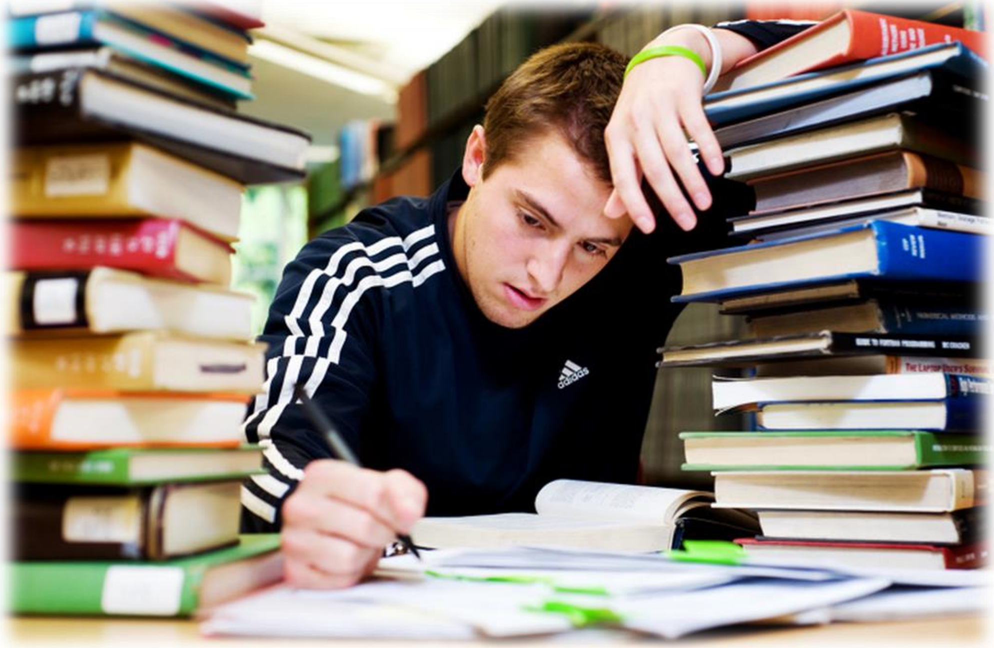


Semi-Supervised Learning: Self-Training and Co-Training

Pilsung Kang

School of Industrial Management Engineering

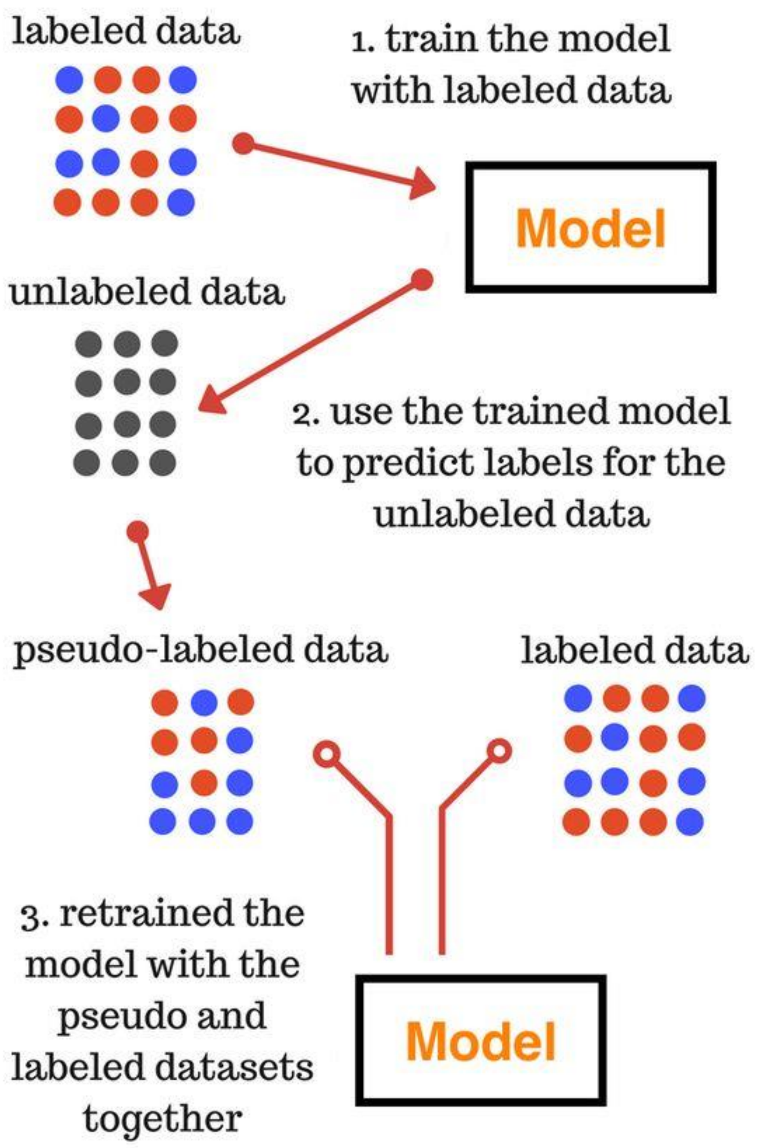
Korea University

Self-Training

- Assumption in the self-training
  - ✓ One's own high confidence predictions are correct
- Basic self-training algorithm
  - ✓ Train from f $(\mathbf{X}_l, \mathbf{y}_l)$
  - ✓ Predict on $\mathbf{x} \in \mathbf{X}_u$
  - ✓ Add $(\mathbf{x}, f(\mathbf{x}))$ to labeled data
  - ✓ Repeat
- Variations in self-training
  - ✓ Add a few most confident $(\mathbf{x}, f(\mathbf{x}))$ to labeled data
  - ✓ Add all $(\mathbf{x}, f(\mathbf{x}))$ to labeled data
  - ✓ Add all $(\mathbf{x}, f(\mathbf{x}))$ to labeled data, weigh each by confidence

Self-Training

- Procedure



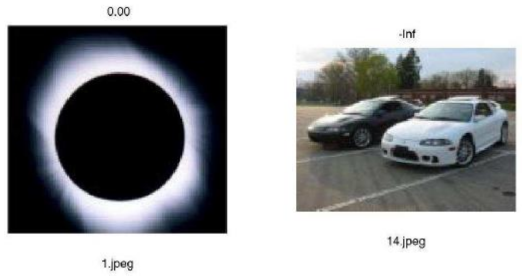
https://www.pinterest.co.kr/pin/540713499003163569/

Self-Training: Example 1

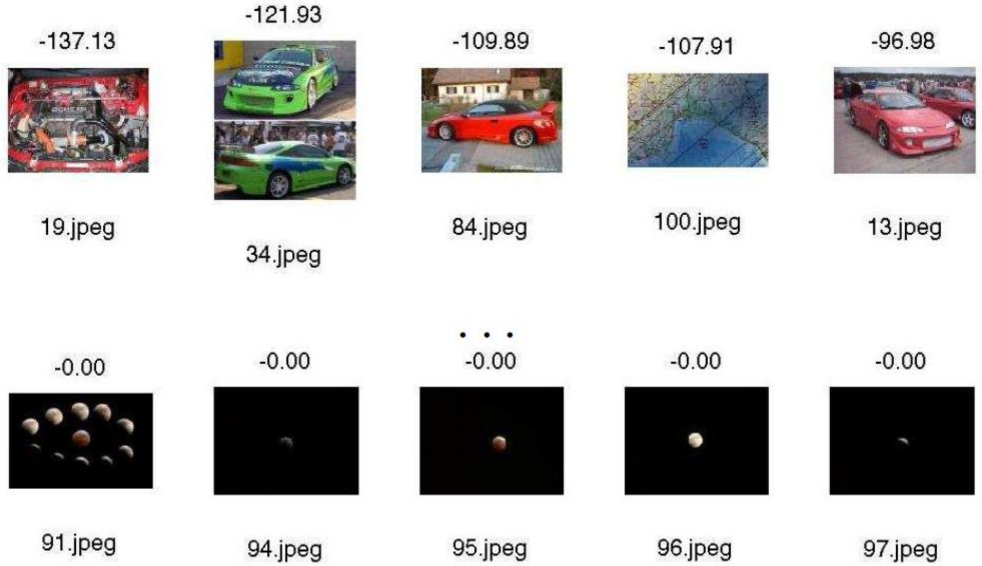
Zhu (2007)

- Image Categorization

1. Train a naïve Bayes classifier on the two initial labeled images



2. Classify unlabeled data, sort by confidence $\log p(y = \text{astronomy} \mid x)$

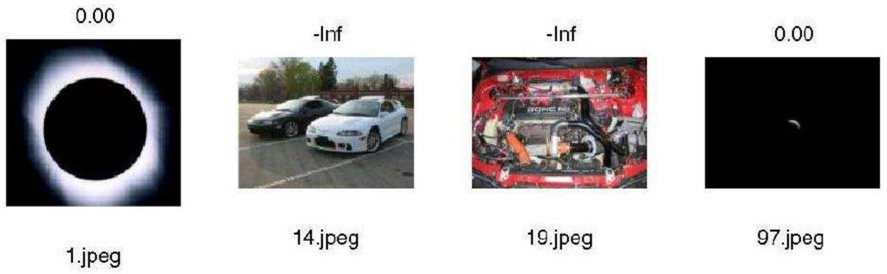


Self-Training: Example 1

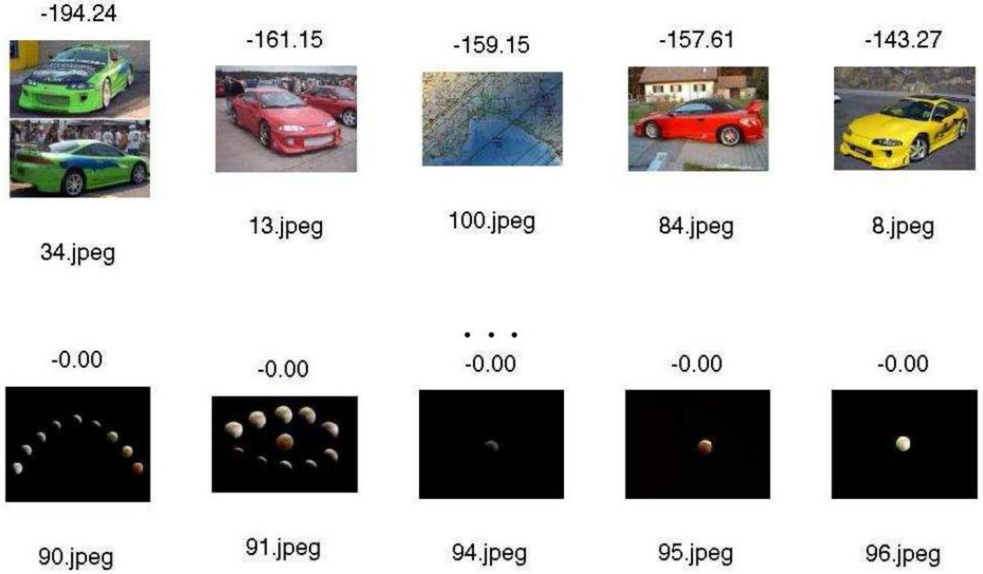
Zhu (2007)

- Image Categorization

3. Add the most confident images and predicted labels to labeled data



4. Re-train the classifier and repeat



Self-Training: Example 2

Zhu (2009)

- Propagating 1-Nearest Neighbor

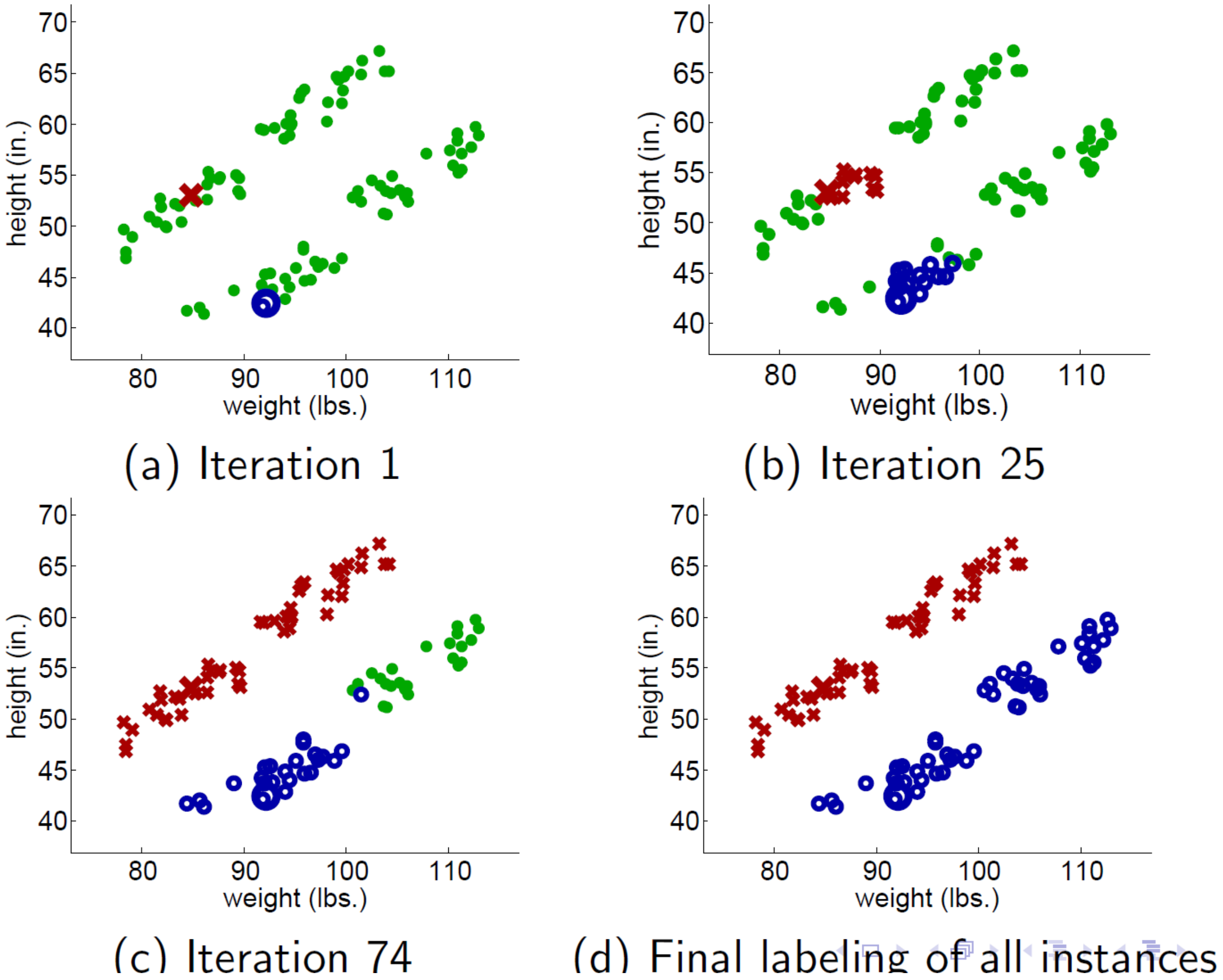
Input: labeled data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, unlabeled data $\{\mathbf{x}_j\}_{j=l+1}^{l+u}$, distance function $d()$.

1. Initially, let $L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ and $U = \{\mathbf{x}_j\}_{j=l+1}^{l+u}$.
2. Repeat until U is empty:
3. Select $\mathbf{x} = \mathrm{argmin}_{\mathbf{x} \in U} \min_{\mathbf{x}' \in L} d(\mathbf{x}, \mathbf{x}')$.
4. Set $f(\mathbf{x})$ to the label of $\mathbf{x}$'s nearest instance in L. Break ties randomly.
5. Remove $\mathbf{x}$ from U; add $(\mathbf{x}, f(\mathbf{x}))$ to L.

Self-Training: Example 2

Zhu (2009)

- Propagating 1-Nearest Neighbor



(a) Iteration 1

(b) Iteration 25

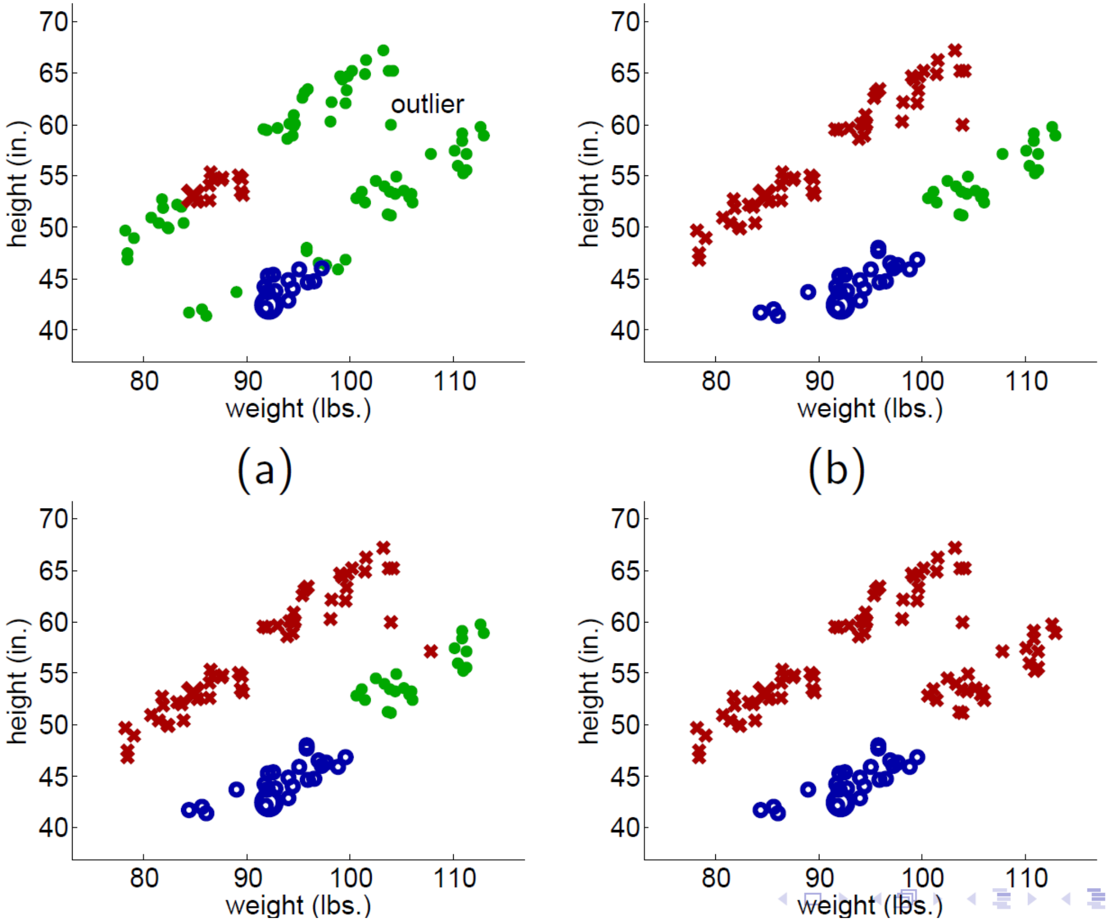
(c) Iteration 74

(d) Final labeling of all instances

Self-Training: Example 2

Zhu (2009)

- Propagating 1-Nearest Neighbor (with a single outlier)



Self-Training: Summary

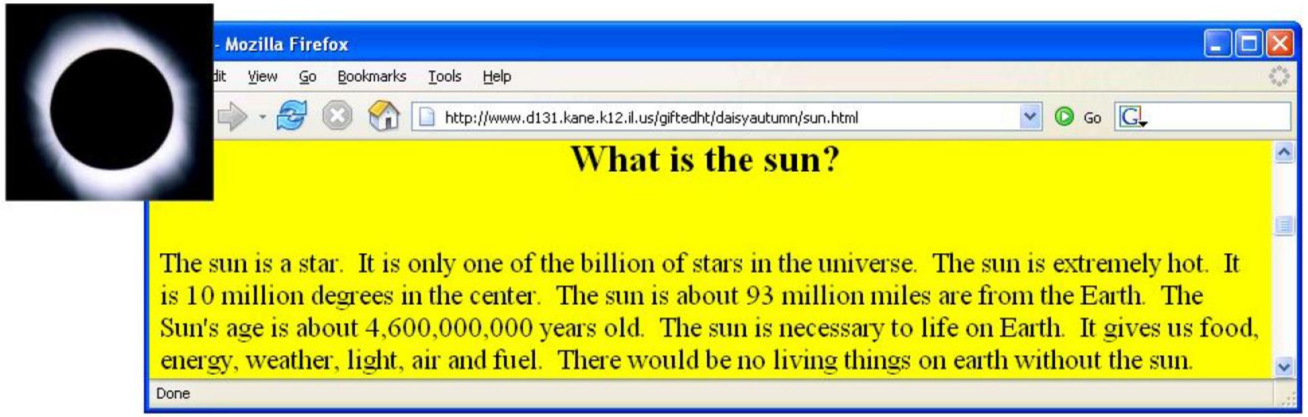
- Advantages
  - ✓ The simplest semi-supervised learning method
  - ✓ A wrapper method, applies to existing (complex) classifiers
  - ✓ Often used in real tasks like natural language processing

- Disadvantages
  - ✓ Early mistakes could reinforce themselves
  - ✓ Cannot say too much in terms of convergence

Multi-view Algorithm: Co-Training

Blum and Mitchell (1998), Yu et al. (2011)

- Co-training
  - ✓ Two views of an item: image and HTML text



Multi-view Algorithm: Co-Training

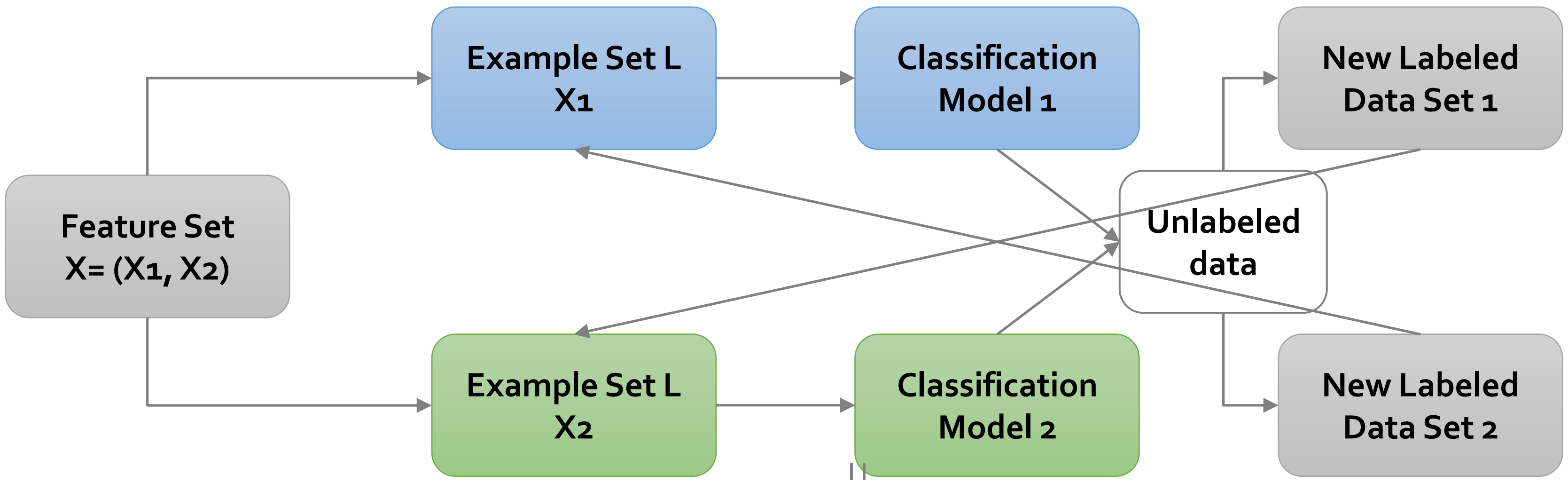
- Feature split

Each instance is represented by two sets of features $x = [x^{(1)}; x^{(2)}]$

- $x^{(1)}$ = image features
- $x^{(2)}$ = web page text
- This is a natural feature split (or multiple views)

Co-training idea:

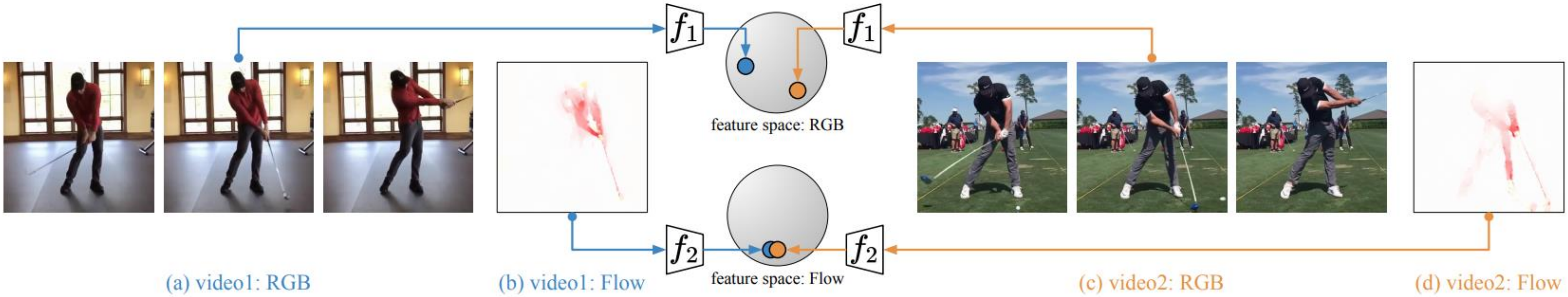
- Train an image classifier and a text classifier
- The two classifiers teach each other



Multi-view Algorithm: Co-Training

Han et al. (2020)

- Feature split



Multi-view Algorithm: Co-Training

Blum and Mitchell (1998)

- Co-Training (Basic algorithm, for classification with Naïve Baye's Classifier)
  - ✓ Given
    - A set L of labeled training examples
    - A set of U of unlabeled examples
  - ✓ Create a pool U' of example by choosing u examples at random from U
  - ✓ Loop for k iterations
    - Use L to train a classifier h_1 that considers only the x_1 portion of x
    - Use L to train a classifier h_2 that considers only the x_2 portion of x
    - Allow h_1 to label p positive and n negative examples from U'
    - Allow h_2 to label p positive and n negative examples from U'
    - Add these self-labeled examples to L
    - Randomly choose 2p+2n examples from U to replenish U'

Multi-view Algorithm: Co-Training

Zhou and Li (2005)

- Co-Training (with k-NN regression)

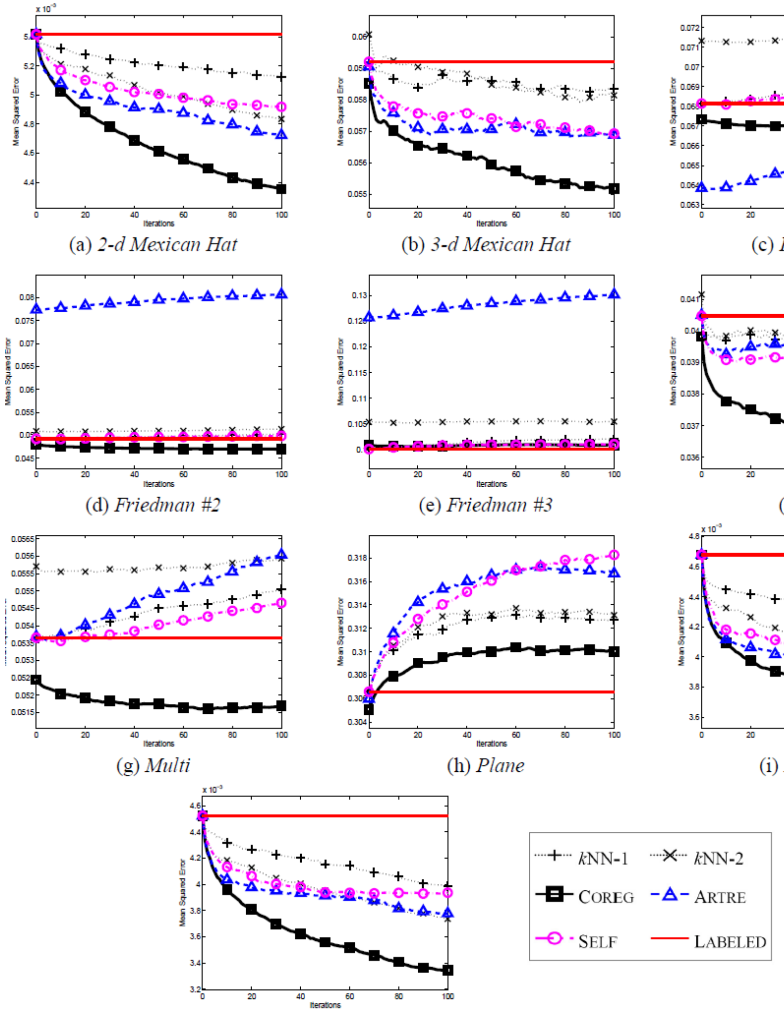
ALGORITHM: COREG

INPUT: labeled example set L, unlabeled example set U,
number of nearest neighbors k,
maximum number of learning iterations T,
distance orders p_1, p_2

PROCESS:

$L_1 \leftarrow L$; $L_2 \leftarrow L$
Create pool U' by randomly picking examples from U
$h_1 \leftarrow kNN(L_1, k, p_1)$; $h_2 \leftarrow kNN(L_2, k, p_2)$
Repeat for T rounds:
  for $j \in \{1, 2\}$ **do**
    for each $\mathbf{x}_u \in U'$ **do**
      $\hat{\mathbf{y}}_u \leftarrow h_j(\mathbf{x}_u)$
      $\Omega \leftarrow Neighbors(\mathbf{x}_u, k, L_j)$
      $h'_j \leftarrow kNN(L_j \cup \{(\mathbf{x}_u, \hat{\mathbf{y}}_u)\}, k, p_j)$
      $\Delta_{\mathbf{x}_u} \leftarrow \sum_{\mathbf{x}_i \in \Omega} \left((\mathbf{y}_i - h_j(\mathbf{x}_i))^2 - (\mathbf{y}_i - h'_j(\mathbf{x}_i))^2 \right)$
    end of for
    if there exists an $\Delta_{\mathbf{x}_u} > 0$
    then $\tilde{\mathbf{x}}_j \leftarrow \arg\max_{\mathbf{x}_u \in U'} \Delta_{\mathbf{x}_u}$; $\tilde{\mathbf{y}}_j \leftarrow h_j(\tilde{x}_j)$
      $\pi_j \leftarrow \{(\tilde{\mathbf{x}}_j, \tilde{\mathbf{y}}_j)\}$; $U' \leftarrow U' - \pi_j$
    else $\pi_j \leftarrow \emptyset$
  end of for
  $L_1 \leftarrow L_1 \cup \pi_2$; $L_2 \leftarrow L_2 \cup \pi_1$
  if neither of L_1 and L_2 changes **then exit**
  else
    $h_1 \leftarrow kNN(L_1, k, p_1)$; $h_2 \leftarrow kNN(L_2, k, p_2)$
    Replenish U' by randomly picking examples from U
end of Repeat

OUTPUT: regressor $h^*(\mathbf{x}) \leftarrow \frac{1}{2}(h_1(\mathbf{x}) + h_2(\mathbf{x}))$

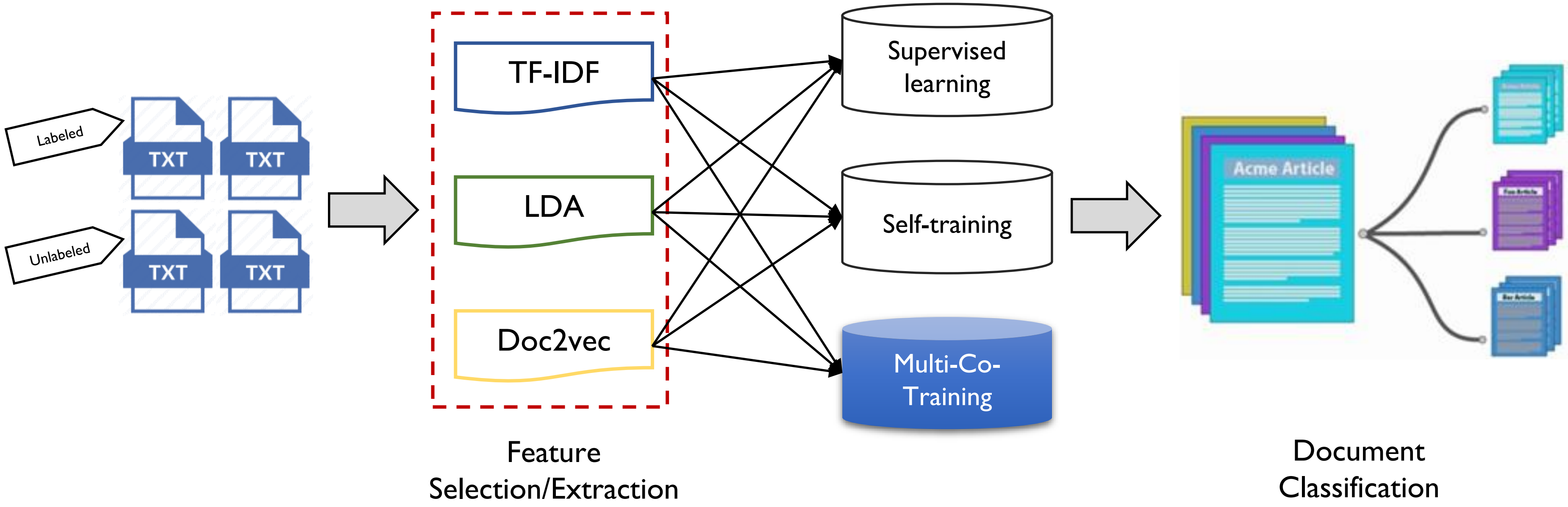


(a) 2-d Mexican Hat
(b) 3-d Mexican Hat
(c) Friedman #1
(d) Friedman #2
(e) Friedman #3
(f) Gabor
(g) Multi
(h) Plane
(i) Polynomial
(j) SinC
Legend

Multi-view Algorithm: Co-Training

Kim et al. (2019)

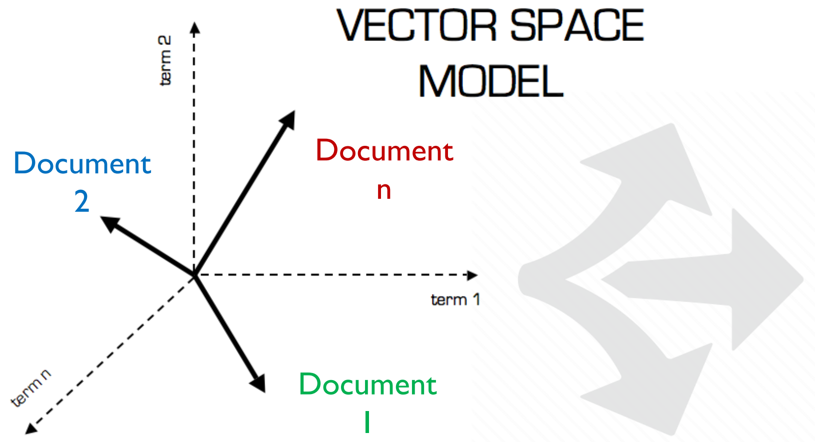
- Multi-Co-Training for Text Classification



Multi-view Algorithm: Co-Training

Kim et al. (2019)

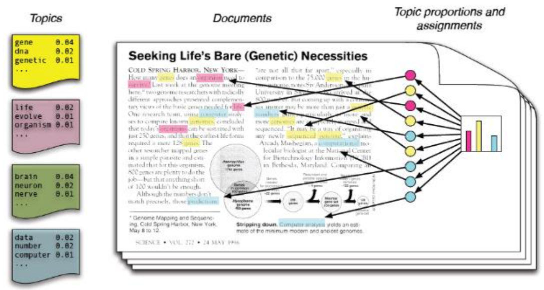
- Multi-Co-Training for Text Classification



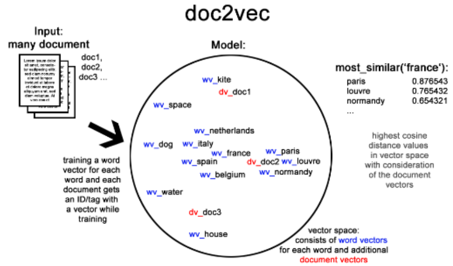
TF-IDF

Document1					
Word	TF	N	DF	IDF	TF-IDF
This	4	3	3	1/3	1.33
is	5	2	3	1/3	1.67
an	3	3	3	1/3	1
Example	2	3	1	1	2

LDA



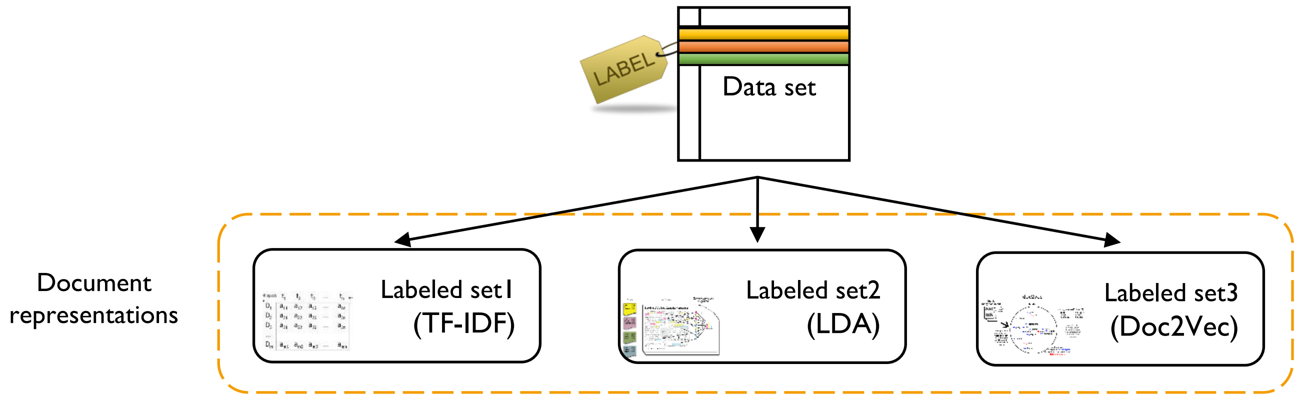
Doc2vec



Multi-view Algorithm: Co-Training

Kim et al. (2019)

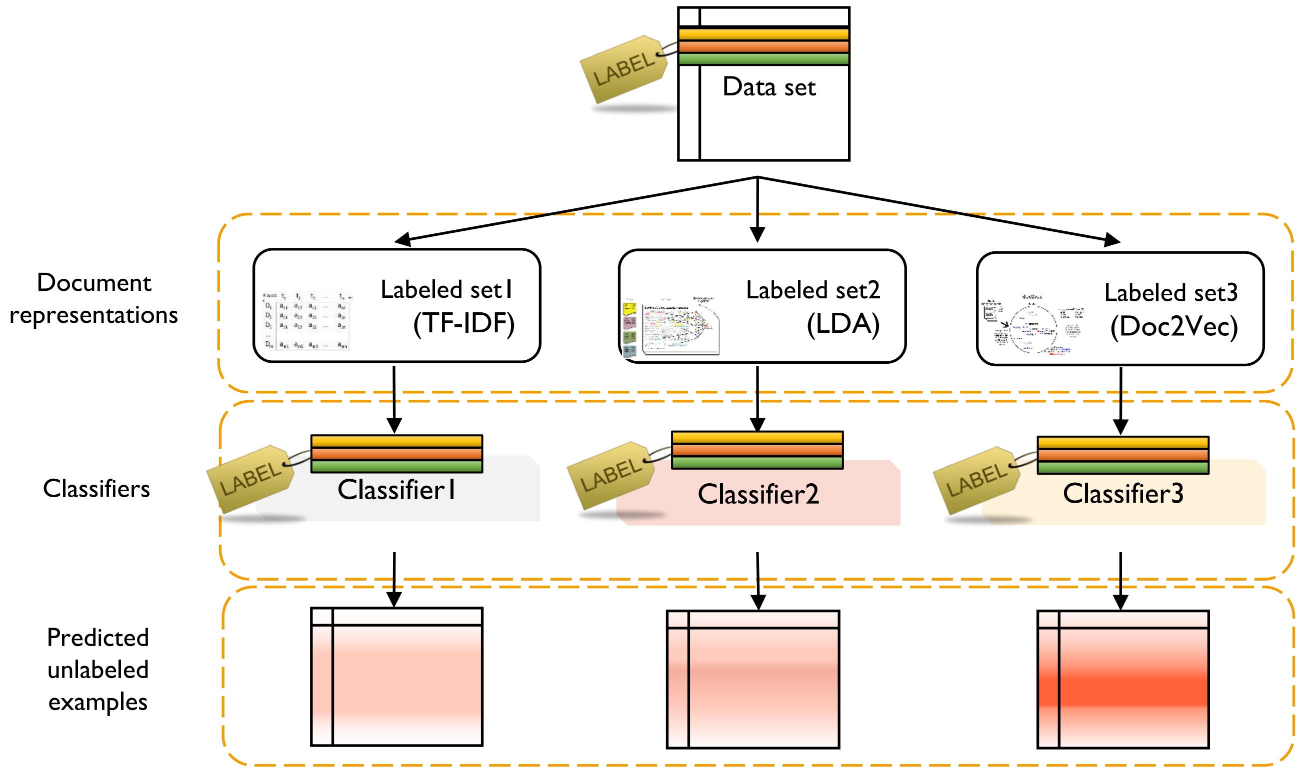
- Multi-Co-Training for Text Classification
  - ✓ Step1) Create multi-views: TF-IDF, LDA and Doc2vec



Multi-view Algorithm: Co-Training

Kim et al. (2019)

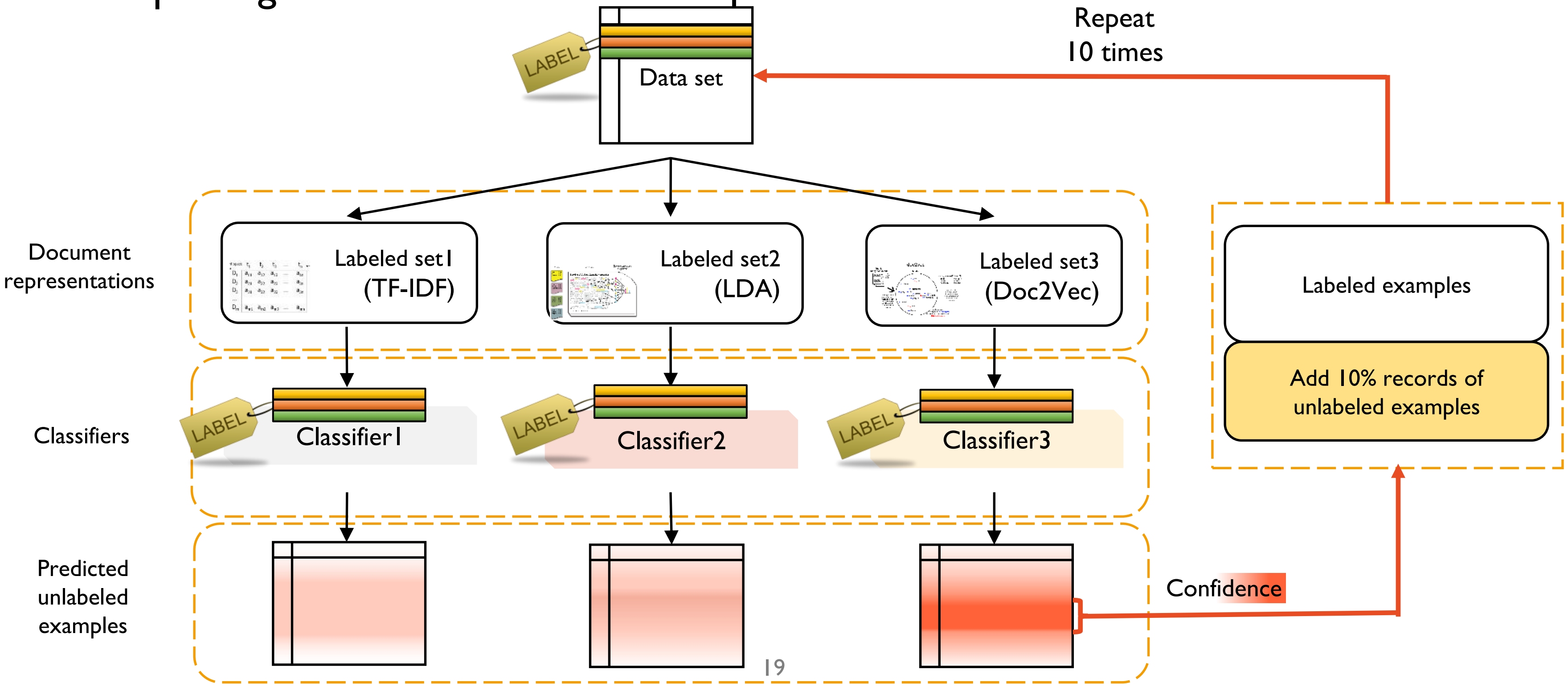
- Multi-Co-Training for Text Classification
  - ✓ Step1) Create multi-views: TF-IDF, LDA and Doc2vec
  - ✓ Step2) Build models and then predict unlabeled examples



Multi-view Algorithm: Co-Training

Kim et al. (2019)

- Multi-Co-Training for Text Classification
  - ✓ Step1) Create multi-views: TF-IDF, LDA and Doc2vec
  - ✓ Step2) Build models and then predict unlabeled examples
  - ✓ Step3) Add the predicted examples with high confidence to labeled examples
  - ✓ Step4) Continue repeating it until all unlabeled examples are annotated



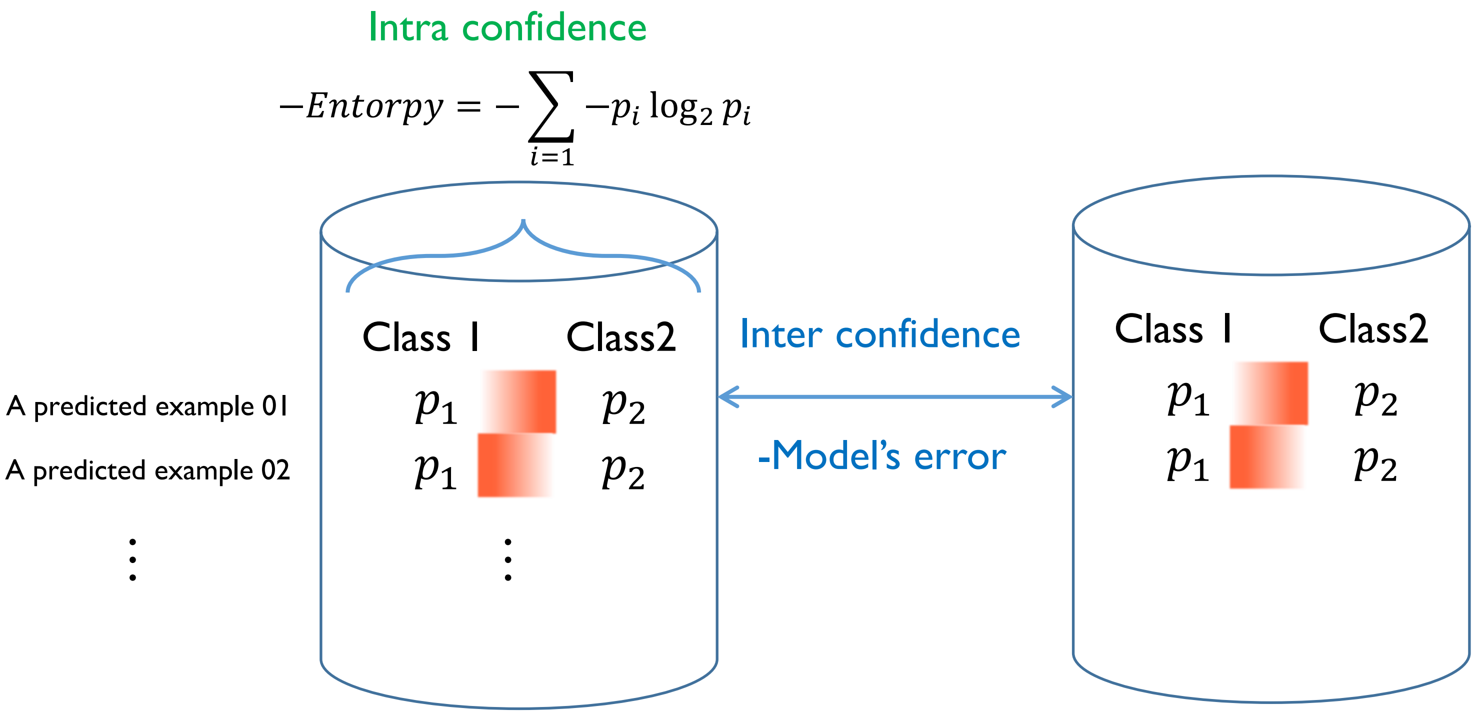
고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Multi-view Algorithm: Co-Training

Kim et al. (2019)

- Confidence measure with Naïve Bayesian
  - ✓ Intra confidence: $-Entropy$
  - ✓ Inter confidence: $-Training\ error$
  - ✓ Confidence measure = $-Entropy \times -Training\ error$



Multi-view Algorithm: Co-Training

Kim et al. (2019)

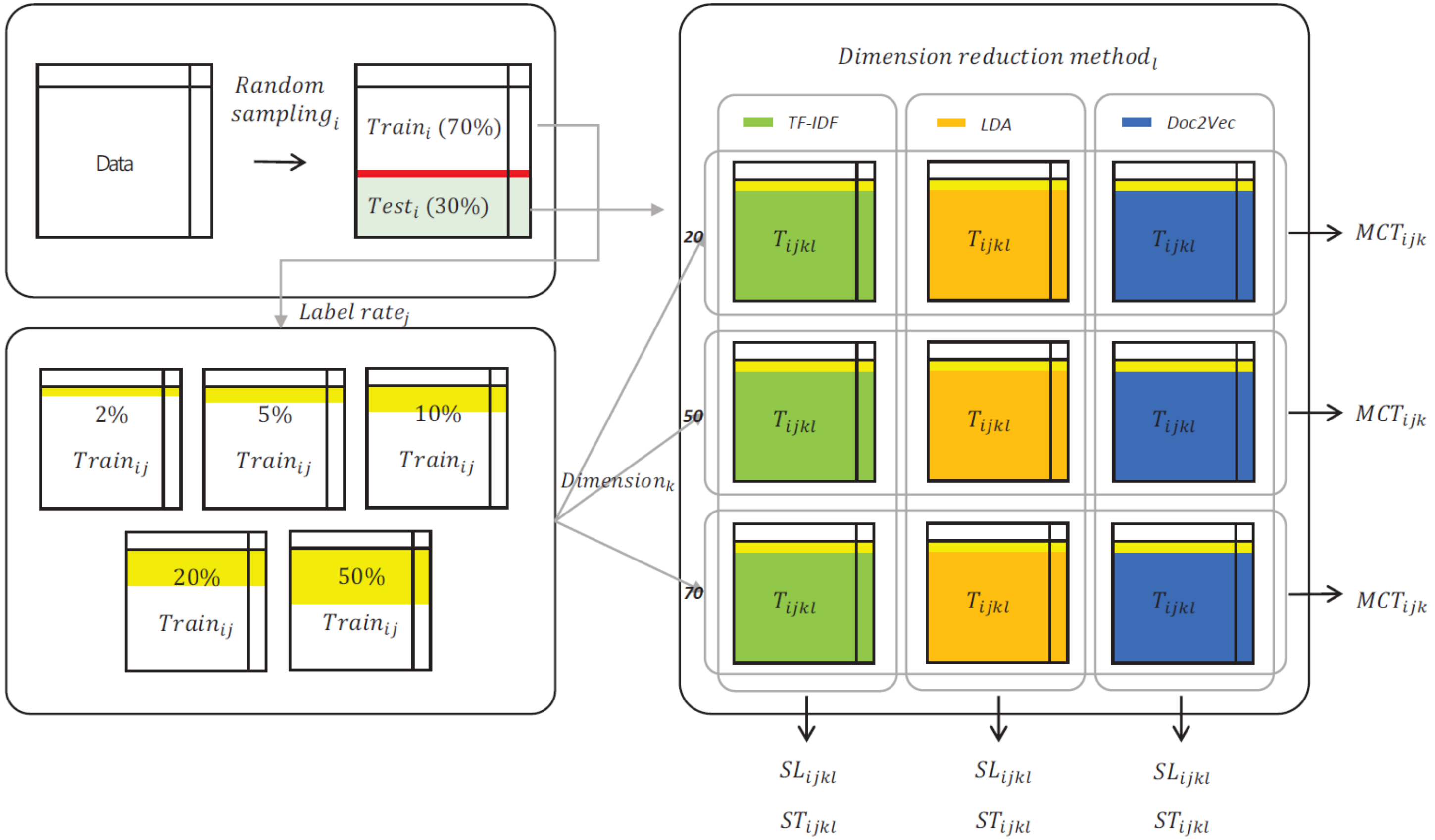
- Experiment: Data sets

Data	Description	Category	No. of documents
Economic	Whether a news article data is associated with the US economy	No : 6,458 (82.12%) Yes : 1,406 (17.88%)	7,864
20 Newsgroup	Data of 20,000 messages collected from 20 different news categories	Computer : 4,863 (30.40%) Recreation : 3,957 (24.74%) Science : 3,933 (24.59%) Talk : 3,244 (20.27%)	15,997
Ohsumed	Article-related abstracts of medical data	C04 : 2,630 (50.77%) C14 : 2,550 (49.23%)	5,180
Reuters	21,578 documents obtained from the Reuters news data	Earn : 3,953 (51.67%) Non-earn : 4,697 (48.33%)	8,650

Multi-view Algorithm: Co-Training

Kim et al. (2019)

- Experiment: Evaluation procedure
  - ✓ Evaluate the average and its standard deviation of Balanced Classification Rate (BCR)



Multi-view Algorithm: Co-Training

Kim et al. (2019)

- Experiment: Results

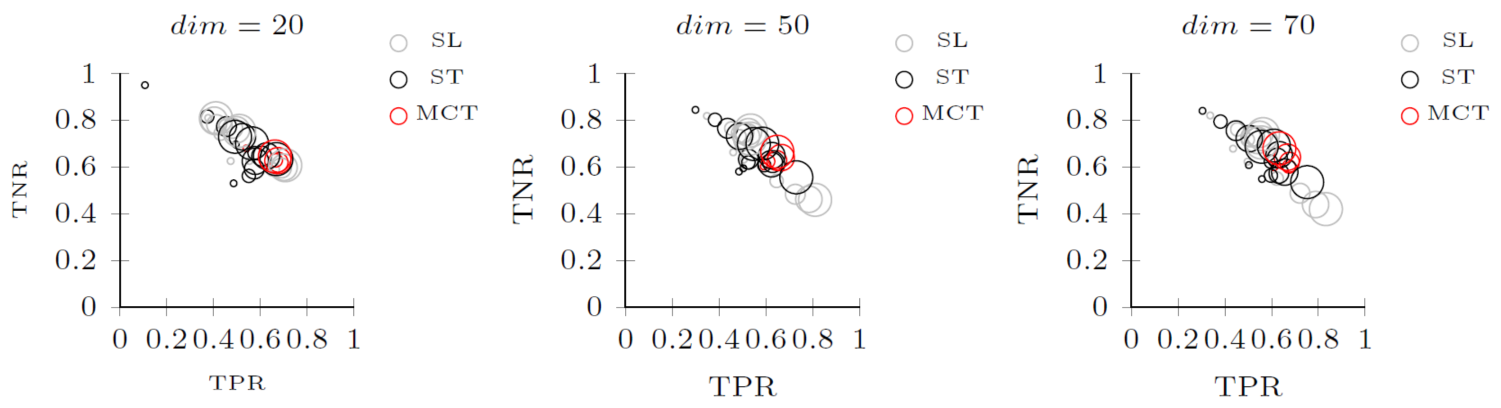


Fig. 7 TPR—TNR plots for SL, ST, and MCT for Economic dataset (Size = label(%))

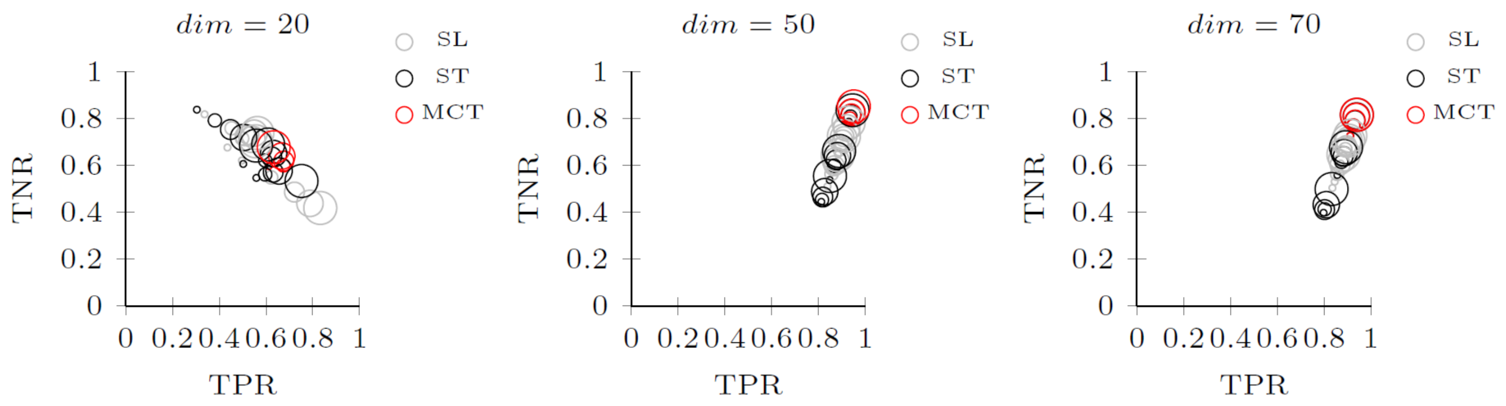


Fig. 8 TPR—TNR plots for SL, ST, and MCT for Newsgroup dataset (Size = label(%))

Multi-view Algorithm: Co-Training

Kim et al. (2019)

- Experiment: Results

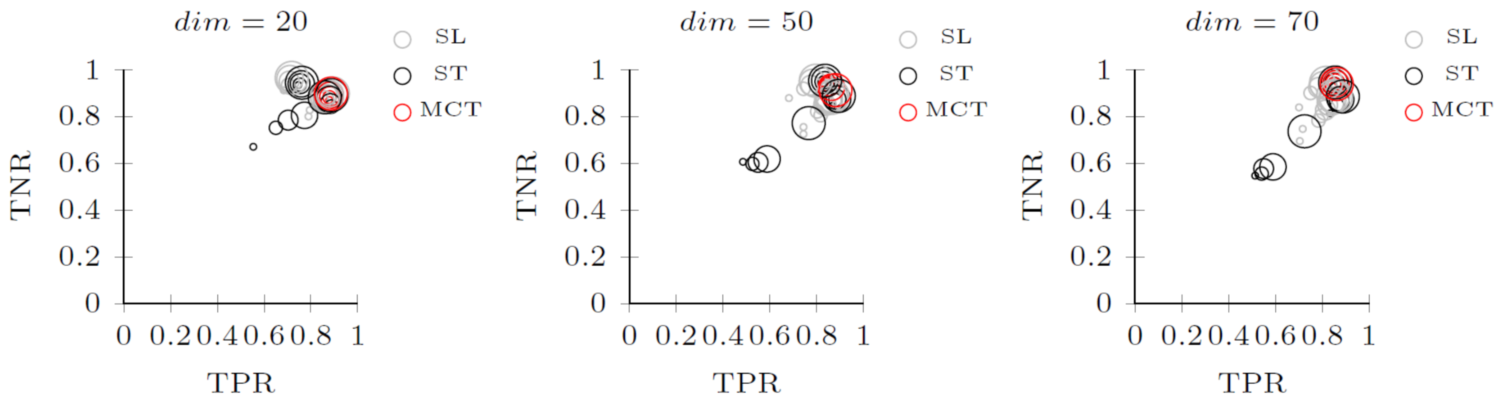


Fig. 9 TPR—TNR plots for SL, ST, and MCT for Ohsumed dataset (Size = label(%))

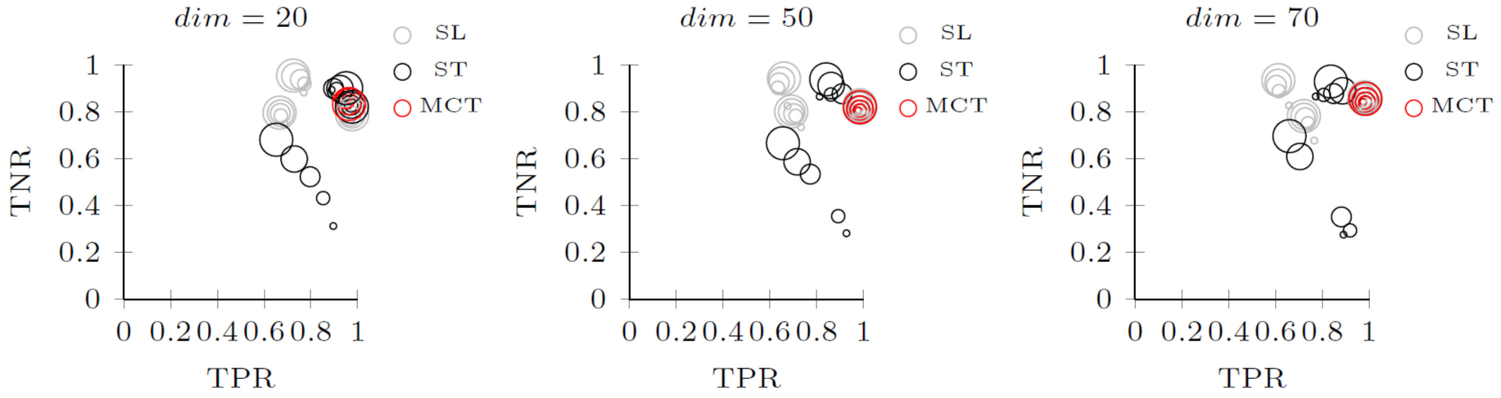


Fig. 10 TPR—TNR plots of SL, ST, and MCT for Reuters dataset (Size = label(%))


Q & A

References

Research Papers

- Ben-David, S., Lu, T., & Pál, D. (2008, July). Does Unlabeled Data Provably Help? Worst-case Analysis of the Sample Complexity of Semi-Supervised Learning. In *COLT* (pp. 33-44).
- Bennett, K., & Demiriz, A. (1999). Semi-supervised support vector machines. *Advances in Neural Information processing systems*, 368-374.
- Blum, A., & Mitchell, T. (1998, July). Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory* (pp. 92-100). ACM.
- Fox-Roberts, P., & Rosten, E. (2014). Unbiased generative semi-supervised learning. *The Journal of Machine Learning Research*, *15*(1), 367-443.
- Han, T., Xie, W., and Zisserman, A.. (2020). Self-supervised Co-training for video representation, *In Advances in Neural Information Processing Systems*.
- Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. Information Science 477, 15-29.
- Kingma, D. P., Mohamed, S., Rezende, D. J., & Welling, M. (2014). Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems* (pp. 3581-3589).
- Singh, A., Nowak, R., & Zhu, X. (2009). Unlabeled data: Now it helps, now it doesn't. In *Advances in neural information processing systems* (pp. 1513-1520).
- Yu, S., Krishnapuram, B., Rosales, R., & Rao, R. B. (2011). Bayesian co-training. *The Journal of Machine Learning Research*, *12*, 2649-2680
- Zhou, Z. H., & Li, M. (2005, July). Semi-Supervised Regression with Co-Training. In IJCAI (Vol. 5, pp. 908-913).

References

Other materials

- Figures in the first page: 하상욱 단편시집 – 서울 시
- Zhu, X. (2007). Semi-Supervised Learning Tutorial. International Conference on Machine Learning (ICML 2007).
- Choi, S. (2015). Deep Learning: A Quick Overview. Deep Learning Workship. KIISE.
- Zien, A. (2008). Semi-Supervised Learning. Summer School on Neural Networks.
- Zhu, X. (2009). Tutorial on Semi-Supervised Learning. Theory and Practice of Computational Learning.