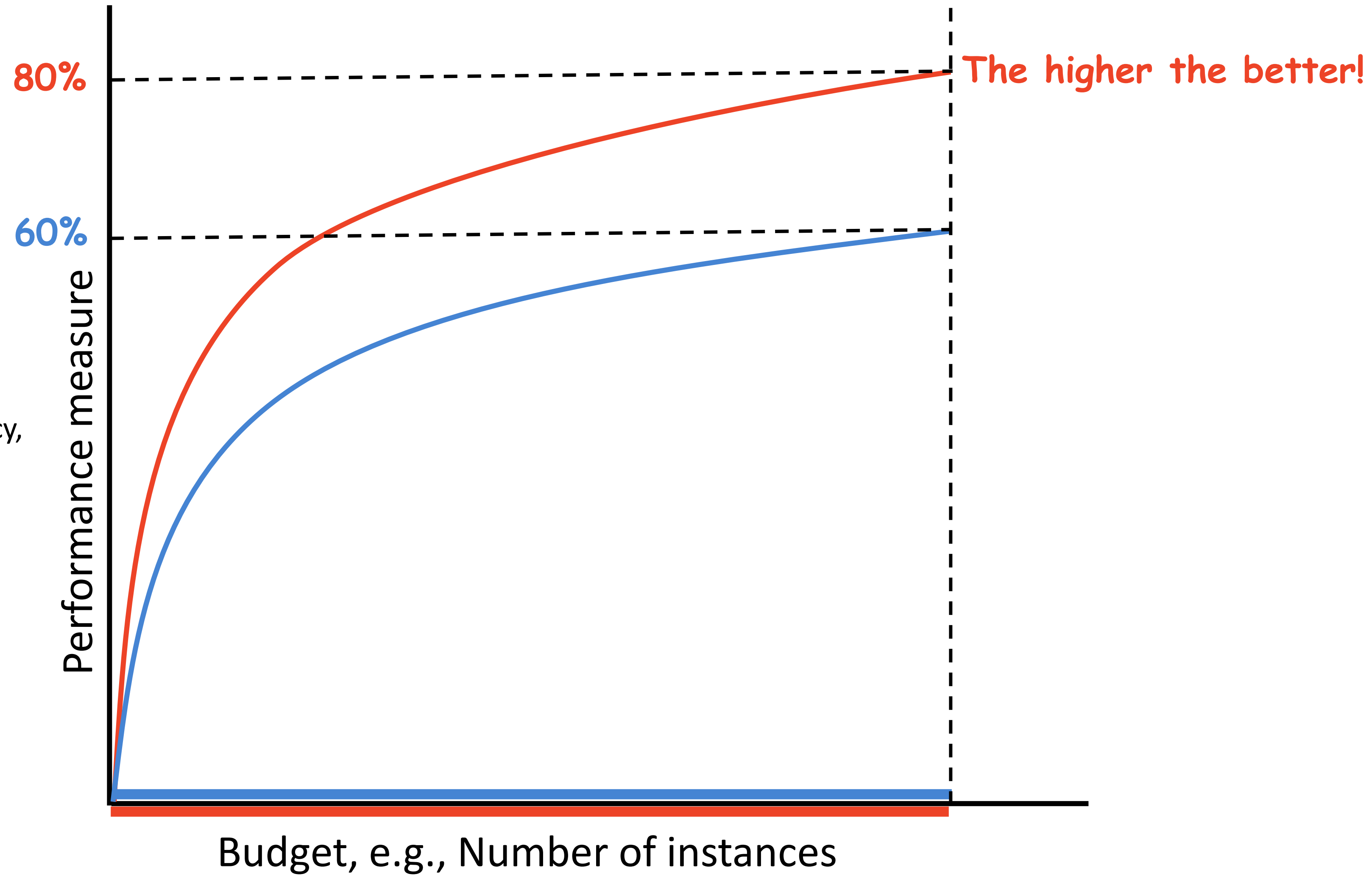
Highly-imbalanced

Performance measures:

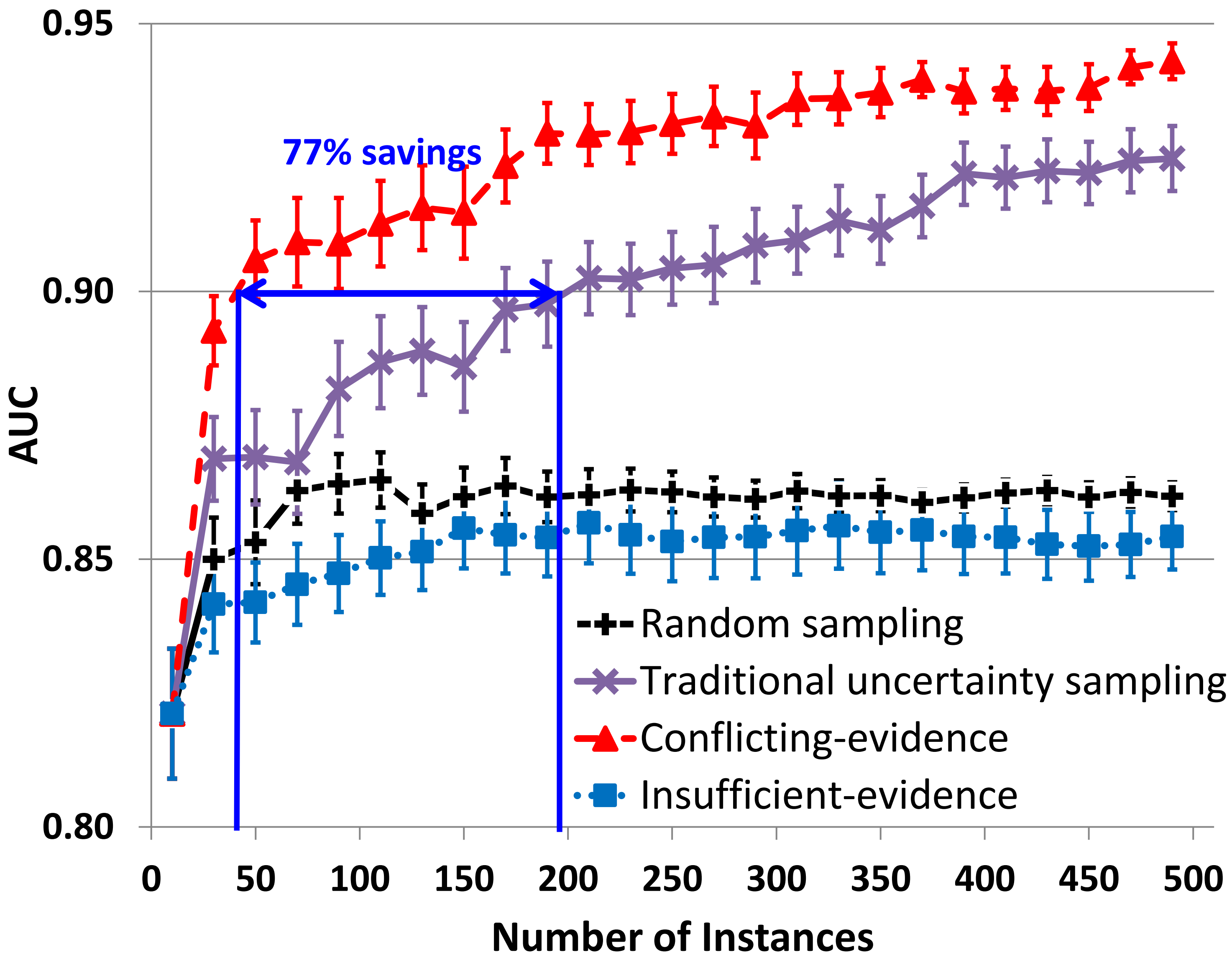
- **AUC**: All datasets
- **Accuracy**: Medium-imbalanced datasets
- **F1**: Highly-imbalanced datasets

How to interpret the results?

E.g., Accuracy, AUC, F1

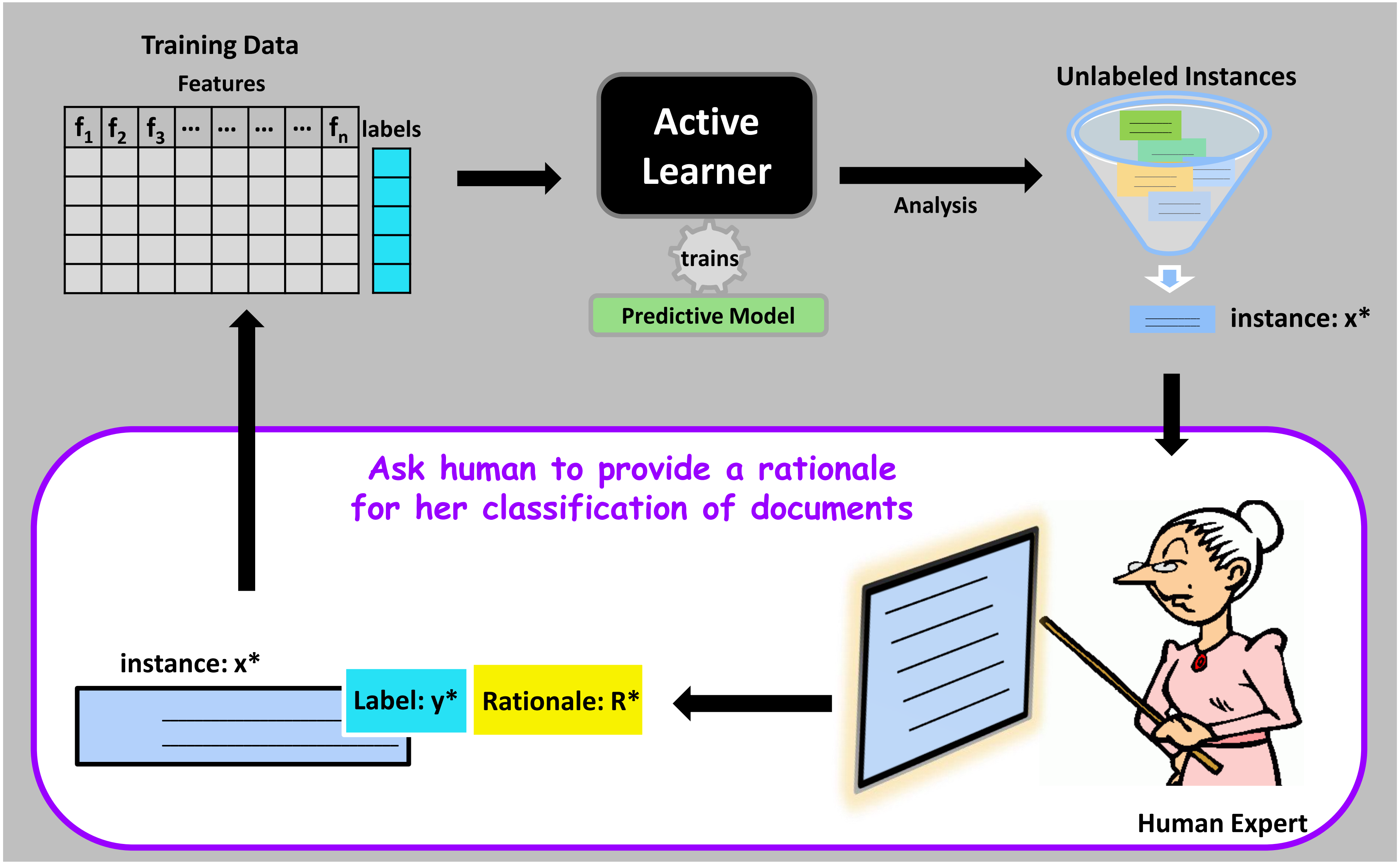


Results – Ibn Sina dataset



2. Learning with Rationales

Ask the humans "why"



2. Learning with Rationales
2.a. Text Classification

The approach

Document 1	This is a great movie.	Positive
Document 2	The plot was great, but the performance of actors was terrible. Avoid it.	Negative
Document 3	I’ve seen this at an outdoor cinema; great atmosphere. The movie was terrific.	Positive

How do we use <x, y, r> for supervised learning?

Datasets & experimental setup

Four text classification datasets:

Dataset	Description	# instances	# Features
IMDB	Sentiment analysis of movie reviews	25,000	27,272
NOVA	20 Newsgroups dataset: Email classification	12,977	16.969
SRAA	UseNet articles: Aviation vs. Auto	48,812	31,883
WvsH	20 Newsgroups dataset: Windows vs. Hardware	1176	4,026

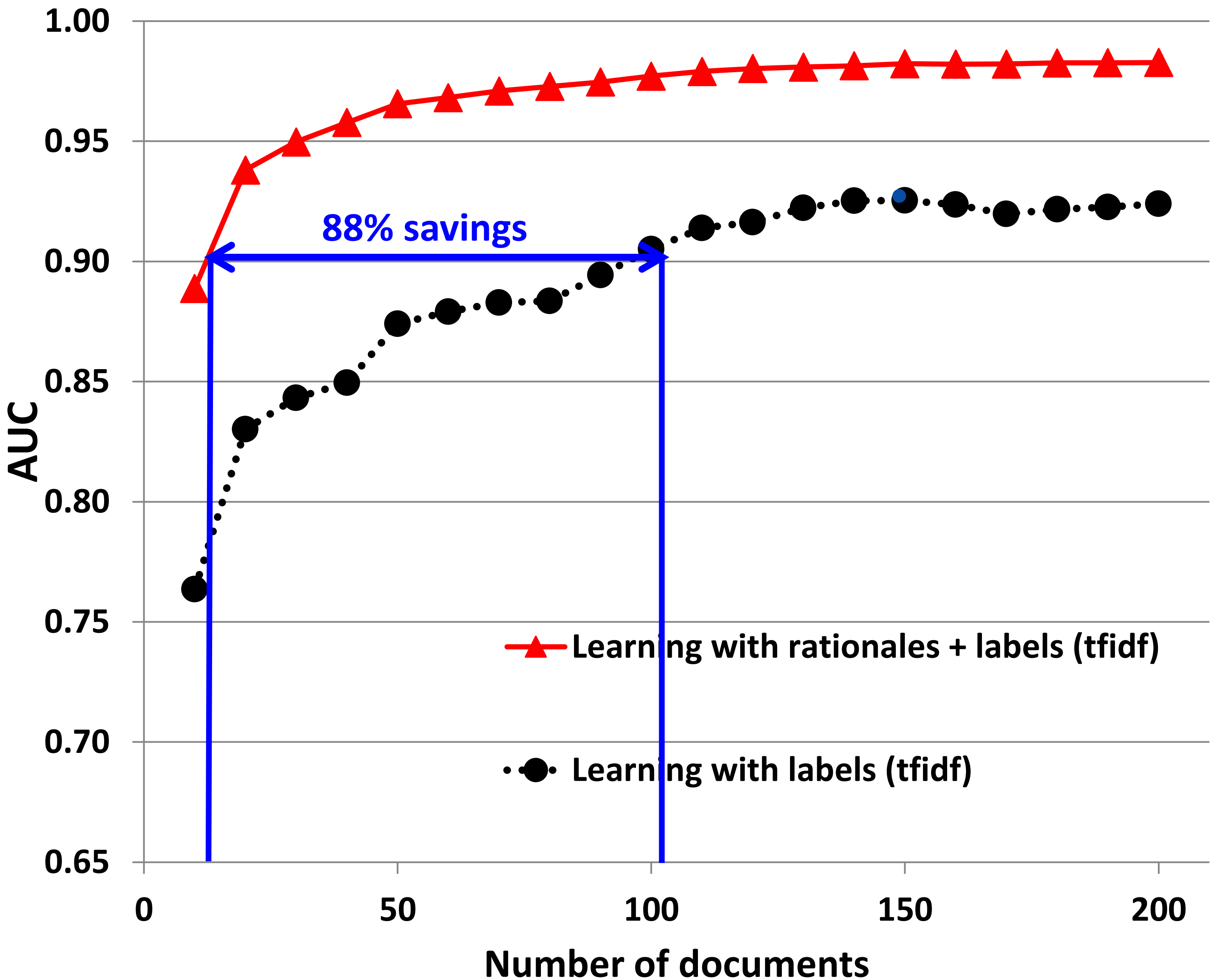
Three classifiers:

- Multinomial naïve Bayes (MNB)
- Logistic regression (LR)
- Support vector machines (SVM)

Two data representations:

- Binary
- Tf-idf

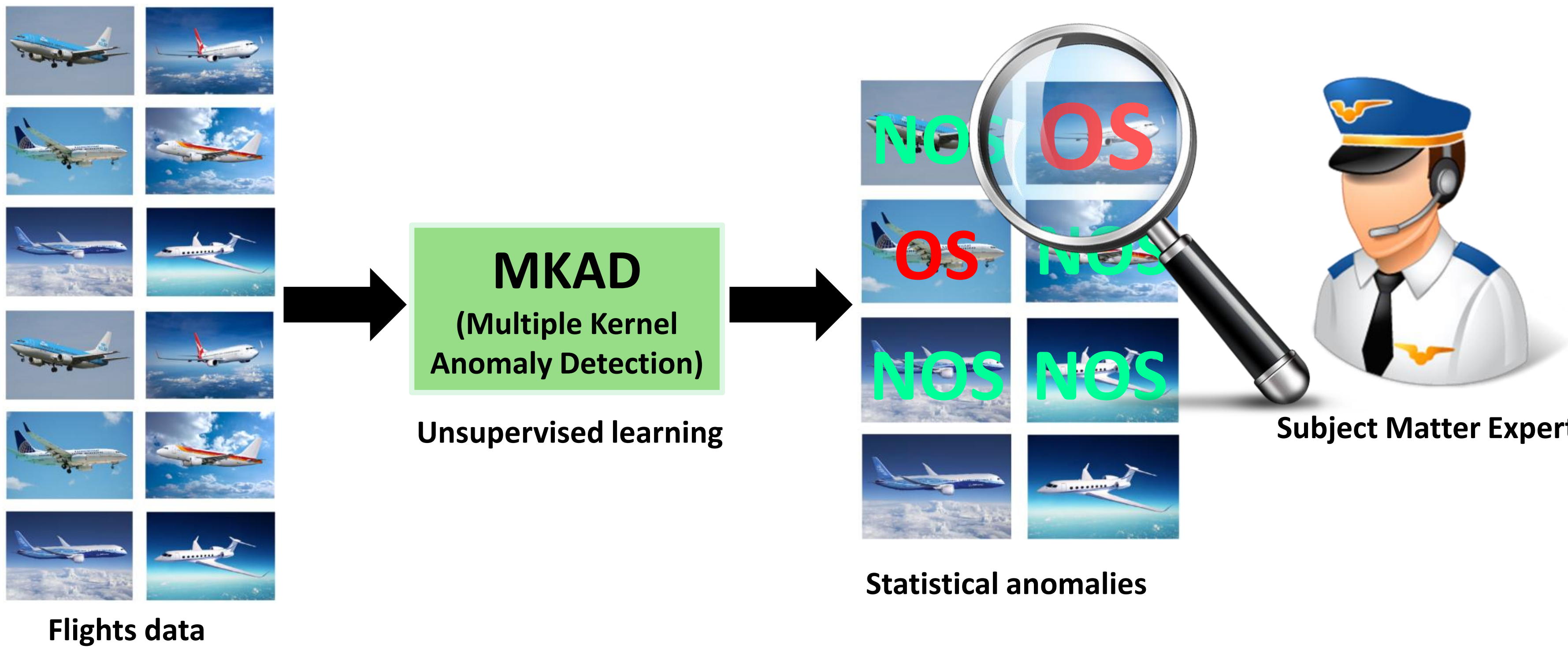
Results – SRAA dataset with MNB



2. Learning with Rationales
2.b. Anomalous Flight Detection

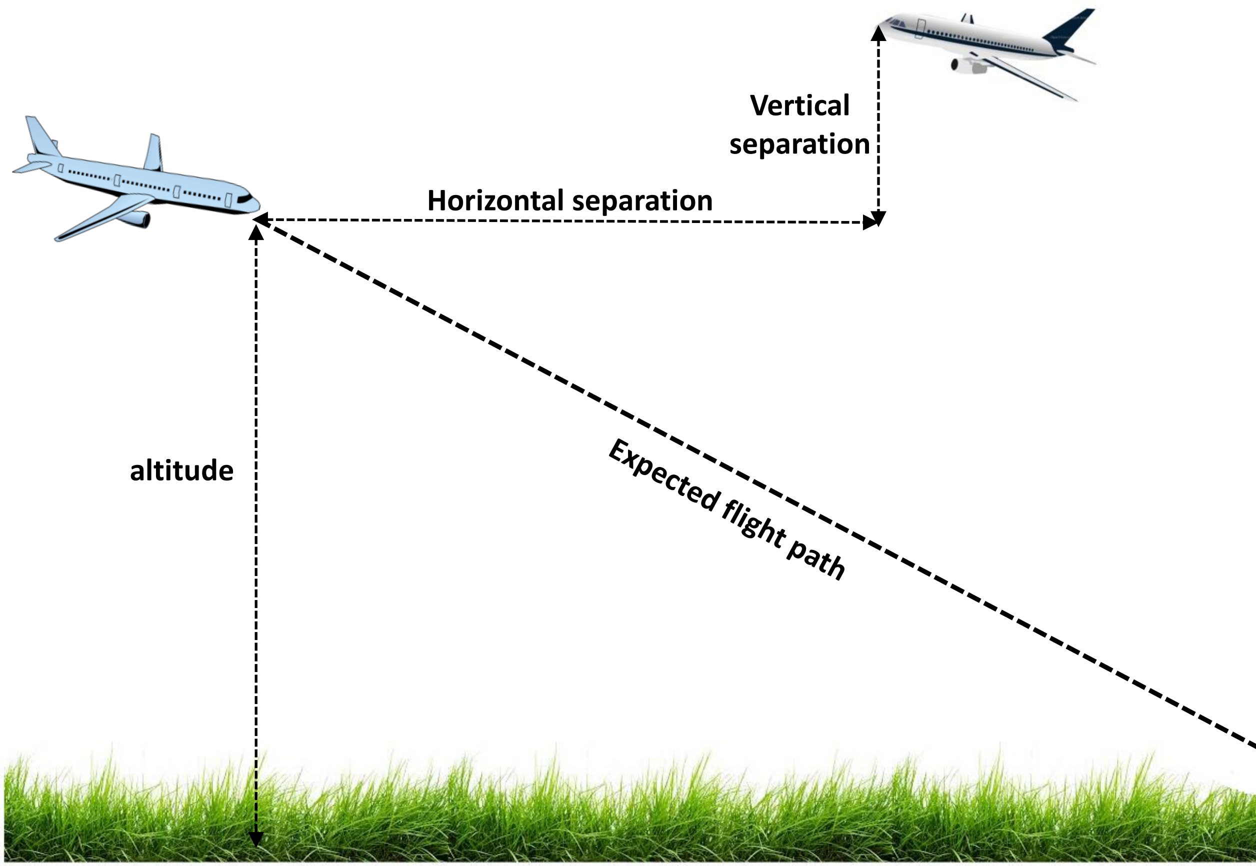
Collaboration w/ NASA

OS: operationally significant
NOS: not operationally significant



GOAL: ***effectively*** train a model to identify operationally significant (OS) anomalies using ***less time*** of experts

Flights data



ORIGINAL FEATURES

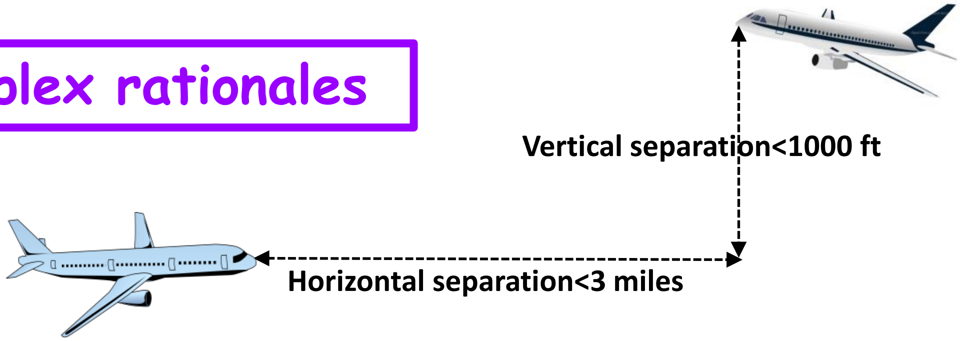
- Latitude
- Longitude
- Altitude
- Ground speed
- Horizontal separation
- Vertical separation
- Aircraft size
- Turn-to-final (TTF) parameters:
    - Maximum overshoot
    - Speed at TTF
    - Distance at TTF
    - Angle at TTF
    - Altitude difference at TTF
- Nearest neighboring (NN) flight info:
    - NN flight on same runway
    - NN flight on parallel runway
    - NN flight part of the same flow

Rationales

More complex rationales

"Loss of separation"

- Horizontal separation < 3 miles AND Vertical separation < 1000 ft AND nearest neighboring flight is not on parallel runways and not part of the same flow



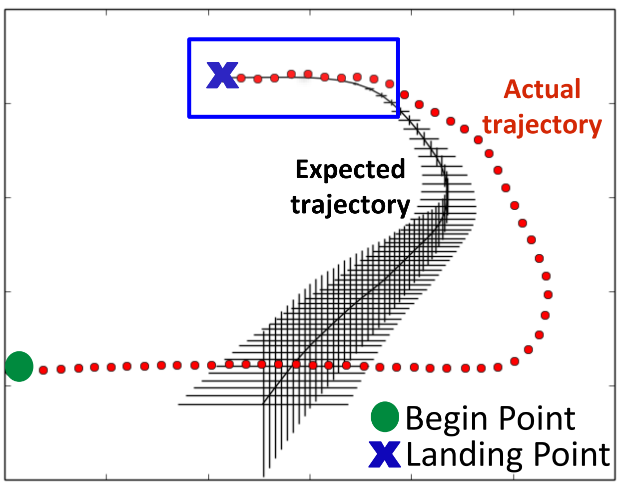
"Large overshoot"

- Maximum overshoot is greater than a threshold based on values of flights with positive labels

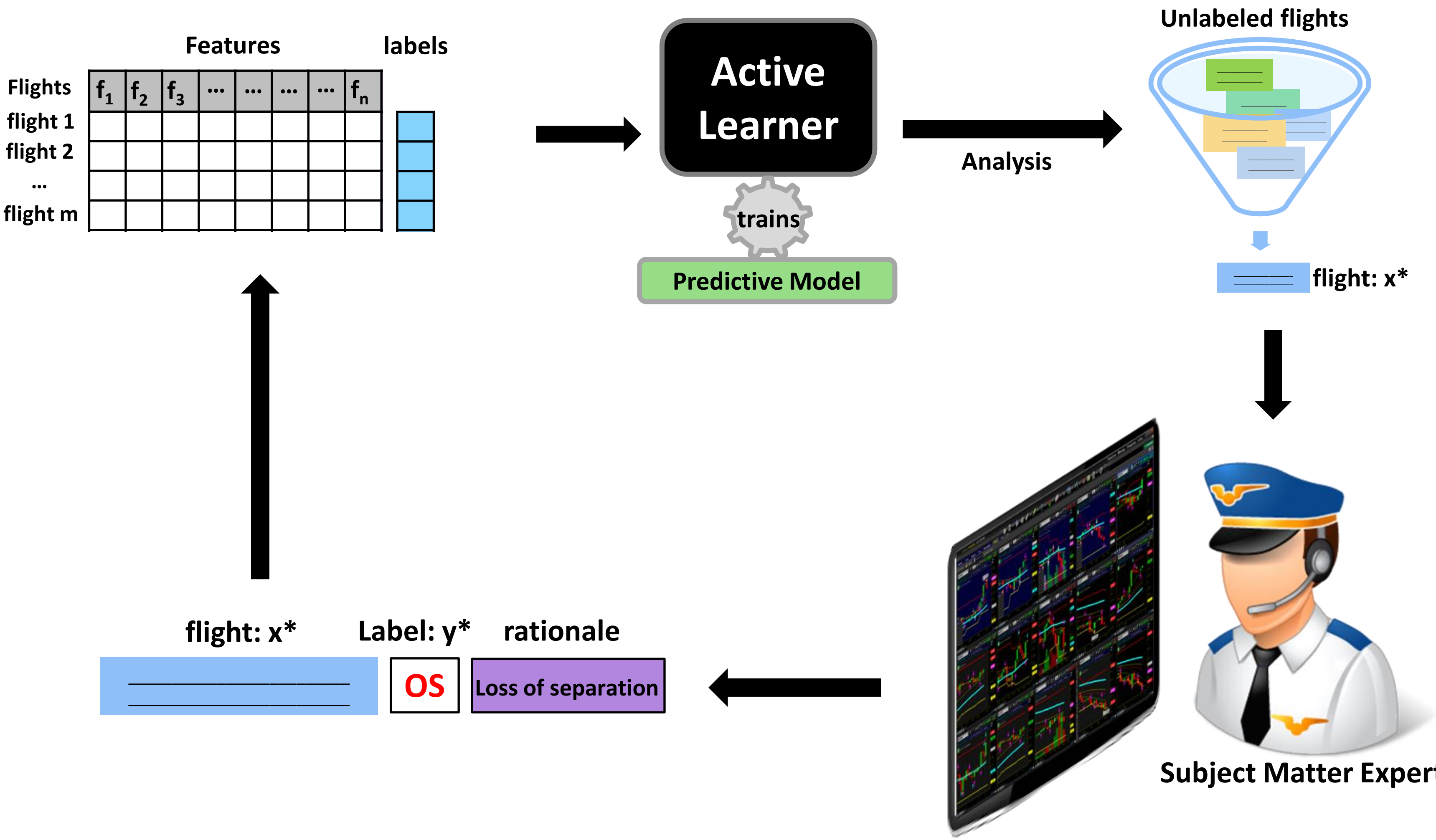
"Unusual flight path"

- Overall deviation from expected (average) trajectory of all landing flights on that runway

Deviation from expected path



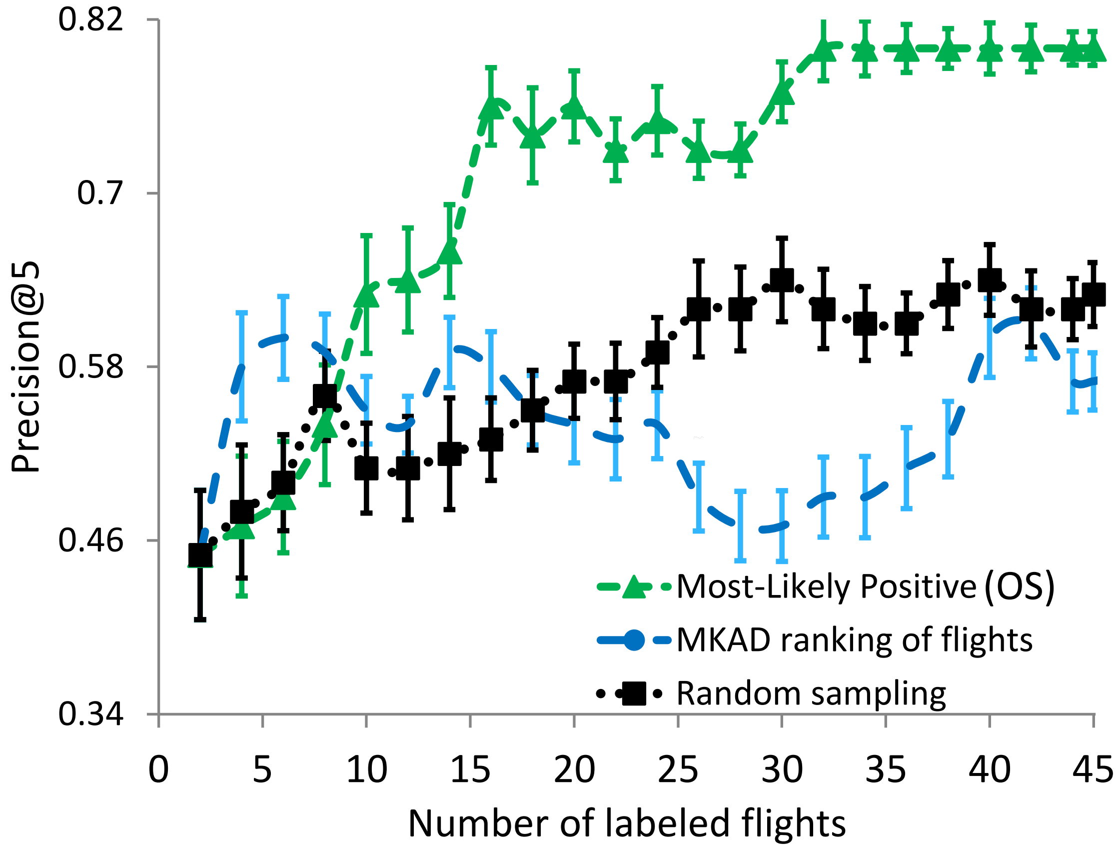
Active learning framework



Selecting informative flights

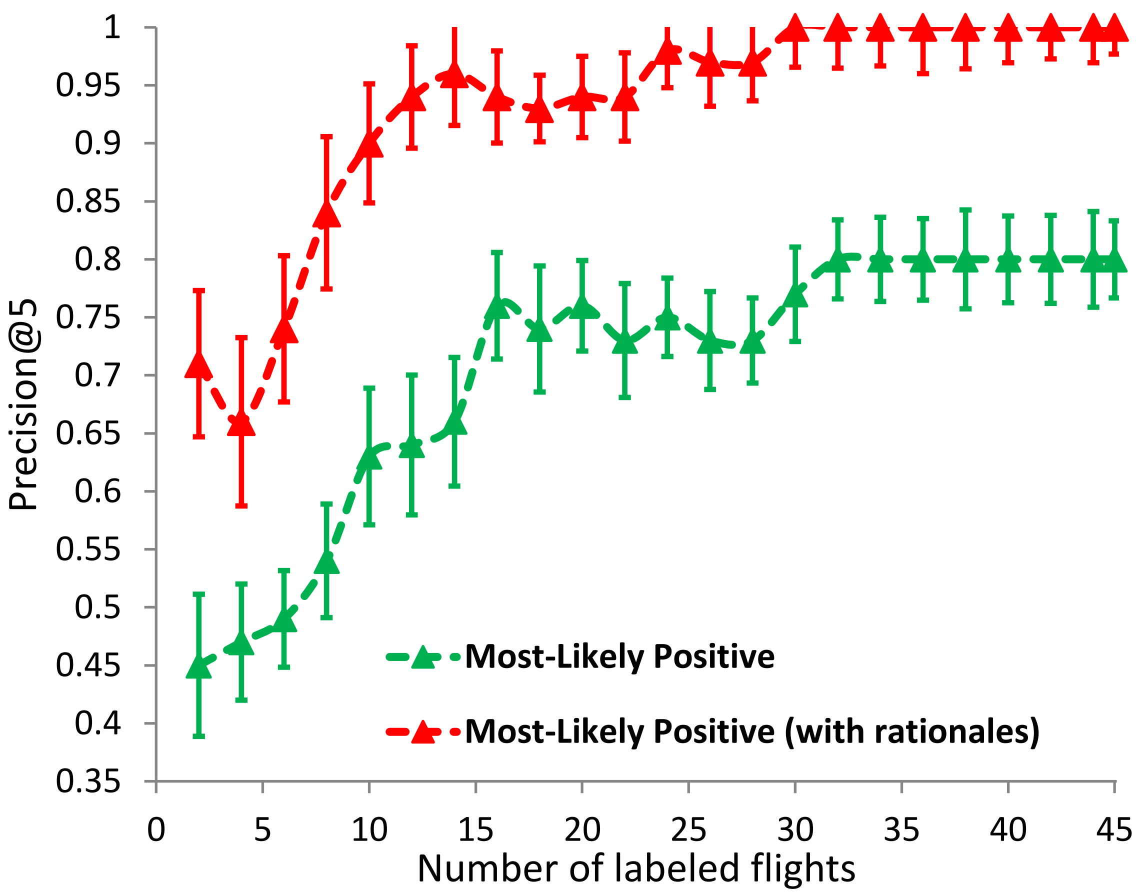
Active learning strategy: Most-likely positive strategy

Objective function: $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathcal{U}} P_\theta(\hat{\mathbf{y}}^+ | \mathbf{x})$



Including rationales into learning

Including rationales improves performance over learning with labels only



3. Fresh, still in the oven, projects

Projects currently in the oven

- Active evaluation
  - Curate a dataset for only evaluation purposes
- Human-like classification
  - Given a case, skim all features but focus on what is most important <u>for</u> that case

Collaboration Opportunities

- We develop methods
  - Active learning, learning with rationales, active evaluation, human-like classification, etc.
- Collaboration opportunities
  1. Application areas
     - If you have problems/datasets where these methods might be applicable (not enough labeled data, experts provide rationales, human-like and interpretable decision making, etc.), I'd be very happy to discuss them and work with you

  AND/OR

  2. Foundational work
     - If you also work on these areas, I'd be very happy to talk to you about potential collaboration opportunities
- Please see the next slide for my contact info

Contact

Email: mbilgic@iit.edu

Lab: http://ml.cs.iit.edu


THANK
YOU
GRACIAS
ARIGATO
SHUKURIA
DANKSCHEEN
TASHAKKUR ATU
YAQHANYELAY
SUKSAMA
EKHMET
TINGKI
BİYAN
SHUKRIA
GRAZIE
MEHRBANI
MAAKE
KOMAPSUMNIDA
GOZAIMASHITA
EFCHARISTO
JUSPAXAR
PALDIES
BOLZİN
MERCI

TEŞEKKÜR