Active Learning of Classification Models from Soft-Labeled Groups

Zhipeng Luo Milos Hauskrecht
Department of Computer Science, University of Pittsburgh

ABSTRACT

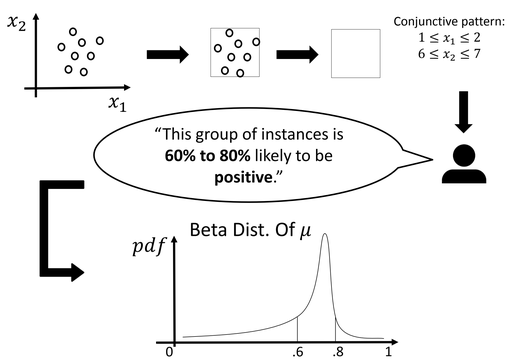
- Challenge: learning classification models from labeled instances often requires huge human annotation efforts.
- Our solution: **actively** learn (instance-based) classification models from (soft) **labeled groups**.
- Our empirical study demonstrates that our method is competitive and can outperform instance-based active learning methods as well as other group- or cluster-based active learning approaches.

MOTIVATION

- Group learning [4,5]: In many real-world situations, it is easier for annotators to provide **one meta-label** on **a set of similar instances** as opposed to annotating individual instance one by one.
  - Electronic health records
  - Image annotation
  - Presidential election results
- Active learning [7]: proper querying strategies can reduce the number of labeled data needed.
- Combining both of the solutions can further reduce human annotation efforts.

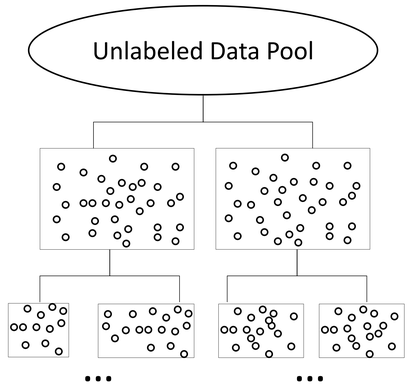
GROUP CONCEPT

- A group is a compact collection of similar instances.
- It is described to humans by using conjunctive patterns over the input space features.
- Annotators assign a **range-based binary label** to the group.
- The range label can be interpreted as a Beta distribution.
  - $P(Group\ is\ positive) = \mu \in [0,1] \sim Beta$



GROUP FORMATION

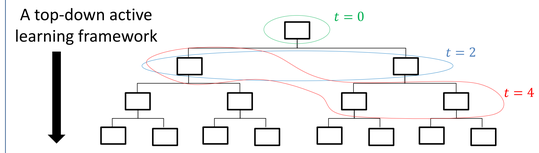
- In practice, groups are not always apparently available.
- Solution: use hierarchical clustering to form groups from general to specific.
- The conjunctive patterns for each group description are learned using a standard decision tree algorithm.



ACTIVE LEARNING FRAMEWORK

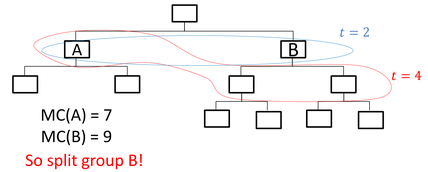
1. Form a hierarchy of groups (unlabeled) by performing hierarchical clustering;
2. Maintain a fringe $F^{(0)} = \{the\ root\ of\ hierarchy\}$;
3. Instantiate an initial classifier with parameters $\boldsymbol{\theta}^{(0)}$;
4. Time $t \leftarrow 0$ is also the number of queries;
5. Repeat :
   i. Actively select a group $G_* \in F^{(t)}$ based on $\boldsymbol{\theta}^{(t)}$;
   ii. Query the labels of G_*'s child groups;
   iii. Replace G_* with its children in the fringe;
      - $F^{(t+\Delta t)} = \{F^{(t)} - G_*\} \cup \{G_*'s\ children\}$.
      - Δt is the number of G_*'s children.
   iv. Retrain the classifier with groups in $F^{(t+\Delta t)}$;
   v. Update $t \leftarrow t + \Delta t$

A top-down active learning framework

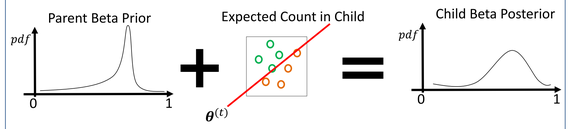


ACTIVE LEARNING STRATEGY

- Maximal expected model change principle [3].
- At each time t, choose the group in current fringe such that if it were split, it would lead to the greatest change to current classifier.



- Use the entire unlabeled data to measure classifier's performance change.
- Essentially we need to **predict** the labels of child groups.
- Method: **Beta Prior** (from parent's label) + **Binomial Counts** (given by classifier) = **Posterior Beta** (child's label predicted)



LEARNING CLASSIFIER FROM GROUPS

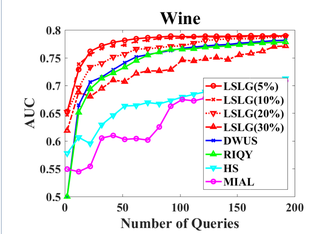
- The learning problem is formulated as a least square regression [4,5] that aims to match (1) the group label given by human with (2) the empirical **averaged** label of all instances in that group estimated by the classifier.
- Suppose at time t, there are N groups $\{G_i\}_{i=1}^N$ in fringe.
  - Each group has n_i instances
  - $n = \sum_{i=1}^N n_i$ is the number of all training instances.
- Each group has a label μ_i following $Beta(a_i, b_i)$.
- Each instance in the group has an estimated μ_{ij} by classifier.
  - $\mu_{ij} = P(the\ j^{th}\ instance\ in\ G_i\ is\ positive|\boldsymbol{\theta})$
- Loss function: $L(\boldsymbol{\theta}) = \sum_{i=1}^N \mathbf{E}_{\mu_i}\{\frac{n_i}{n}(\frac{\sum_j \mu_{ij}}{n_i} - \mu_i)\}^2 + R(\boldsymbol{\theta})$
  - u_i is a given label for group G_i;
  - $\frac{\sum_j \mu_{ij}}{n_i}$ is the averaged instance labels given by model;
  - $R(\boldsymbol{\theta})$ is a regularizing term;
  - $\mathbf{E}_{\mu_i}(\cdot)$ integrates out all possible μ_i.
- The $\boldsymbol{\theta}^*$ that minimizes $L(\boldsymbol{\theta})$ can be solved by:
  - Newton's optimization for simple models;
  - gradient-based method for more complex models.

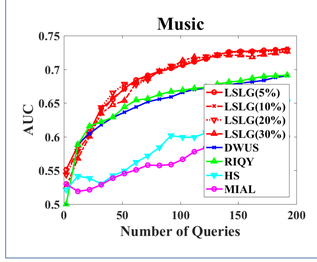
EXPERIMENT SETTING

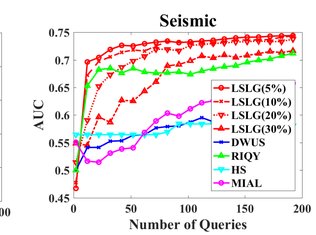
- 3 binary classification datasets from UCI ML repository [1]:
  - **Wine**, which has been used widely;
  - **Music**, which is high-dimensional with 68 features;
  - **Seismic Bump**, which has unbalanced class distribution (7%);
- Active learning methods tested:
  - [7] Density-Weighted Uncertainty Sampling (**DWUS**);
  - [6] **RIQY**, a state-of-the-art group-based active learning approach;
  - [2] Hierarchical Sampling (**HS**);
  - [8] Multi-Instance Active Learning (**MIAL**);
  - Our method: Learning from Soft-Labeled Groups (**LSLG**);
- Group range label simulation:
  - Our group labels are simulated by counting class proportion based on instance labels within each group. The proportion is further expanded to an range label with certain level of precision: 5%, 10%, 20%, 30%.
  - E.g. LSLG(10%) means all the simulated range labels have a fixed interval width = 10%, like '60%~70% positive'.
- The base classifier: Logistic Regression
- Evaluation: We show AUC (Area under ROC curve) of the base classifier on test data for all methods after $k < 200$ queries.

EXPERIMENT RESULTS

- Our methods outperform others when range labels are fairly precise (≤20%).
- For all datasets, the performance of our method gradually drops as the range label uncertainty increases from 5% to 30%, as expected.
- We recommend a robust range width as 20% for future studies.







Contact Information

Zhipeng Luo, Milos Hauskrecht
University of Pittsburgh
Email: {zpluo, milos}@cs.pitt.edu
Website: http://people.cs.pitt.edu/~milos/
Phone: 412-624-8845

References

1. Arthur Asuncion and David Newman. **UCI machine learning repository**, 2007.
2. Sanjoy Dasgupta and Daniel Hsu. **Hierarchical sampling for active learning**. In Proceedings of the 25th international conference on Machine learning, pages 208–215. ACM, 2008.
3. Alexander Freytag, Erik Rodner, and Joachim Denzler. **Selecting influential examples: Active learning with expected model output changes**. In European Conference on Computer Vision, pages 562–577. Springer, 2014.
4. Dimitrios Kotzias, Misha Denil, Nando De Freitas, and Padhraic Smyth. **From group to individual labels using deep features**. In Proceedings of the 21th ACM SIGKDD, pages 597–606. ACM, 2015.
5. Novi Quadrianto, Alex J Smola, Tiberio S Caetano, and Quoc V Le. **Estimating labels from label proportions**. Journal of Machine Learning Research, 10(Oct):2349–2374, 2009.
6. Parisa Rashidi and Diane J Cook. **Ask me better questions: active learning queries based on rule induction**. In Proceedings of the 17th ACM SIGKDD, pages 904–912. ACM, 2011.
7. Burr Settles. **Active learning**. Synthesis Lectures on Artificial Intelligence and Machine Learning, 6(1):1–114, 2012.
8. Burr Settles, Mark Craven, and Soumya Ray. **Multiple-instance active learning**. In Advances in neural information processing systems, pages 1289–1296, 2008.

Acknowledgements

The work presented in this paper was supported in part by grants R01GM088224 and R01LM010019 from the NIH. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.