北京航空航天大學
BEIHANG UNIVERSITY

Graph-based active Semi-Supervised Learning: a new perspective for relieving multi-class annotation labor

Lei Huang, Yang Liu, Xianglong Liu, Bo Lang

2014-11-01

outline

- Introduction
  - Learning-based automatic Annotation
  - Semi-Supervised Learning
  - Active learning
- Graph-based active semi-supervised learning framework
  - Graph-based semi-supervised learning
  - Active learning: Minimize expected global uncertainty
  - Real scene efficiency
    - Inductive extensity
    - Incrementally update
    - Sample candidate points
- Experiment
- conclusion

Introduction(1)

- Explosion of the unstructured data (image, text, video …)

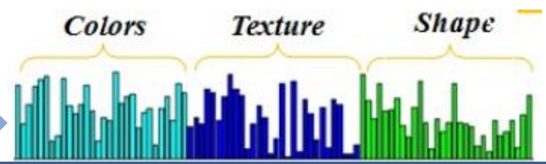
- Labeling or tagging data
  - Manual annotation by human
  - Automatic annotation

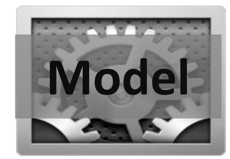
Introduction(2)

- Learning-based automatic annotation

car

Feature extraction



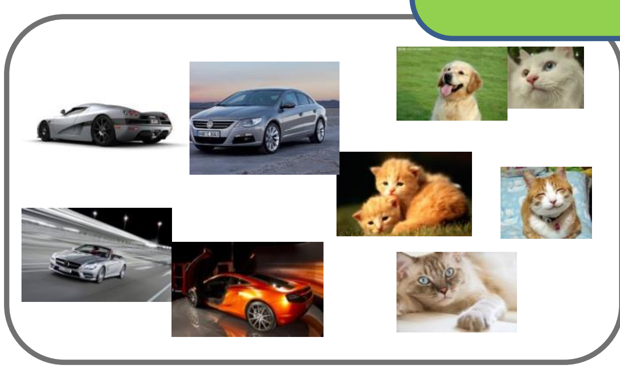


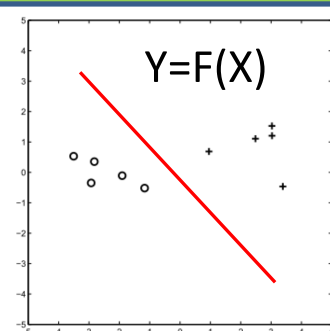
relation

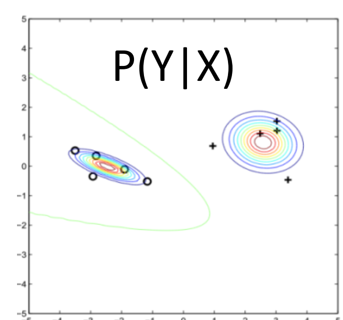
car

bel(Y)

Learning Y=F(X)or P(Y|X)

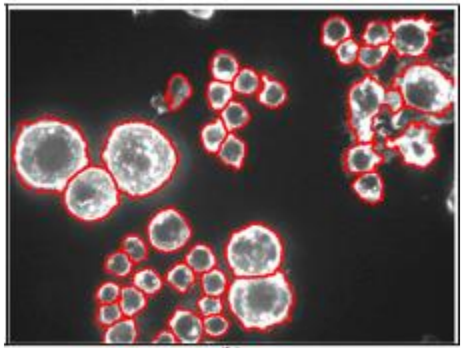




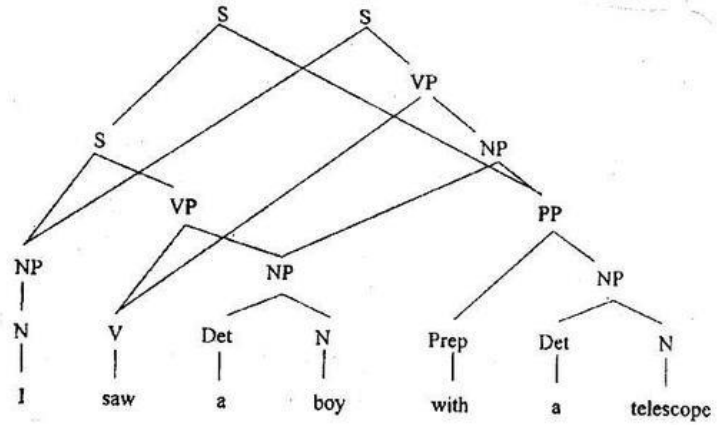


Introduction(3)

- Accurate model
  - The amounts of the training data (labeled data)
- Labeled data is expensive and hard to obtain
  - Microscopic image
  - Text parsing
  - …



(b)

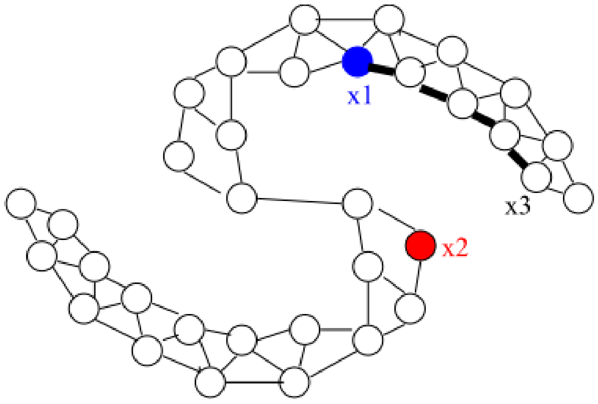


- How to build more accurate model with as few as labeled data?

Introduction(4)

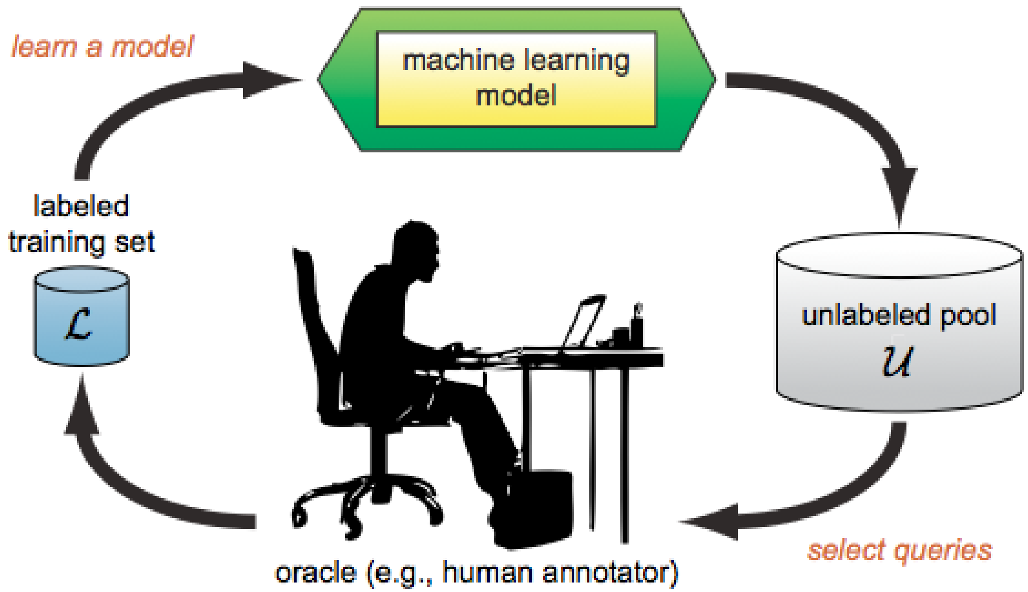
- Semi-supervised Learning (SSL)
  - Make use of unlabeled data to boost the performance of supervised learning.

- Graph-based Semi-supervised Learning (GB-SSL)
  - Use graph to approximate the “manifold structure” P(X), which is used to boost the conditional distribution P(Y|X)



Introduction(5)

- Active Learning
  - Learn a model in an interactive way, which is able to select the most representative data based on the model learned in each iteration.



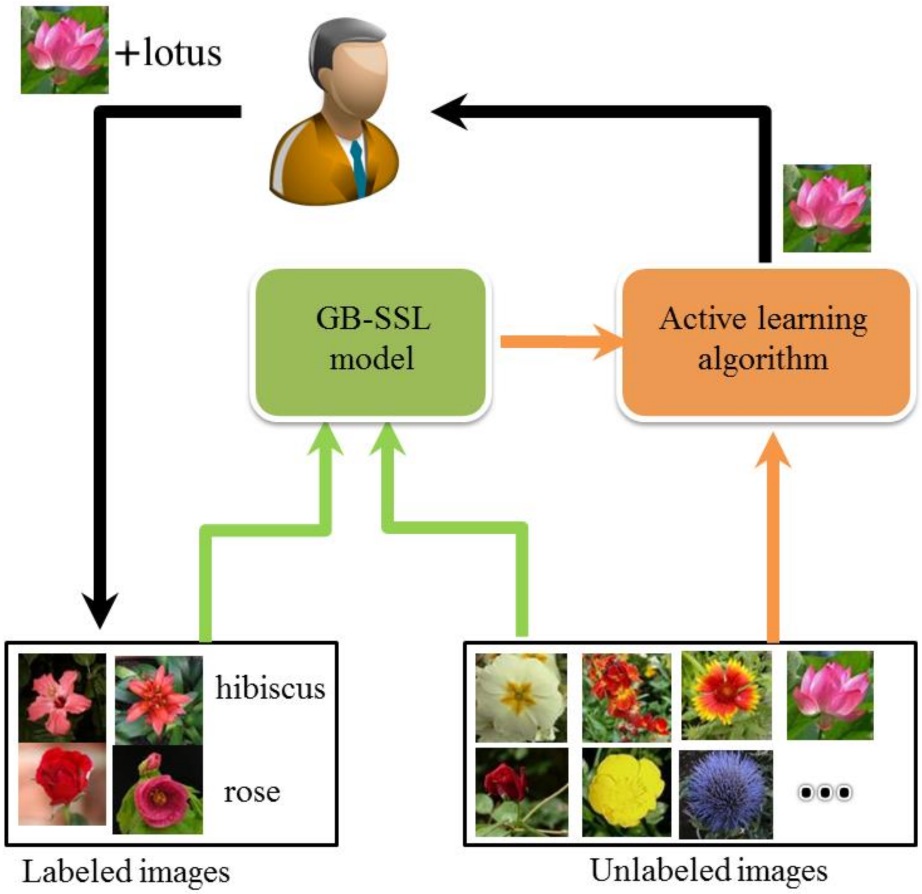
[Settles 2010]

outline

- Introduction
  - Learning-based automatic Annotation
  - Semi-Supervised Learning
  - Active learning
- Graph-based active semi-supervised learning framework
  - Graph-based semi-supervised learning
  - Active learning: Minimize expected global uncertainty
  - Real scene efficiency
    - Inductive extensity
    - Incrementally update
    - Sample candidate points
- Experiment
- conclusion

Graph-based active semi-supervised learning framework

- Workflow
  - Initialize the annotation model by using GB-SSL
  - Using the active learning algorithm to select the most informative examples to query the user for label
  - Update the model by incorporating the selected examples into training set



Graph based Semi-supervised Learning(1)

- Local and Global Consistency (LGC) , Zhou, NIPS 2003

  - **Notation**
    - A point set $\mathcal{X} = (\mathcal{X}_L, \mathcal{X}_U) = \{\mathbf{x}_1, \dots, \mathbf{x}_l, \mathbf{x}_{l+1}, \dots, \mathbf{x}_n\}$
    - Points $\mathcal{X}_L = \{\mathbf{x}_1, \dots, \mathbf{x}_l\}$ are labeled $y_i \in \mathcal{L} = \{1, \dots, c\}$
    - Predict the label of unlabeled points $\mathcal{X}_U = \{\mathbf{x}_{l+1}, \dots, \mathbf{x}_n\}$
  - **Method**

1. Form the affinity matrix $\mathbf{W}$ with its entries $w_{ij} = \exp(-||\mathbf{x}_i - \mathbf{x}_j||/2\sigma^2)$ if $i \neq j$ and $w_{ii} = 0$.
2. Construct the normalized Laplacian Matrix $\mathbf{S} = \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$, in which $\mathbf{D}$ is a diagonal matrix with its (i,i)-element equal to the sum of the i-th row of $\mathbf{W}$.
3. Iterate $\mathbf{F}(t+1) = \alpha\mathbf{S}\mathbf{F}(t) + (1-\alpha)\mathbf{Y}$ until convergence, where α is a parameter in (0,1). Let $\mathbf{F}^*$ denote the limit of sequence $\mathbf{F}(t)$, which has a closed solution form:

$$\mathbf{F}^* = \lim_{t\to\infty} \mathbf{F}(t) = (1-\alpha)(\mathbf{I} - \alpha\mathbf{S})^{-1}\mathbf{Y}$$

4. We can assign each point $\mathbf{x}_i \in \mathcal{X}_U$ with the label $y_i = \arg\max_{j \le c} \mathbf{F}^*_{ij}$

Active Learning: Minimize Expected Global Uncertainty (1)

- **The intuition:** For a certain unlabeled example, if we incorporate it along with its assumed label (It can be empirically evaluated by the current model predictor) and re-train the model, which can make the new predictor has minimal uncertainty for the other unlabeled examples.

- Howe to measure the uncertainty of unlabeled examples
  - Entropy (Y_i)
  - Using **F** to approximate $P(Y_U|X)$
  - The global uncertainty can be calculated as:

$$H(\mathbf{F}) = \sum_{i=1}^{n} H(Y_i) = -\sum_{i=1}^{n}\sum_{j=1}^{c} \mathbf{F}_{ij} \log_2 \mathbf{F}_{ij}$$

Active Learning: Minimize Expected Global Uncertainty (2)

- If we select an unlabeled example $\mathbf{x}_k$ to query the oracle and we receive the assumed label y_k, adding $(\mathbf{x}_k, y_k)$ to the training set and retraining , we will get the new predictor $\mathbf{F}^{+(\mathbf{x}_k, y_k)}$

$$H(\mathbf{F}^{+(\mathbf{x}_k, y_k)}) = -\sum_{i=1}^{n}\sum_{j=1}^{c} \mathbf{F}_{ij}^{+(\mathbf{x}_k, y_k)} \log_2 \mathbf{F}_{ij}^{+(\mathbf{x}_k, y_k)} \quad (2)$$

- In fact, we don't know the true label y_k before we query the oracle. So we empirically assume the label y_k=j is given with the probability $\mathbf{F}_{kj}$. Hence the expected global uncertainty is:

$$H(\mathbf{F}^{+\mathbf{x}_k}) = \sum_{j=1}^{c} \mathbf{F}_{kj} H(\mathbf{F}^{+(\mathbf{x}_k, j)}) \quad (3)$$

- We greedily select the example $\mathbf{x}_k$ that minimizes the expected global uncertainty to query the oracle, which can be calculated as:

$$\mathbf{x}_k = \arg\min_{\mathbf{x}_{k'} \in \Omega_U} H(\mathbf{F}^{+\mathbf{x}_{k'}}) \quad (4)$$

Active Learning: Minimize Expected Global Uncertainty (3)

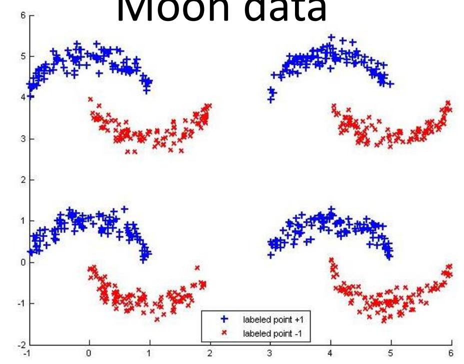
Algorithm 1 Minimize Expected Global Uncertainty

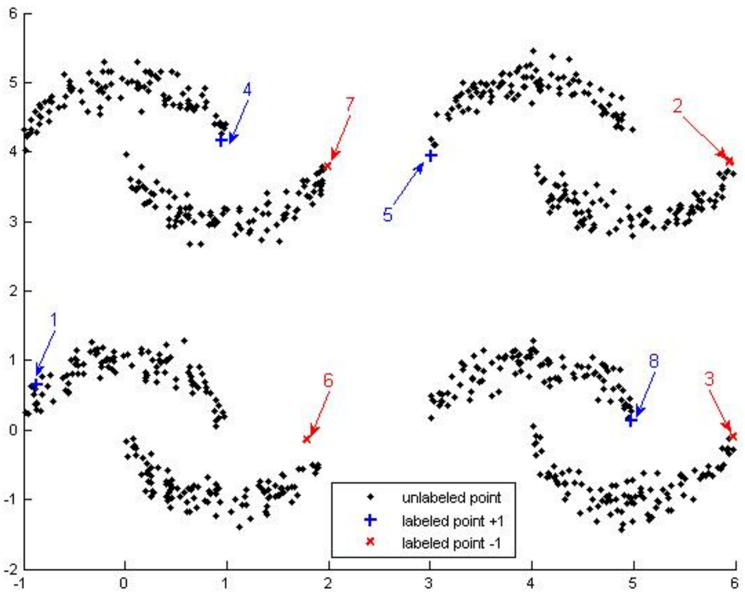
Input: Ω_L, Ω_U, normalized Laplacian Matrix $\mathbf{S}$;
Initialize $\mathbf{F}$ using formula (1);
for each round k **do**
  for each example $\mathbf{x}_{k'} \in \Omega_U$ **do**
    for each possible label $j \in \{1, 2, ...c\}$ **do**
      Compute $\mathbf{F}^{+(\mathbf{x}_{k'}, j)}$ with $\Omega_L \cup \{(\mathbf{x}_{k'}, j)\}$
      Compute $H(\mathbf{F}^{+(\mathbf{x}_{k'}, j)})$ using formula (2)
    end for
    Compute $H(\mathbf{F}^{+\mathbf{x}_{k'}})$ using formula (3)
  end for
  Find $\mathbf{x}_k$ based on (4)
  Query $\mathbf{x}_k$ for label y_k
  Add $(\mathbf{x}_k, y_k)$ to Ω_L, remove $\mathbf{x}_k$ from Ω_U
  Update $\mathbf{F}$ with the new Ω_L
end for
Output: Ω_L and $\mathbf{F}$.

Active Learning: Minimize Expected Global Uncertainty (4)

- How does the proposed active Learning method works?
  - tend to select the examples which is more representative.

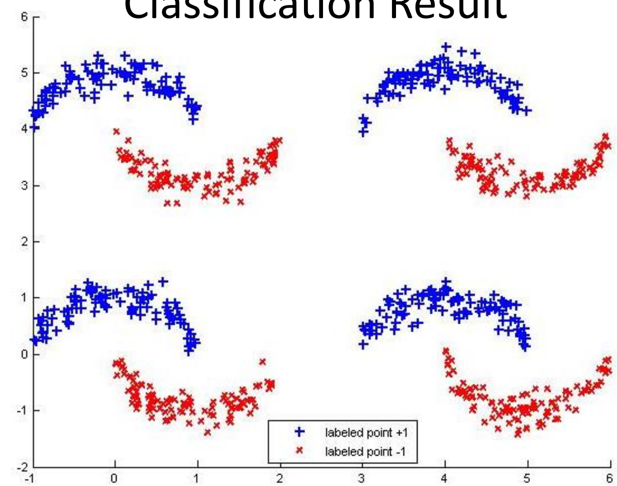
Moon data





Ideally classification with 8 labeled data

Classification Result

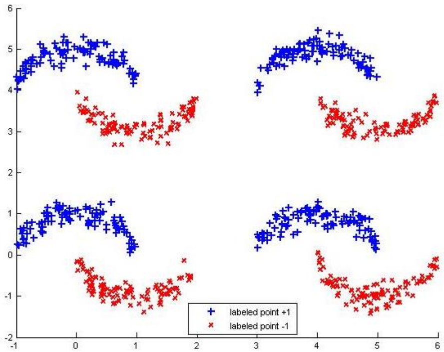


Active Learning: Minimize Expected Global Uncertainty (5)

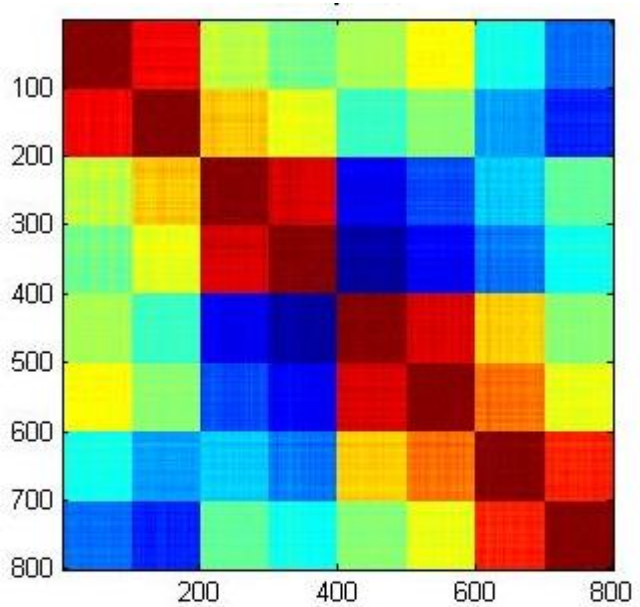
- Why does the proposed active Learning method works?

  - $\mathbf{F} = \mathbf{W}_{LGC}\mathbf{Y}$,
    where $\mathbf{W}_{LGC} = (\mathbf{I} - \alpha\mathbf{S})^{-1}$

Moon data

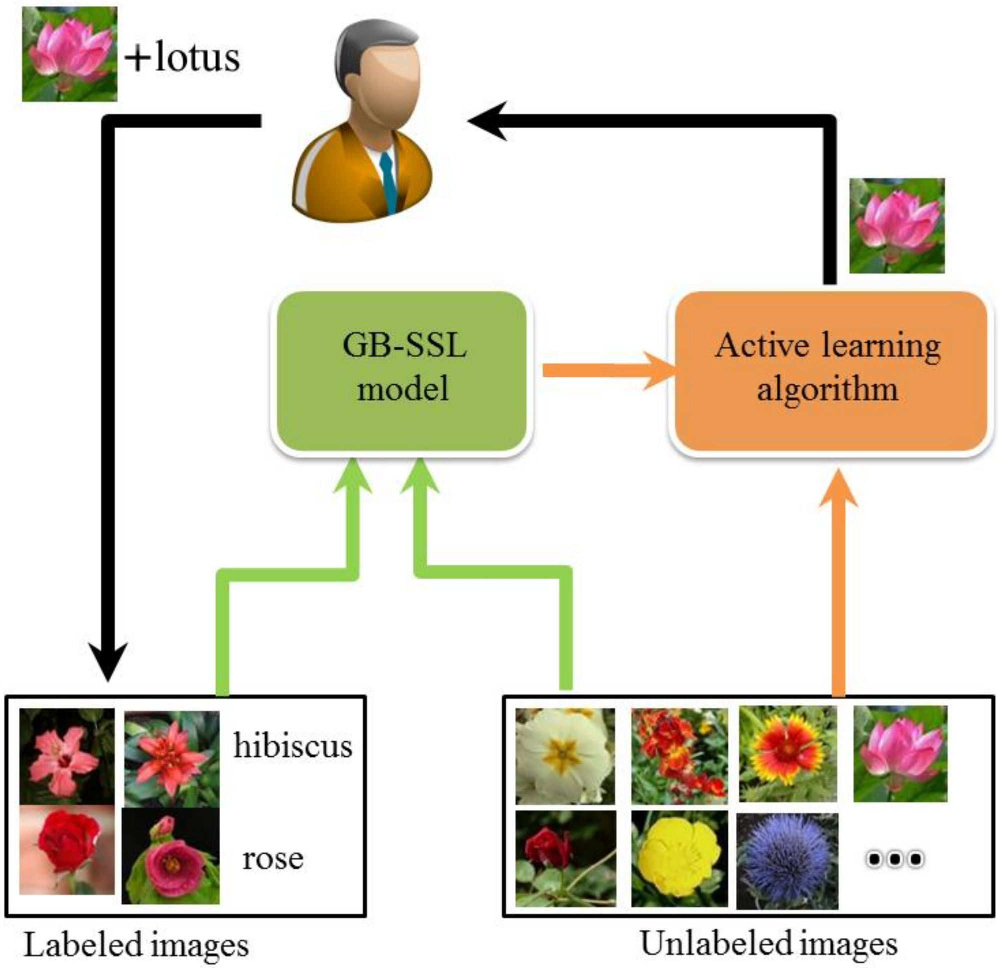


Log ($\mathbf{W}_{LGC}$)



Efficiency: For Real-world Scenario

- For Real-world Scenario
  - Incrementally update
  - Sample Candidate points



Efficiency: For Real-world Scenario(2)

- Incrementally update
  - Problem: use formula (1) to update **F** or $\mathbf{F}^{+}$**:** O(n^3)
  - Method: decomposed formulation
    - New example $\mathbf{x}_k$ with label $y_k = j$
      - ✓ $\mathbf{F}^{+} = \mathbf{F} + \mathbf{T}\mathbf{e}_k \cdot \mathbf{e}_j^T$, where $\mathbf{T} = (1-\alpha)(\mathbf{I} - \alpha\mathbf{S})^{-1}$

        $\Delta\mathbf{F}_{.j} = \mathbf{T}_{.k}$
  - The time complexity is O(*n)*

Efficiency: For Real-world Scenario(3)

- Sample Candidate points
  - In each Iteration, the complexity of the algorithm : $O(c^2 \cdot n^2)$
  - Sample the candidate point set with size *m (m<<n)*.
  - The reduced complexity: $O(c^2 \cdot m \cdot n)$, linear to n

outline

- Introduction
  - Learning-based automatic Annotation
  - Semi-Supervised Learning
  - Active learning
- Graph-based active semi-supervised learning framework
  - Graph-based semi-supervised learning
  - Active learning: Minimize expected global uncertainty
  - Real scene efficiency
    - Inductive extensity
    - Incrementally update
    - Sample candidate points
- Experiment
- conclusion

Experiments(1): setup

- Baselines
  - Random example selection (Random)
  - Maximize Entropy-Based (MEB)
  - Best-versus-Second-Best (BvSB, Joshi, CVPR 2009)
  - Minimize the Risk (Risk, Zhu, ICML 2004)
- Evaluation criteria
  - Accuracy
- Datasets
  - USPS : 4000, 10 classes, 256D Pixel
  - Flower-102: 1963, 12 classes, 1500D Bow
  - MNIST: 70000, 10 classes, 784D Pixel

Experiments(2): accuracy comparison

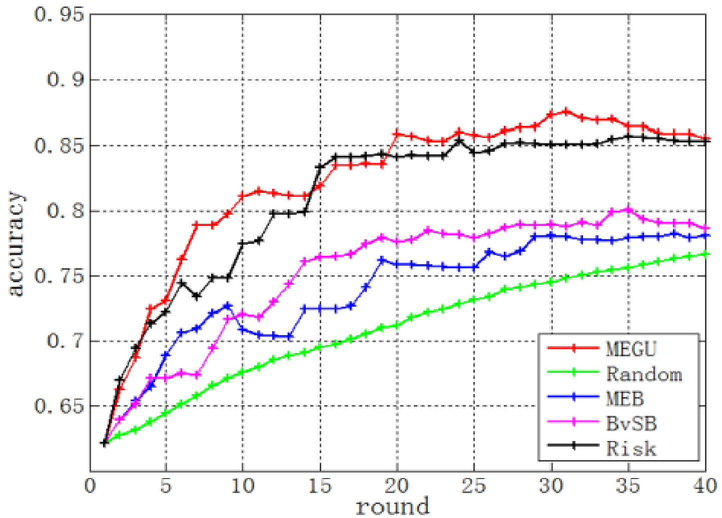


Fig. 1 Accuracy on USPS

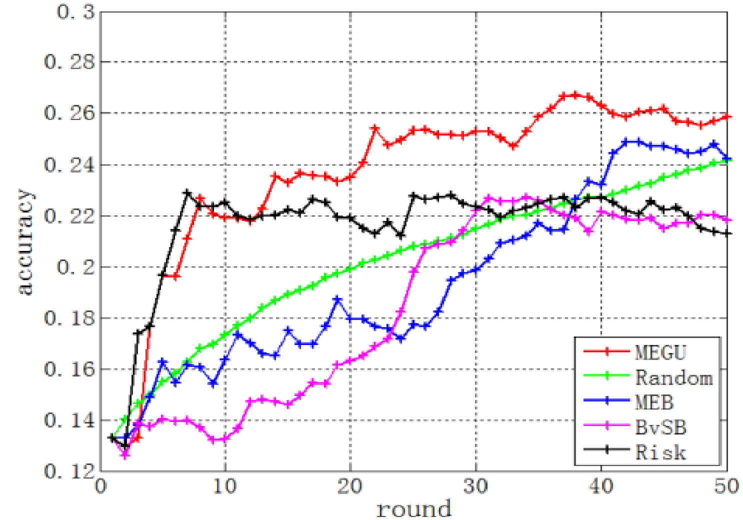


Fig. 2 Accuracy on Flower-102

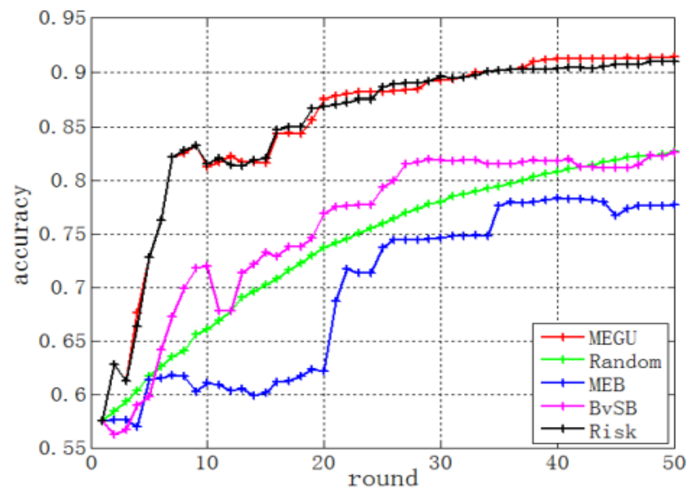


Fig. 3 Accuracy on MNIST

Experiments(3)

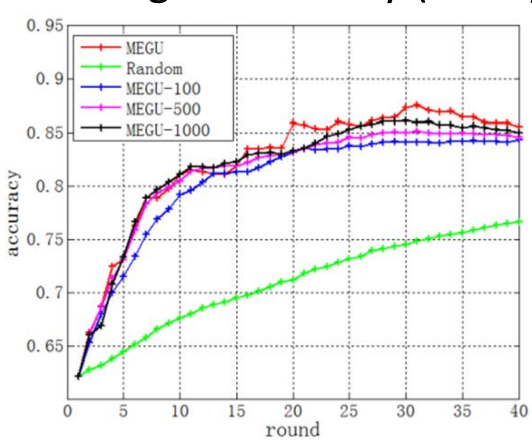
- Reduction in annotation

Tab. 1 Quantitative comparison

accuracy	dataset	#MEGU	#Random
80%	USPS	19	75
85%	USPS	40	154
80%	MNIST	16	46
85%	MNIST	25	75
90%	MNIST	42	178

- Sample candidate points

Fig. 4 Accuracy (USPS)



Tab. 2 time cost (USPS)

method	time cost (s)
MEGU-100	2.5
MEGU-500	12.8
MEGU-1000	25.7
MEGU	103
Random	0.024

Conclusion

- Propose a novel graph-based active semi-supervised learning framework which can learn a multi-class model efficiently with minimal human labor

- Propose Minimize Expected Global Uncertainty (MEGU) algorithm to actively select example, which naturally combine the probabilistic outputs of GB-SSL methods

- propose an incremental model updating method, which has the time complexity of O(n), compared to the original re-training of O(n^3).

Thank you !