Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Incremental Learning of Visual Categories

Stephan Kirstein, Heiko Wersing

2012

Preprint:

This is an accepted article published in Encyclopedia of the Sciences of Learning. The final authenticated version is available online at: https://doi.org/[DOI not available]

Incremental Learning of Visual Categories

Stephan Kirstein
Honda Research Institute Europe GmbH
Offenbach
Germany
Stephan.Kirstein@honda-ri.de

Heiko Wersing
Honda Research Institute Europe GmbH
Offenbach
Germany
Heiko.Wersing@honda-ri.de

Definition

Incremental learning of visual categories denotes the capability of a visual perceptual system to build up an increasing repertoire of visual concepts based on a sequence of experiences. A visual category is here defined as a possibly large group of individual objects that share similar properties like shape, appearance, or color. Biological visual systems achieve this function very efficiently for their behaviorally relevant categories, where an appropriate generation and selection of features is considered to be responsible for good generalization. Static visual learning models with a fixed a priori set of trainable parameters face severe problems for dynamically changing training sets. In contrast to non-incremental visual approaches incremental ones approach this problem by growing categorical representations that adapt to successively available training stimuli.

Theoretical Background

Visual categories can be differentiated according to their abstraction level into superordinate, basic level, and subordinate categories (Rosch et al. 1976). Superordinate categories share several visual properties ("sheepdogs"), whereas basic level categories ("dogs") may have a larger visual variability. Subordinate categories ("animals") are rather defined by behavioral concepts instead of visual attributes.

Similarity or dissimilarity of categories is generally formalized through the concept of visual features: Local features can be generic object parts or substructures of parts like specific edge combinations. Global features characterize holistic properties like orientation histograms or distribution moments of the object silhouette. Learning appropriate features is often the key problem in visual categorization algorithms. Most established visual feature learning methods are constrained to a stationary ensemble of training patterns, delivering a fixed feature set. On the contrary, the flexibility of incremental human visual category

learning has been explained by the capability of perceptual learning (Goldstone 1998): Visual features are added, pruned, combined or differentiated based on the requirements of the visual categorization task.

Incremental learning for human visual category perception is assumed to be the result of interacting anatomically segregated brain areas: The prefrontal cortex contributes strongly to working memory for attending a learning task, while the transfer from short-term memory (STM) to long-term memory (LTM) is associated with the mediotemporal lobe (MTL) (O'Reilly and Norman, 2002). The MTL is responsible for the memory consolidation process of the visual knowledge into neocortical areas like inferotemporal cortex, being selective to shape categories.

Algorithmic solutions for visual category learning can be distinguished according to the way training data is made available to the algorithm: stationary ensemble, increasing training set with exhaustive memory, and a continuously changing limited training set. Especially the last case requires memory architectures incorporating both STM and LTM components.

The first case of a stationary ensemble is effectively dealt with by modern machine learning theory (Vapnik 2000). Although the theory is well-developed, it cannot be directly generalized to incremental learning with non-stationary training sets.

For the second case of increasing training sets with exhaustive memory, training patterns are added and all previously seen data remains available to the learning approach. The main problem is to achieve the capability of revising feature representations that were learned initially, but have to be adapted in later stages of learning. A typical example is the correlated appearance of a category-specific feature with an unrelated feature in the initial training. Later stages of learning should allow the pruning of the corresponding feature representation.

The third case of a continuously changing training set best matches the situation for human visual category learning, allowing only a limited short-term memory capacity. This task has also been phrased as the life-long learning problem, where an exhaustive memory of all training examples is prohibitive. For the incremental learning of categories this changing training set requires to approach the well-known *"Stability-Plasticity Dilemma"* (Carpenter and Grossberg, 1987): The fundamental conflict between the addition or correction of categorical knowledge (plasticity) and the long-term conservation of previously acquired information (stability).

Important Scientific Research and Open Questions

Categorization approaches can be partitioned into generative, discriminative, and hybrid models (Fritz 2008). Generative probabilistic methods first model the underlying joint probability for each category individually. Based on this

model and the Bayes theorem the posterior class probability is determined. The advantages of generative models are that prior knowledge can easily be incorporated and that typically few training examples are sufficient to reach good performance. In contrast to this, discriminant models like support vector machines (Vapnik 2000) directly learn the mapping from the training vectors to the desired category output. Such discriminant models tend to achieve a better categorization performance compared to generative models if a large training ensemble is available (Ng and Jordan 2001). But purely discriminative determination of category features induces a strong specialization towards already seen examples that causes a strong "catastrophic" forgetting effect for data that was removed from the visible training window. Life-long learning of categories therefore also requires a hybrid combination of discriminative and generative components that allows to keep non-redundant information of previously seen training examples efficiently in memory.

Category-specific features should neglect individual object-specific details and concentrate on reoccurring and stable features. The closer this feature representation is to the shared visual attributes of the categories (e.g. cars have wheels) the easier the category learning becomes. Parts-based methods (Leibe et al., 2004, Hasler et al., 2007) dominate most of the current work in this field. To obtain and describe relevant parts, local feature descriptors are computed and clustered across the category training examples. Good features deliver an optimal compromise between selectivity to the category and stability across different views and category exemplars. Another prominent feature learning approach for visual categorization is agglomerative clustering (Mikolajczyk et al., 2006). The basic idea is to start with a large set of local features (e.g. line segments). During each iteration of the clustering process the most similar features are merged together. The iterative clustering converges into a tree-like structure, where at the lowest level the original local features are represented and in the root node all local features are clustered together. For the learning of visual categories typically the intermediate clusters of the generated tree are used, because they offer a good compromise between complexity and category specificity.

Compared to static learning models with a fixed finite set of adaptable parameters, incremental learning models can increase their representation complexity in a self-optimizing way. On the downside, this greater flexibility has made rigorous mathematical treatment very difficult and has kept incremental learning within a small niche of machine learning theory. Most incremental approaches therefore use heuristics to achieve a good trade off between representation effort, memory capacity and discrimination capability. To achieve self-optimization a learning method has to decide when and where the categorical representation has to be enhanced, based on direct or indirect error measures. Direct error measures are derived from categorization errors so that the representation is adapted based on erroneous training vectors (Kirstein et al. 2009). In contrast to this indirect measures are more connected to the representational accuracy like the locally accumulated quantization error. The representation is then extended in the vicinity of such local error hot spots. The definition of

incremental insertion rules is crucial with respect to the categorization performance but also for convergence speed and allocated resources.

Due to the rapid advancement in mobile and cognitive robotics interactive learning techniques are becoming increasingly popular. The challenges typically lie in the speed of the learning method, to allow human interaction and a quick incorporation of newly acquired data. To achieve high learning speed typically dimensionality reduction or feature selection techniques are used, so that learning only takes place in low-dimensional subspaces. Additionally incremental learning techniques are popular in this context, because of their greater flexibility with respect to a priori unknown categories. Interactivity during the learning process is also efficient for bootstrapping of the representation, because the tutor directly obtain feedback what the learning system already knows and therefore can concentrate on the remaining errors. Thus typically much less training data is required compared to traditional offline learning.

One of the general open questions of visual category learning is the scalability of the approaches to an arbitrary number of categories. So far most approaches concentrate on single (e.g. pedestrians) or very few categories. This is caused by the large amount of a priori knowledge that is required for good categorization performance, but also by the strong focus on the development of isolated methods for feature extraction or learning. A stronger emphasis on larger integrated systems, addressing the co-development of feature and categorical representation for life-long learning may allow a better scaling to much larger categorization systems. Also the scaling of current categorization approaches to basic or subordinate categories is still an open question, because such categories commonly share distinctly less visual properties compared to the more treated superordinate categories. For the categorization of every-day objects, however, most objects may belong to several categories at the same time. Nevertheless categories are trained independently in most approaches. This is sufficient if the training time is irrelevant, but for interactive learning the consideration of co-occurring categories is necessary. For the handling of co-occurring categories the learning approach should be able to extract the defining features for each category based on a feature set reflecting all possible categories. Especially if the training data is incrementally acquired, errors in the feature selection process can not be prevented so that additional correction mechanisms are required.

Cross-References

- Incremental Learning vs. Non-Incremental Learning
- Visual Perception Learning
- Self-organized Learning
- Autonomous Learning
- Learning Algorithms

- Learning-based Knowledge Representation
- Learning in Artificial Neural Networks

References

Goldstone, R.L. (1998). Perceptual Learning. Annual Review of Psychology, 49, 585–612.

Vapnik, V. (2000). The nature of statistical learning theory. New York: Springer.

Fritz, M. (2008). Modeling, Representation and Learning of Visual Categories. Ph. D. thesis, Technical University of Darmstadt.

Kirstein, S., Denecke, A., Hasler, S., Wersing, H., Gross, H.-M., & Körner, E. (2009). A vision architecture for unconstrained and incremental learning of multiple categories. Memetic Computing 1(4), 291–304.

Ng, A. Y. and M. I. Jordan (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In Proc. Advances in Neural Information Processing Systems.

Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem (1976). Basic objects in natural categories. Cognitive Psychology 8, 382–439.