Open-world Machine Learning: Supplementary Material

JITENDRA PARMAR and SATYENDRA S. CHOUHAN, Malaviya National Institute of Technology Jaipur, INDIA
VASKAR RAYCHOUDHARY, Miami University, USA
SANTOSH S. RATHORE, ABV-IIITM Gwalior, INDIA

This file is created as a supplementary material for the review paper "Open-world Machine Learning: Applications, Challenges, and Opportunities" . In this file, the first part discusses the benchmark dataset used in open-world learning. It also lists all datasets based on the proposed year and provides the links for publicly available datasets. The next part discusses related areas such as Transfer Learning, Active Learning, Lifelong Machine Learning, and Multi-Task Learning.

ACM Reference Format:
Jitendra Parmar, Satyendra S. Chouhan, Vaskar Raychoudhary, and Santosh S. Rathore. 2021. Open-world Machine Learning: Supplementary Material. 1, 1 (November 2021), 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Datasets Used in Open-world Machine Learning

Most of the researchers employed benchmark datasets to evaluate the performance of their proposed algorithms. Some of the researchers built their datasets or altered the existing dataset and evaluated the methods. In this section, we discuss some of the datasets primarily used in both the domain of open-world machine learning.

Figure 1 shows the classification of datasets for CV-IP and NLP with their proposed years. Next we discussed publicly available datasets that are used in OWML.

Caltech-256 [17]: $caltech - 256$ has set of 256 categories of object and the total 30607 images in this dataset. Each category contains minimum 80 and maximum 827 images, these categories are further labeled with three tags on the basis of image quality. the labels are *good*, *bad* and *none* (out of the category). the *good* indicates clear vision and *bad* indicates clutters or artistic example where *none* indicates the image does not belong to the particular category.

MNIST [26, 27]: Modified National Institute of Standard and Technology, wildly known as *MNIST* is a handwriting dataset. It is a modified version of NIST. MNIST is used in optical character reorganization and is also used as a test case in pattern recognition and machine learning. We have analyzed, MNIST has become a standard for testing machine learning algorithms. There are 60000 training images; some may use for validation and 10000 images for testing purposes. All the digits are black and white and normalized in seize, the center intensity with $28 * 28$ pixels; thus, the dimension of the image is $28 * 28 = 784$, and each element is a binary. The MNIST has tested for almost all the benchmark baseline algorithms and well-known Fields of classification such as

Authors' addresses: Jitendra Parmar, 2019rcp9044@mnit.ac.in; Satyendra S. Chouhan, sschouhan@mnit.ac.in, Malaviya National Institute of Technology Jaipur, JLN Marg, Jaipur, Rajsthan, INDIA, 302017; Vaskar Raychoudhary, Miami University, 510 E. High St., Oxford, Ohio, USA, raychov@miamioh.edu; Santosh S. Rathore, ABV-IIITM Gwalior, , Gwalior, MadhyaPradesh, INDIA.

XXXX-XXXX/2021/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

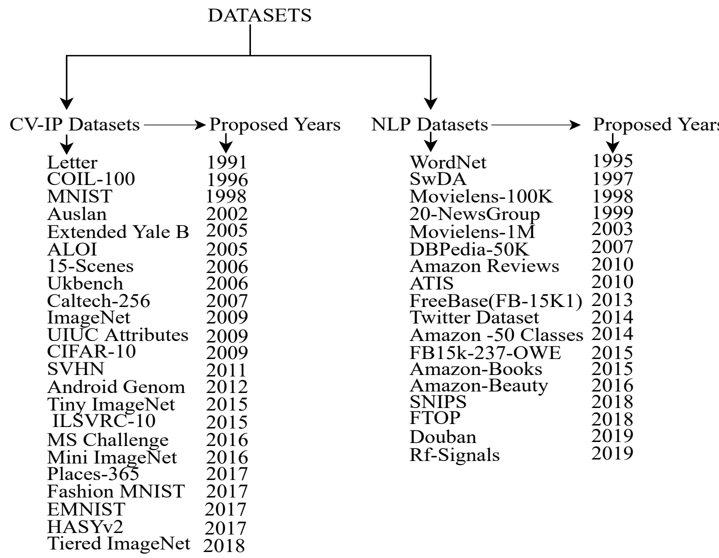

Fig. 1. Classification of Datasets Used in CV-IP and NLP with Proposed Years

linear classification, convolution neural networks, simple neural networks, K-Nearest neighbors, support vector machines (SVMs), boosted stamps, and nonlinear classification.

Fashion-MNIST [54]: The Fashion-MNIST is a dataset of Zalando's article images which containing a training set of 60,000 images, and a test set of 10,000 images. Each image is a 28x28 pixel and grayscale image related to a label from 10 different classes. Zalando aims for Fashion-MNIST to serve as a substitution for the original MNIST dataset, which comprises many handwritten digits, for benchmarking artificial intelligence and machine learning algorithms.

ImageNet [14]: $ImageNet$ is a large scale ontological dataset of visual objects. The structure of $ImageNet$ inspired from $WordNet$ dataset thus it constructs on backbone of $WordNet$. $ImageNet$ has 80000 synsets of $WordNet$ with around 1000 full resolution cleaned images and its updating continuously. The basic $ImageNet$ contains 3.2 million images with 12 sub-trees and $5247 synsets$. $ImageNet$ is hierarchical dataset like $WordNet$ which contains the synonym's of world in tree structure. The 12 sub-trees consist of the following categories: bird, reptile, vehicles, musical instruments, tools, fruits, mammals, fish, amphibians, geological formulations, furniture, and flower with 5247 synsets. The continual updating of this data aimed at 50 million images in a hierarchical structure. The evolution with various baseline methods showed that the $ImageNet$ has a 99.7 percent average precision rate.

Tiny-ImageNet [25]: $Tiny - ImageNet$ is a collection of 100000 images that are retrieved from internet. The resolution of all these images is $32 * 32$ pixels and $64 * 64$. Tiny-ImageNet has 200 categories of images, of which $100,000$ images for training, 10000 for validation, and 10000 images are reserved for testing. Images are collected by sending all search words in $WordNet$ to the image in the search engine. it is a successful dataset tested on application-specific algorithms because of a high level of noise and low resolution. $Tiny - ImageNet$ is suitable for general-purpose algorithms. It also contains synsets of high quality with an average resolution of $400 * 350$.

CIFAR-10 [21]: The CIFAR-10 data set was developed by the Canadian Institute for Advanced Research. It contains 10 categories (dog, frog, automobile, bird, horse, ship, truck, airplane, cat, and deer) of images, as it is a subset of CIFAR-100, which consists of 100 categories of images. Total

60000 color images are in CIFAR-10 with the resolutions of $32 * 32$ pixels, and every class has 600 images. The dataset is divided into training and test sets, which consist of 50000 and 10000 images, respectively. The entire dataset is divided into batches, 5 batches for training and 1 for testing. The testing batch has 1000 random images, and the rest of the images randomly contain by training batches.

SVHN [36]: The Street View House Number (SVHN) contains 600000 labeled digits that are cropped from actual street view images. The initial goal of this dataset is to identify house numbers from original street view images. There are two types of images one is whole numbers, and another is cropped digits. The whole numbers contain high-resolution full-size original images with character-level bounding boxes for the house number. The cropped digits are character-level ground truth, and all these digits are resized with the resolution of $32 * 32$ pixels. The SVHN is further divided for training and testing, and there are 73257 digits images for training and 26032 digits images for testing. The rest of the images also reserve as extra for training.

20-NewsGroup [23, 40]: It is a collection of near about 20000 new documents which are collected from different newsgroups. It is one of the most popular datasets for the application that is based on text classification in machine learning. All the newsgroups are different, but some of the new groups are related to each other. Generally, 90 percent of documents are used for training and 10 percent for testing. 20-NewsGroup is publicly available in different forms, and the original dataset is not sorted but later on is sorted by date. The headers and duplicate data are also removed in this version. The latest version of this data is available with 18828 documents with only "From" and "Subject" headers.

Amazon Product Reviews [35]: Amazon.com is one of the most successful e-commerce web site across the globe since it emerges. The amazon product review dataset contains 5.8 million reviews, written by 2.14 million for 6.7 million products from 9600 different categories when extracted from these reviews. The dataset has 8 different headers such as Product ID, Reviewer ID, Rating, Date, Review Title, Review Body, Number of Helpful Feedbacks, and Number of Feedbacks. It can be used for feature identification and construction of both reviewer and reviews, and the features can be Review Centric or Product-Centric. The amazon product reviews.

50-Class Reviews [9]: The 50-Class review dataset is a collection of reviews, and there are 50 different categories of products. The data set has two versions, one has reviews of 50 different electronic items, and the other has 50 different non-electronic items. There are 1000 reviews for each product or domain.

WordNet [34]: WordNet is a multi-language (Approx 200 languages) lexical dataset of semantic relationships among words, including meronyms, synonyms, and hyponyms. Some synsets contain synonyms in a group with short definitions and examples. The WordNet is a popular dataset for text analysis applications in artificial intelligence and machine learning. Initially, it has created for the English language only; later on, it extended for other languages, and updating is continuous to add a new language in WordNet. The WordNet contains approx 175979 words which are organized in 175979 synsets, and there is a 207016 pairs which are word-sense pair. All synsets are connected with semantic relations.

SwDa [19, 43, 46]: The Switchboard Dialog Act Corpus (SwDA) covers the SwDA-1 Telephone Speech Corpus, and some tags recapitulate semantic, syntactic, and pragmatic information about the related turn.

Some publicly available datasets with their repository link are shown in Table 1.

2 Related Areas

Some related areas that are closely associated with open-world machine learning are mentioned and discussed briefly in this section.

Table 1. Publicly Available Benchmark Datasets Repositories

Dataset	Link
Caltech-256 [17]	http://www.vision.caltech.edu/Image_Datasets/Caltech256
MNIST [26, 27]	http://yann.lecun.com/exdb/mnist
Extended Yale B [28]	http://vision.ucsd.edu/ leekc/ExtYaleDatabase/ExtYaleB.html
ALOI [16]	https://aloi.science.uva.nl
UIUC Attributes [15]	https://vision.cs.uiuc.edu/attributes
Mini ImageNet [53]	https://cseweb.ucsd.edu/ weijian/static/datasets/mini-ImageNet
Fashion-MNIST [54]	https://github.com/zalandoresearch/fashion-mnist/tree/master/data
HASYv2 [48]	https://zenodo.org/record/259444.YX94vBzhXIU
ImageNet [14]	http://image-net.org/download
Tiny-ImageNet [25]	http://cs231n.stanford.edu/tiny-imagenet-200.zip
CIFAR-10 [21]	https://www.cs.toronto.edu/~kriz/cifar.html
RF Signal Dataset [18]	https://github.com/xguo7/MDCC-for-open-world-recognition
Twitter Dataset [39]	https://github.com/xguo7/MDCC-for-open-world-recognition
SVHN [36]	http://ufldl.stanford.edu/housenumbers
20-NewsGroup [23, 40]	http://qwone.com/~jason/20Newsgroups
Amazon Product Reviews [35]	https://jmcauley.ucsd.edu/data/amazon
WordNet [34]	https://wordnet.princeton.edu/download
SwDa [19, 43, 46]	https://web.stanford.edu/~jurafsky/ws97/
ATIS [52]	https://rasa.com/docs/rasa/nlu-training-data/#json-format
FB15k [5]	https://www.microsoft.com/en-us/download/ confirmation.aspx?id=52312
DBpedia [3]	https://wiki.dbpedia.org/datasets
EMNIST [11]	https://www.nist.gov/itl/products-and-services/emnist-dataset
Auslan [20]	https://archive.ics.uci.edu/ml/datasets/Australian +Sign+Language+signs+(High+Quality)
Ukbench [37]	https://archive.org/download/ukbench
Places-2 [56]	http://places2.csail.mit.edu/download.html

2.1 Transfer Learning

Conventional Machine Learning (ML) techniques perform predictions on the expected data by applying analytical principles trained on previously accumulated unlabeled or labeled training examples [4, 22, 55]. Analysis on transfer learning has brought attention since 1995 in several titles: knowledge transfer, learning to learn, multitask learning, knowledge consolidation, inductive transfer, knowledge-based inductive bias, context-sensitive learning, cumulative learning [49, 51], and multitask learning framework [31]. Transfer learning involves interpreting data for a reference task to provide a productive basis for a new task. Transfer learning is often applied to specific data sets, which have some labeled value. For example, an actual demonstrative prototype of one virus would have a significant advantage to developing a distinguishing prototype for another virus, for which fewer training samples are available. While all learning involves generalization across all queries, transfer learning illustrates the transfer of information across comparable but non-consistent fields, tasks, and distributions. In distinction, the unlabeled data does not require to be obtained from a similar task in the transfer learning framework. In the prior decade, there has been substantial development in improving cross-task transfer utilizing both discriminative and generative strategies in a broad category of frames.

2.2 Active Learning

Active Learning [50] is a discipline of machine learning where the algorithm is designed for learning, can choose the data for learning, or learning strategy generated during learning (Figure 2). The active learning methodologies can play an essential role in domains that are dealing with real-time data such as speech recognition, information extraction [42], classification, and filtering. Moreover, active learning provides high accuracy with a small testing size of labeled data.

There are different types of scenarios of active learning, such as membership query synthesis [1], stream-based selective sampling [12], and pool-based active learning [29]. The standard data mining methods learn models with isolated data and make a prediction based on static models [4, 22, 55].

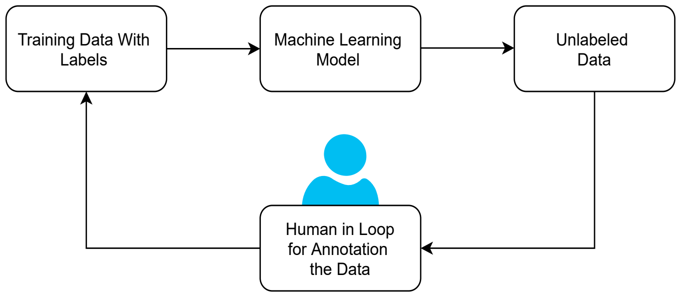

Fig. 2. Basic Framework of Active Learning [41]

It needs to use previous knowledge, or a learning model should transfer knowledge, and it must be used to predict future learning. It is termed as transfer learning [38]. The knowledge can be transferred in various forms such as transferring knowledge of instances [13], knowledge of feature representations [2] (for both supervised and unsupervised), knowledge of parameters [24] and relational knowledge [33].

2.3 Lifelong Learning/Continual Machine Learning

Lifelong machine learning is a system that can continuously learn from different domains, and this knowledge can be used effectively on future tasks in an efficient manner [45]. The selective knowledge is transferred when learning a novel task. Knowledge is retained from a different source and improves learning. The various techniques of lifelong learning as prior works in knowledge retention and improves learning a new task. The major tasks of lifelong machine learning are shown in Figure 3.

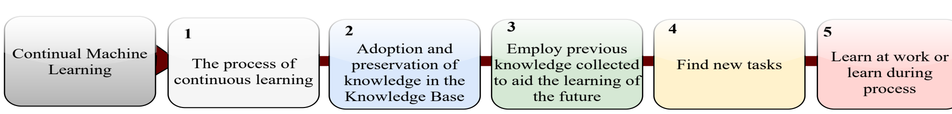

Fig. 3. Tasks of Continual Machine Learning

There are different names: constructive induction, incremental and continual learning, explanation-based learning, sequential task learning, and never-ending learning. These methods are further divided into different categories: lifelong machine learning is supervised learning, continual learning is reinforcement learning, and self-taught learning or never-ending learning is unsupervised learning.

Supervised learning lifelong learning uses Explanation-based Neural Network (EBNN) using back propagation gradient. Whenever new learning tasks occur, EBNN uses prior domain information of the task. EBNN gives more accurate results even with fewer amounts of data. In [44], authors suggested knowledge-based cascade correlation neural networks. This method uses prior trained networks and concealed units to set the new bias for a novel task. Unsupervised Lifelong learning is used to increase the system's scalability, and adaptive resonance theory had been used to map the bottom and top nodes of clusters. Set threshold to consider a new example node, a map with vigilance parameter (below threshold).

In [47], authors proposed a novel approach to ensemble clusters from the primary partition of objects; it uses labels of the cluster deprived of accessing the original features. The self-taught learning models build high-level features by using unlabeled data and test such models for various

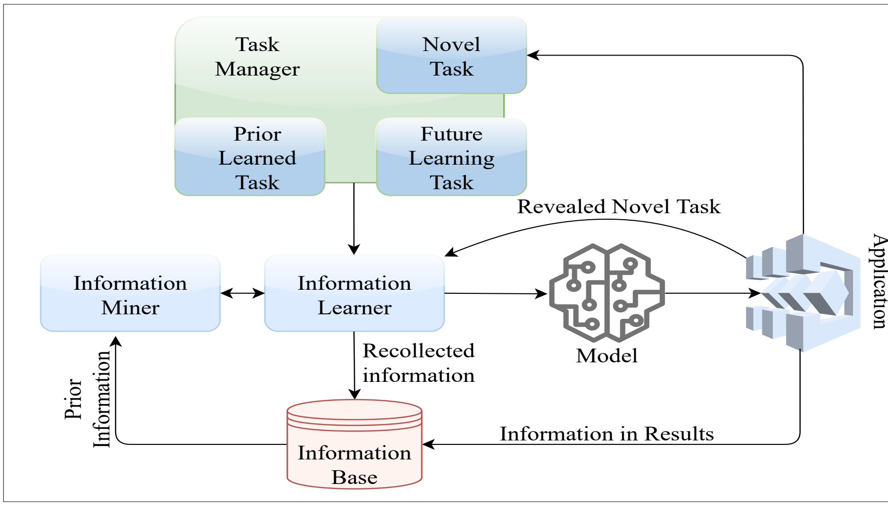

Fig. 4. Basic Framework of Lifelong Machine Learning [10]

classification applications in image, web, and song genres. Lifelong learning goals can be achieved by another popular method is Never-ending Language Learner (NELL) [6]. NELL extracts data or reads information from the web and increases its knowledge, then learns how to perform a new task better than the same task done in the earlier day. Rather than focus on conventional machine learning, the system should retain knowledge and transfer this to the system to the learning agent. The system should learn sequential tasks and increase their magnitude.

2.3.1 Challenges and Benefits of Lifelong Learning Models [45].

- **Input /output Type, Complexity and Cardinality :** The real-time environment has a variety of data from different domains; it can differ in nature. The attributes of each input may vary according to their source and required task.
- **Training Examples Vs Prior Knowledge :** In life-long learning systems, prior knowledge is a crucial part of the end-to-end system to achieve accuracy while performing a new task. There is a need to retain valid data from the knowledge base that must have information act as a training example.
- **Effective and Efficient knowledge Retention :** The system must retain efficient information that must not be erroneous. Furthermore, it must use finite memory to store knowledge with limited computational capacity. The system must be capable of handling duplicate data and increase the accuracy of the prior knowledge.
- **Effective and Efficient knowledge Transfer :** Prior knowledge should not increase computational time and effort. Moreover, the transfer of knowledge shout not generates less accurate inputs/models for new tasks. There are three major components of lifelong learning.
  (1) Retention of learned task knowledge,
  (2) Selective transfer of prior while learning a new task, and
  (3) The system must ensure that retention and transfer of knowledge must be efficient.
- **Scalability :** Scalability is one of the most challenging and essential aspects of almost all fields of computer science. The system must be able to adapt increments in volumes of input data. The lifelong learning systems must be able to address the space and time complexity of all these factors.

- **Heterogeneous Domain of Task :** The lifelong learning systems must handle data from different domains by establishing relations among the origin domain and targeted domain. There are so many features that are common between but diversity in transferred knowledge data also exists. The system must have the ability to map features in transferred knowledge.

2.4 Multi-task Learning

Multi-task learning (MTL) [7, 8, 30] acquires various associated tasks concurrently, beaming at delivering a more reliable representation by using the associated knowledge yielded by various jobs. The motive behind introducing inductive bias in Multi-task learning is to joint hypothesis space of every job by utilizing the task-relatedness building. It additionally inhibits over-fitting in the specific job and therefore has a more immeasurable generalization capability. Unlike transfer learning, it mainly does various jobs preferably than various areas as much of the area's existing research is based on several comparable jobs of the identical application area. Multi-task learning allows those jobs are strictly associated with each other. There are several hypotheses in terms of job-relatedness, which drive to another modeling strategy. Many researchers continue to hypothesize that all job data come from the same sources and are correlated to the standard or global models. According to this hypothesis, they created the association among jobs employing a task-coupling parameter, including regularization. In [32], authors proposed multi-task learning for the deep neural network. They classify multi-tasking tasks into two categories for deep learning. First is classification, and second is ranking; in classification, the model identifies the queried domain, whereas the ranking model finds the relevant queries.

References

[1] Dana Angluin. 1988. Queries and concept learning. *Machine learning* 2, 4 (1988), 319–342.

[2] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2008. Convex multi-task feature learning. *Machine learning* 73, 3 (2008), 243–272.

[3] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The semantic web*. Springer, 722–735.

[4] Elena Baralis, Silvia Chiusano, and Paolo Garza. 2007. A lazy approach to associative classification. *IEEE Transactions on Knowledge and Data Engineering* 20, 2 (2007), 156–171.

[5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-Relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada). Curran Associates Inc., Red Hook, NY, USA, 2787–2795.

[6] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI conference on artificial intelligence.*

[7] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997), 41–75.

[8] Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye. 2009. A convex formulation for learning shared structures from multiple tasks. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 137–144.

[9] Zhiyuan Chen and Bing Liu. 2014. Mining topics in documents: standing on the shoulders of big data. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1116–1125.

[10] Zhiyuan Chen and Bing Liu. 2018. Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, 3 (2018), 1–207.

[11] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. 2017. EMNIST: Extending MNIST to handwritten letters. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2921–2926.

[12] David Cohn, Les Atlas, and Richard Ladner. 1994. Improving generalization with active learning. *Machine learning* 15, 2 (1994), 201–221.

[13] Wenyuan Dai, Ou Jin, Gui-Rong Xue, Qiang Yang, and Yong Yu. 2009. Eigentransfer: a unified framework for transfer learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 193–200.

[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. 248–255.

[15] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. 2009. Describing objects by their attributes. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 1778–1785.

[16] Jan-Mark Geusebroek, Gertjan J Burghouts, and Arnold WM Smeulders. 2005. The Amsterdam library of object images. *International Journal of Computer Vision* 61, 1 (2005), 103–112.
[17] Gregory Griffin, Alex Holub, and Pietro Perona. 2007. Caltech-256 object category dataset. (2007).
[18] Xiaojie Guo, Amir Alipour-Fanid, Lingfei Wu, Hemant Purohit, Xiang Chen, Kai Zeng, and Liang Zhao. 2019. Multi-stage deep classifier cascades for open world recognition. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 179–188.
[19] Dan Jurafsky. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. *Institute of Cognitive Science Technical Report* (1997).
[20] Mohammed Waleed Kadous et al. 2002. *Temporal classification: Extending the classification paradigm to multivariate time series*. Citeseer.
[21] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).
[22] Ludmila I Kuncheva and Juan J Rodriguez. 2007. Classifier ensembles with a random linear oracle. *IEEE Transactions on Knowledge and Data Engineering* 19, 4 (2007), 500–508.
[23] Ken Lang. 1995. Newsweeder: Learning to filter netnews. In *Machine Learning Proceedings 1995*. Elsevier, 331–339.
[24] Neil D Lawrence and John C Platt. 2004. Learning to learn with the informative vector machine. In *Proceedings of the twenty-first international conference on Machine learning*. 65.
[25] Ya Le and Xuan Yang. 2015. Tiny imagenet visual recognition challenge. *CS 231N* 7, 7 (2015), 3.
[26] Yann LeCun. 1998. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/* (1998).
[27] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
[28] Kuang-Chih Lee, Jeffrey Ho, and David J Kriegman. 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on pattern analysis and machine intelligence* 27, 5 (2005), 684–698.
[29] David D Lewis and William A Gale. 1994. A sequential algorithm for training text classifiers. In *SIGIR'94*. Springer, 3–12.
[30] Hui Li, Xuejun Liao, and Lawrence Carin. 2009. Multi-task Reinforcement Learning in Partially Observable Stochastic Environments. *Journal of Machine Learning Research* 10, 5 (2009).
[31] Qiuhua Liu, Xuejun Liao, Hui Li, Jason R Stack, and Lawrence Carin. 2008. Semisupervised multitask learning. *IEEE transactions on pattern analysis and machine intelligence* 31, 6 (2008), 1074–1086.
[32] Xiaodong Liu, Jianfeng Gao, Xiaodong He, Li Deng, Kevin Duh, and Ye-Yi Wang. 2015. Representation learning using multi-task deep neural networks for semantic classification and information retrieval. (2015).
[33] Lilyana Mihalkova, Tuyen Huynh, and Raymond J Mooney. 2007. Mapping and revising markov logic networks for transfer learning. In *Aaai*, Vol. 7. 608–614.
[34] George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39–41. https://doi.org/10.1145/219717.219748
[35] Susan M Mudambi and David Schuff. 2010. Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly* (2010), 185–200.
[36] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. 2011. Reading digits in natural images with unsupervised feature learning. (2011).
[37] David Nister and Henrik Stewenius. 2006. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. Ieee, 2161–2168.
[38] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.
[39] Hemant Purohit, Carlos Castillo, Fernando Diaz, Amit Sheth, and Patrick Meier. 2014. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday* (2014).
[40] Jason Rennie and Ken Lang. 2008. The 20 Newsgroups data set. *Available in web page< URL: http://qwone. com/jason/20Newsgroups* (2008).
[41] Burr Settles. 2009. Active learning literature survey. (2009).
[42] Burr Settles, Mark Craven, and Lewis Friedland. 2008. Active learning with real annotation costs. In *Proceedings of the NIPS workshop on cost-sensitive learning*. Vancouver, CA:, 1–10.
[43] Elizabeth Shriberg, Andreas Stolcke, Daniel Jurafsky, Noah Coccaro, Marie Meteer, Rebecca Bates, Paul Taylor, Klaus Ries, Rachel Martin, and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and speech* 41, 3-4 (1998), 443–492.
[44] Thomas R Shultz and Francois Rivest. 2001. Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science* 13, 1 (2001), 43–72.
[45] Daniel L Silver, Qiang Yang, and Lianghao Li. 2013. Lifelong machine learning systems: Beyond learning algorithms. In *2013 AAAI spring symposium series*.

[46] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics* 26, 3 (2000), 339–373.
[47] Alexander Strehl and Joydeep Ghosh. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research* 3, Dec (2002), 583–617.
[48] Martin Thoma. 2017. The hasyv2 dataset. *arXiv preprint arXiv:1701.08380* (2017).
[49] Kristinn R Thórisson, Jordi Bieger, Xiang Li, and Pei Wang. 2019. Cumulative learning. In *International Conference on Artificial General Intelligence*. Springer, 198–208.
[50] Sebastian Thrun. 1995. Exploration in active learning. *Handbook of Brain Science and Neural Networks* (1995), 381–384.
[51] Sebastian Thrun and Lorien Pratt. 1998. Learning to learn: Introduction and overview. In *Learning to learn*. Springer, 3–17.
[52] Gokhan Tur, Dilek Hakkani-Tür, and Larry Heck. 2010. What is left to be understood in ATIS?. In *2010 IEEE Spoken Language Technology Workshop*. IEEE, 19–24.
[53] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. *Advances in neural information processing systems* 29 (2016), 3630–3638.
[54] Han Xiao, Kashif Rasul, and Roland Vollgraf. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747* (2017).
[55] Xiaoxin Yin, Jiawei Han, Jiong Yang, and Philip S Yu. 2006. Efficient classification across multiple database relations: A crossmine approach. *IEEE Transactions on Knowledge and Data Engineering* 18, 6 (2006), 770–783.
[56] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40, 6 (2017), 1452–1464.