Interactive Semantics for Knowledge Transfer

Jonghyun Choi JHCHOI@UMIACS.UMD.EDU
Institute of Advanced Computer Studies, University of Maryland, MD USA

Sung Ju Hwang SJHWANG@UNIST.AC.KR
Ulsan National Institute of Science and Technology, Ulsan, Korea

Leonid Sigal LSIGAL@DISNEYRESEARCH.COM
Disney Research. Pittsburgh, PA USA

Larry S. Davis LSD@UMIACS.UMD.EDU
Institute of Advanced Computer Studies, University of Maryland, MD USA

Abstract

We propose a novel learning framework for object categorization with interactive semantic feedback. In this framework, a discriminative categorization model improves through human-guided iterative semantic feedbacks. Specifically, the model identifies the most helpful relational semantic queries to discriminatively refine the model. The user feedback on whether the pattern is valid or not is incorporated back into the model, in the form of regularization, and the process iterates. We validate the proposed model in a few-shot multi-class classification scenario, where we measure classification performance on a set of 'target' classes, with few training instances, by leveraging and transferring knowledge from 'anchor' classes, that contain large set of labeled instances.

1. Approach

Given a labeled dataset $D = \{(\boldsymbol{x}_i, y_i) \in (\mathbb{R}^d, \mathcal{Y})\}_{i=1}^N$, where $\boldsymbol{x}_i$ is a d-dimensional feature vector of i^{th} example, y_i is its class label and N is the number of examples, we learn a model that minimizes classification error for new, unknown, example $\boldsymbol{x}^*$ at test time. We adopt an efficient and scalable discriminative embedding approach (Bengio et al., 2010) to classification, where both the samples, $\boldsymbol{x}_i$, and their labels, y_i, are projected into a common low dimensional space $\mathbb{R}^m$, where $m \ll d$. We denote the projected version of $\boldsymbol{x}_i$ as $\boldsymbol{z}_i = f(\boldsymbol{x}_i)$ and class label $y_i = c \in \mathcal{Y}$ as $\boldsymbol{u}_c$. The goal is then to learn both the embedding function $f(\cdot)$ and the location of the prototypes $\boldsymbol{u}_c$ for all classes such that the projected version of the test instance $f(\boldsymbol{x}^*)$ would be closer to the correct class prototype than to others.

If one assumes existence of semantic information, the classification space can be further improved through graph-based regularization, *i.e.*, semantic relationships as constraints on the placement of prototypes in the embedding space (Hwang et al., 2013; Law et al., 2013). However, as the number of entities increase, the number of possible relationships between them increases rapidly, making it very expensive to annotate all semantic relationships. Further, even if one has a complete set of semantic information, not only using all semantic relationships lead to an unjustifiable computational expense, but also not all semantics would be equally useful for discriminative classification, which suggests that using all of the semantics may even degrade the classification performance. One often needs to trade off discriminative classification accuracy for the ability to encode all the semantics entities in the knowledge set with a fixed dimensional manifold. To address this, we aim to actively identify a compact subset of semantic relations that are most helpful in learning a discriminative classification model.

Specifically, we propose an interactive approach to encode semantics in the form of relative distance: "class a is more similar to class b than to class c." for reducing prohibitive cost of attaining a complete semantic knowledge base whose number of such triplet relationships is cubic in the number of category labels. We summarize the overall procedure of our method in Algorithm 1 and describe detailed steps in the following subsections.

ICML Workshop on Advances in Active Learning - Bridging Theory and Practice., Lille, France, 2015. Copyright 2015 by the author(s).

Algorithm 1 Interactive Learning with Semantic Feedback

```
Input: $(x_i, y_i) \in \mathbb{R}^d \times \mathcal{Y},\ \forall i \in \{1, \dots N\}$.
Output: $\boldsymbol{W} \in \mathbb{R}^{m \times d}, \boldsymbol{U} \in \mathbb{R}^{m \times C}$.
$\mathcal{R} \leftarrow \emptyset$
Initialize $\boldsymbol{W}_{prev}, \boldsymbol{U}_{prev}$ with random matrices
$\boldsymbol{W}^A$ and $\boldsymbol{U}^A \leftarrow$ Solve Eq.(1)
$\delta\boldsymbol{W} = \boldsymbol{W}^A - \boldsymbol{W}_{prev}$, $\delta\boldsymbol{U} = \boldsymbol{U}^A - \boldsymbol{U}_{prev}$
while $\delta\boldsymbol{W} > \epsilon$ and $\delta\boldsymbol{U} > \epsilon$ do
    $\boldsymbol{W}$ and $\boldsymbol{U} \leftarrow$ Solve Eq.(2) with $\mathcal{R}, \boldsymbol{W}^A$
    $\mathcal{P} \leftarrow GenerateOrderedQueries(\boldsymbol{W}, \boldsymbol{U}, \mathcal{R})$ (Sec. 1.3)
    $R \leftarrow Feedback(\mathcal{P})$ (Sec. 1.4)
    $\mathcal{R} \leftarrow \mathcal{R} \cup R$
    $\delta\boldsymbol{W} = \boldsymbol{W} - \boldsymbol{W}_{prev}$, $\delta\boldsymbol{U} = \boldsymbol{U} - \boldsymbol{U}_{prev}$
    $\boldsymbol{U}_{prev} = \boldsymbol{U}$, $\boldsymbol{W}_{prev} = \boldsymbol{W}$
end while
```

1.1. Discriminative Semantic Embedding

Our method is based on the feature embedding approach that embeds both the image features $\boldsymbol{x}_i$ and corresponding class labels y_i into a common low-dimensional space such that the projection of $\boldsymbol{x}_i$, denoted as $\boldsymbol{z}_i$, is closer to the corresponding category embedding $\boldsymbol{u}_{y_i}$ than the embeddings for all the other categories (Weinberger & Chapelle, 2008). This is accomplished by constructing a linear projection $\boldsymbol{W} \in \mathbb{R}^{m \times d}$ such that $\boldsymbol{z}_i = \boldsymbol{W}\boldsymbol{x}_i$, and $\|\boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{u}_{y_i}\|_2^2 + 1 \leq \|\boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{u}_c\|_2^2, \forall c \neq y_i$.

For knowledge transfer, we first build a reference model with well-defined *anchor* classes. Then we build a model on the *target* classes by transferring semantic information from the *anchor* classes.

Reference Model with Anchor classes. The desired objective for categorizing semantic embeddings in the *anchor* classes can be expressed as minimization of the large-margin constraints above for all anchor class instances indexed by $i \in \{1, \dots, N^A\}$ with respect to $\boldsymbol{W}^A$ and prototypes $\boldsymbol{u}_c$:

$$\begin{aligned} \min_{\boldsymbol{W}^A, \boldsymbol{U}^A} & \sum_{i=1}^{N^A} \sum_{c \in \mathcal{C}^A} \mathcal{L}\left(\boldsymbol{W}^A, \boldsymbol{x}_i, \boldsymbol{u}_c\right) + \lambda_1 \|\boldsymbol{W}^A\|_F^2 + \lambda_2 \|\boldsymbol{U}^A\|_F^2, \\ \text{s.t.} \quad & \mathcal{L}(\boldsymbol{W}^A, \boldsymbol{x}_i, \boldsymbol{u}_c) \\ & \quad = \max(\|\boldsymbol{W}^A\boldsymbol{x}_i - \boldsymbol{u}_{y_i}\|_2^2 - \|\boldsymbol{W}^A\boldsymbol{x}_i - \boldsymbol{u}_c\|_2^2 + 1, 0), \\ & \quad \forall i, \forall c \neq y_i, \end{aligned} \tag{1}$$

where N^A is number of training samples in anchor classes ($\mathcal{C}^A$), $\boldsymbol{U}^A$ is a column stacked matrix of label prototypes $\{\boldsymbol{u}_c\}$ of the anchor classes and λ_1 and λ_2 are hyperparameters for scale regularization terms; $\|\cdot\|_F$ refers to a Frobenius norm.

Knowledge Transfer via Relational Semantics. From the learned *anchor* class categorization model, $\boldsymbol{W}^A$ and $\boldsymbol{U}^A$, we transfer the knowledge to the *target* classes that have only a few training samples. Specifically, we use interactively provided semantic relationships $R \in \mathcal{R}$ (the set containing all semantic constraints) to regularize an objective function to learn the discriminative embeddings of target classes as:

$$\begin{aligned} \min_{\boldsymbol{W}, \boldsymbol{U}} & \sum_{i=1}^{N^T} \sum_{c \in \mathcal{C}^T} \mathcal{L}\left(\boldsymbol{W}, \boldsymbol{x}_i, \boldsymbol{u}_c\right) + \lambda_1 \|\boldsymbol{W}\|_F^2 + \lambda_2 \|\boldsymbol{U}\|_F^2 \\ & \qquad + \lambda_3 \|\boldsymbol{W} - \boldsymbol{W}^A\|_F^2 + \gamma \sum_j \Omega\left(R_j, \boldsymbol{U}\right), \\ \text{s.t. } & R_j \subset \mathcal{R}, \\ & \mathcal{L}(\boldsymbol{W}, \boldsymbol{x}_i, \boldsymbol{u}_c) \qquad \forall i, \forall c \neq y_i \\ & \quad = \max(\|\boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{u}_{y_i}\|_2^2 - \|\boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{u}_c\|_2^2 + 1, 0), \end{aligned} \tag{2}$$

where N^T is number of training samples in target classes ($\mathcal{C}^T$), $\{R_j\}$ is a subset of $\mathcal{R}$, and $\boldsymbol{U} = [\boldsymbol{U}^A, \boldsymbol{U}^T]$ is a concatenation of all class prototypes. We regularize the data embedding $\boldsymbol{W}$ with $\boldsymbol{W}^A$, and semantic embedding with $\Omega(R_j, \boldsymbol{U})$, which is a regularizer defined on the relationship R_j.

Optimization. Eq. (1) and Eq. (2) are not jointly convex on $\boldsymbol{W}$ and $\boldsymbol{U}$, but are bi-convex in terms of each variable. We use alternating optimization to solve the problem, where we alternate between the optimization of $\boldsymbol{W}$ and $\boldsymbol{U}$ while fixing the other. We use stochastic sub-gradient method to optimize for each variable.

1.2. Encoding Relational-Semantics by Geometric Topologies

It is shown that the semantic relationships effectively regularize the embedding space for better classification generalization (Hwang et al., 2013; Law et al., 2013). We particularly use the *triplet-based relationships* in which human feedback is of the form of 'object a is more similar to b than to c' since the triplet based relationship is good for the less need of reconciling feedback scales due to its relativity (Tamuz et al., 2011; Kendall & Gibbons, 1990).

Specifically, suppose an target entity, $\boldsymbol{u}_t$, is semantically closer to the anchor entity $\boldsymbol{u}_{a_1}$ than to another anchor entity $\boldsymbol{u}_{a_2}$; we denote such relationship by $R = (t, (a_1, a_2))$ and define its geometric regularizer as a hinge loss type of regularizer that encourages moving $\boldsymbol{u}_t$ closer to $\boldsymbol{u}_{a_1}$ and farther from $\boldsymbol{u}_{a_2}$:

$$\max\left(1 - \|\boldsymbol{u}_{a_2} - \boldsymbol{u}_t\|_2^2 / \|\boldsymbol{u}_{a_1} - \boldsymbol{u}_t\|_2^2, 0\right). \tag{3}$$

Eq.(3), however, is neither differentiable nor convex in terms of $\boldsymbol{u}_*$'s thus makes the optimization difficult if it is used as a regularization term. So, we relax the regularizer by introducing a scaling constant σ_1 as a proxy of $\frac{1}{\|\boldsymbol{u}_{a_1} - \boldsymbol{u}_t\|_2^2}$ by a reciprocal of the distance between the sample mean of class a_1 and t with a smoothed $\max(x, 0)$ around $x = 0$ denoted as $h_\rho(\cdot)$, to make the regularizer differentiable everywhere (Amit et al., 2007):

$$\Omega(R, \boldsymbol{U}) = \sigma_1 h_\rho \left(\|\boldsymbol{u}_{a_1} - \boldsymbol{u}_t\|_2^2 - \|\boldsymbol{u}_{a_2} - \boldsymbol{u}_t\|_2^2 \right). \quad (4)$$

Even though the relationships are local with respect to the associated entities, solving the optimization using the relationships, Eq.(2), changes the topology of the class prototype embeddings globally, which results in a semantically more meaningful model overall.

1.3. What Questions to Ask First?

To reduce the number of semantic relationships in the regularizer, while aiming for better classification, the order of questions to ask is very important.

1.3.1. Generating a Pool of Queries

As a proper ordering all possible candidate triplets is very expensive, we first generate a pool of candidate triplet-based semantic relationships, $\mathcal{R} = \{R | R = (t, (a_1, a_2))\}$, from the label prototypes U and the R, we denote target as $\boldsymbol{u}_t$, and two anchors as $(\boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2})$. Especially, we prioritize it to improve the classification of the target entity that are least confident in classification by transferring knowledge from the anchor entities, that are highly confident in classification. Thus, we choose the target as highly confused ones (*i.e.*, classification accuracy in the current model is low) and the anchor as highly confident ones (*i.e.*, classification accuracy in the current model is high).

Specifically, for each $R = (t, (a_1, a_2))$, we define a scoring function, $S(R, \boldsymbol{U})$, for mining semantic relationship by favoring the most confusing (the least confident) target entity and the least confusing (the most confident) anchor entities. For the measure of confusion of each entity, we regard each entity as a random variable over the class labels and use its entropy, $H(\boldsymbol{u}_c)$. The higher the entropy, the higher the confusion. The scoring function is as the conditional entropy of a target entity, $\boldsymbol{u}_t$, given anchor entities $(\boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2})$ as:

$$\begin{aligned} S(R, \boldsymbol{U}) &= H(\boldsymbol{u}_t | \boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2}) \\ &= H(\boldsymbol{u}_{t_1}, \boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2}) - H(\boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2}). \end{aligned} \quad (5)$$

Given the label of the target entity $\boldsymbol{u}_t$ of the candidate relationship R, we want the anchor entities to be even more certain. In other words, we assume the uncertainty of anchor entities given the target entity label, $H(\boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2} | \boldsymbol{u}_t)$, is 0. Then, we can reduce (5) to:

$$S(R, \boldsymbol{U}) = H(\boldsymbol{u}_t) - H(\boldsymbol{u}_{a_1}, \boldsymbol{u}_{a_2}). \quad (6)$$

Intuitively, the score favors choosing target entities that have high classification confusion and the anchor entities that have low classification confusion.

1.4. Feedback

Feedback can be obtained from human expert(s). We simulate human feedback by an oracle that gives answers based on the distance of attribute description. Since the attribute description is an agglomerative feedback of different criteria from a number of human annotators, it is a reasonable measure for the semantic decision regarding validity of relational queries. Specifically, for each triplet-based relationships, we compute the distance of attribute description of $\boldsymbol{u}_t$ and $\boldsymbol{u}_{a_1}$ and $\boldsymbol{u}_t$ and $\boldsymbol{u}_{a_2}$. If the distance between $\boldsymbol{u}_t$ and $\boldsymbol{u}_{a_1}$ is smaller than the distance between $\boldsymbol{u}_t$ and $\boldsymbol{u}_{a_2}$, the oracle gives an answer to the system of 'Yes', otherwise 'No'. We only use the relationships that are answered as 'Yes' as constraints.

Interactive Learning. Note that the key to our approach is to adaptively update the query generation, which we refer to this as 'interactive' model. We iterate the process multiple times, updating the embedding manifold (model) and use the updated model to generate a new pool and prioritize the queries for the next iteration's feedback.

2. Experiments

2.1. Datasets and Experimental Details

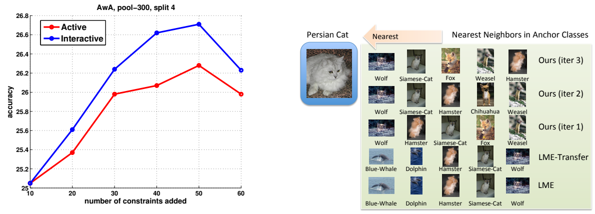
We validate our method on two object categorization datasets: 1) Animals with Attributes (AWA) (Lampert et al., 2009), which consists of 50 animal classes and 30,475 images, 2) ImageNet-50 (Hwang et al., 2013), which consists of 70,380 images of 50 categories.

For visual features, we use the features provided by dataset authors (Lampert et al., 2014; Hwang et al., 2013). For dimension of the embedding space, we choose 75, which is slightly bigger than the number classes (50) for encoding additional semantic information.

We evaluate the performance of knowledge transfer by classification accuracy on target classes in a challenging set-up that has very small number of training samples (2, 5 and 10 samples per class, few-shot learning) with a prior learned with anchor classes that have comparatively more numbers of training samples (30 samples per classes). For testing and validation set, we use a 50/50 split of remaining samples, excluding the training samples. In both datasets, we use 40 classes as anchor classes and 10 classes as target classes. We configure the anchor/target classes, following the configuration of training/test classes in zero-shot/few-shot learning set-up in (Lampert et al., 2014).

2.2. Results

Effect of Interaction. Our interactive learning scheme continuously updates the model to select a better set of questions in terms of classification accuracy. We use a mini batch size of 10 for interactive setting. The interactively mined constraints provide better classification accuracy over an equivalent sized set of constraints produced in a batch. Fig. 1-(a) shows the classification accuracy as a function of number of constraints added by the iteratively updated model and by a batch model. In both cases same measure for selection and ordering was used. Interestingly, as iterations continue, the accuracy starts to drop. We believe it is because there is not much helpful semantics to be added for classification past that iteration.



(a) (b)

Figure 1. Effect of Interaction. (a) Classification accuracy as a function of number of constraints added by active or interactive scoring. (b) Qualitative result of nearest neighbor of target class.

As a qualitative result, we present the nearest neighbor of a target class in the anchor set in Fig. 1-(b). As baseline models (LME, LME-Transfer) do not explicitly enforce the semantic relationships of categories, the nearest neighbors obtained by the baseline models are not semantically meaningful. The nearest neighbors obtained using our model, however, are semantically meaningful from the first iteration onward. As iterations proceed, the nearest neighbor is further refined to be semantically more meaningful, *e.g.*, *Siamese-cat* appears as the second nearest neighbor in the iteration 2 and 3 where as it was a third-nearest neighbor at the first iteration.

# Iter	Positively answered query at its highest rank
1	|fox - persian cat| < |blue whale - persian cat|
2	|grizzly bear - persian cat| < |horse - persian cat|
3	|dalmatian - persian cat| < |beaver - persian cat|
4	|dalmatian - persian cat| < |german shepherd - persian cat|

Table 1. Top Ranked Query as Interaction (Iter) Proceeds. As interactions continue, top ranked query whose target class is 'Persian Cat' becomes semantically more meaningful.

As interaction proceeds, the embedding space becomes semantically more meaningful so do the generated queries. Table 1 shows top positive query related to *Persian-cat* as a function of iterations. In early iterations, the questions try to relate *Persian-cat* to *fox* and *blue whale*. But in the later iterations, the question becomes more semantically meaningful, comparing *Persian-cat* with *dalmatian* and *german shepherd*.

Comparison Among Query-Scoring Metrics. Scoring metric for query is one of the most important components in the interactive framework. In Table 2, we compare the accuracy obtained by adding the constraints by the various scoring schemes. Number of constraints added and other hyper-parameters are determined by cross validation. 'Random'–random ordering of query from the selected pool. 'Entropy'–Entropy-based scores. 'Active'–classification accuracy based score by a batch-mode model. 'Active-Regression'–regressed score of the classification accuracy obtained by a batch-mode model. 'Interactive'–classification accuracy based score by an adaptively updated model, which is our proposal. 'Interactive-UB' refers to a upper bound that our framework can achieve; we score and add the queries based on classification accuracy with test set itself in our interactive model. Note that except 'Interactive', all other scoring metrics are in a batch-mode. The interactive model outperforms the batch mode model, which we denote as 'Active', and other scoring schemes, and is tight to the upper bound. We also present the baseline results of 'LME' and 'LME-Transfer' for reference.

Dataset	Animals with Attribute			ImageNet-50		
# samples/class	2	5	10	2	5	10
LME	22.51±2.48	29.85±1.90	34.52±1.33	23.20±2.97	28.22±2.43	34.67±1.62
LME-Transfer	24.59±2.23	32.17±1.53	35.39±1.67	23.47±2.66	28.78±2.05	34.94±1.03
Random	24.75±2.11	31.32±1.31	35.96±1.66	24.23±1.92	28.72±2.26	34.74±2.26
Entropy	24.96±2.24	31.81±1.27	35.92±1.91	24.60±2.80	28.88±2.43	35.64±0.99
Active-Regression	25.43±1.90	32.49±1.58	36.18±0.88	23.34±2.76	28.99±2.34	35.49±0.89
Active	26.62±1.67	32.42±1.45	36.40±1.33	24.35±2.42	28.55±2.07	35.60±1.01
Interactive	**27.24±1.82**	**33.31±1.28**	**36.46±1.60**	**24.95±2.20**	**29.08±1.88**	**35.62±1.01**
Interactive-UB	28.57±1.85	33.61±2.15	36.86±1.83	25.15±2.13	29.23±1.85	35.95±1.53

Table 2. Classification Accuracy (%) for Comparing Quality of Scoring Function. For ease of comparison, we provide two baselines of the method (LME and LME-Transfer) and the upperbound of our interactive model (Interactive-UB), which is obtained by adding constraints scored by the test set.

Note that all methods use the same validation set to tune parameters. Our scoring metric in 'Active' and 'Interactive', in addition, uses it to prioritize queries to the user as this is the most direct way to measure the effect of adding a particular constraint on the recognition accuracy without using the testing set. While this perhaps makes direct comparison to the baselines slightly less transparent, the comparison of 'Active' and 'Interactive' variants, which both use this criterion, clearly points to the fact that 'Interactive' learning is much more effective in selecting and ordering of constraints.

3. Conclusion

We propose an interactive learning framework that takes human feedback to iteratively refine the learned model. Our method detects recurring relational patterns from a semantic manifold and translate them into semantic queries to be answered and retrain the model by imposing positively feed-backed semantic relationships as constraints. We validate our method against batch learning methods on classification accuracy of target classes with transferred knowledge from anchor classes via relational semantics.

References

Amit, Y., Fink, M., Srebro, N., and Ullman, S. Uncovering Shared Structures in Multiclass Classification. In *ICML*, 2007.

Bengio, S., Weston, J., and Grangier, D. Label Embedding Trees for Large Multi-Class Tasks. In *NIPS*, 2010.

Hwang, S. J., Grauman, K., and Sha, F. Analogy-preserving semantic embedding for visual object categorization. In *International Conference on Machine Learning (ICML)*, pp. 639–647, 2013.

Kendall, M. and Gibbons, J. D. *Rank Correlation Methods*. 5 edition, 1990.

Lampert, C., Nickisch, H., and Harmeling, S. Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

Lampert, C. H., Nickisch, H., and Harmeling, S. Attribute-Based Classification for Zero-Shot Visual Object Categorization. *IEEE Trans. on PAMI*, 2014.

Law, M. T., Thome, N., and Cord, M. Quadruplet-wise Image Similarity Learning. In *CVPR*, 2013.

Tamuz, O., Liu, C., Belongie, S., Shamir, O., and Kalai, A. T. Adaptively Learning the Crowd Kernel. In *ICML*, 2011.

Weinberger, K. and Chapelle, O. Large Margin Taxonomy Embedding with an Application to Document Categorization. In *NIPS*, 2008.