EDITOR: San Murugesan, san@computer.org

On the Persistence of Multilabel Learning, Its Recent Trends, and Its Open Issues

Nikolaos Mylonas and Ioannis Mollas, *Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece*

Bin Liu, *Chongqing University of Posts and Telecommunications, Chongqing, 400065, China*

Yannis Manolopoulos, *Open University of Cyprus, Nicosia, 2220, Cyprus*

Grigorios Tsoumakas, *Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece*

Multilabel data comprise instances associated with multiple binary target variables. The main learning task from such data is multilabel classification, where the goal is to output a bipartition of the target variables into relevant and irrelevant ones for a given instance. Other tasks involve ranking the target variables from the most to the least relevant one or even outputting a full joint distribution for every possible assignment of values to the binary targets.

Multilabel learning started gaining traction as a research topic about 15 years ago. Two early events that got it more widely known were the First and Second Workshops on Multilabel Classification held at Bled, Slovenia, with ECML PKDD 2009 and Haifa, Israel, with ICML 2010, respectively. Despite years of progress, multilabel learning continues to attract the interest of researchers (see Figure 1). The ECML PKDD test-of-time awards for 2017 and 2019 were both given to multilabel learning papers of 2007 and 2009, respectively. This may have contributed to a renewed interest and could partly explain the steep increase in the number of papers after 2017. The ECML PKDD 2022 workshop on current trends and open challenges of multilabel learning at Grenoble, France, was a well attended event with a full room of 60-person capacity. This article looks into what makes multilabel learning such a persistent research topic, discusses some of the important recent trends in this area, and points to open issues worth addressing.

A PERSISTENT RESEARCH TOPIC

One of the main reasons for the continuous popularity of multilabel learning is the ubiquity of multilabel data. Indeed, such data arise in a multitude of domains, including biology, health care, e-commerce, (social) media, sentiment analysis, energy, transportation, and robotics, mainly in the context of annotating unstructured data (documents, audio, images, video, and biological sequences). Multilabel data can even arise as part of another artificial intelligence (AI) task.[1] Therefore, a significant proportion of multilabel learning papers and articles focus on discussing the multilabel nature of particular applications and how they deal with them via existing multilabel learning algorithms.

A second important reason for the persistence of multilabel learning is the compelling research challenges that it offers, with the steam engine being the modeling and exploitation of label dependencies. When these dependencies are considered independently of the input features, they are called *marginal*, while, otherwise, they are called *conditional*. In terms of how they model such dependencies, learning algorithms can be categorized as *first-order* (modeling each different label separately), *second-order* (modeling pairs of labels), and *higher-order* (simultaneously modeling more than two labels). The relationship of label dependence to the type of loss to be minimized was one of the important theoretical results in this topic. The challenge of label dependence continues to engage researchers up to today, both on the practical side and, more importantly, on the theoretical side.[2]

Another important challenge of multilabel learning is dealing with a very large number of labels, also known as *extreme multilabel classification*.[3] While this challenge can also arise in multiclass classification tasks, mutual exclusivity is less common when there is a large number of classes. Therefore, typical extreme classification datasets are multilabel.[a] Interestingly, first-order

Digital Object Identifier 10.1109/MIS.2023.3255591
Date of current version 4 April 2023.

[a] The Extreme Classification Repository: http://manikvarma.org/downloads/XC/XMLRepository.html.

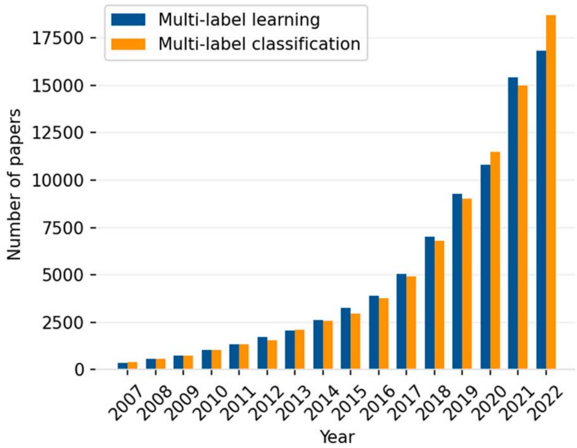

FIGURE 1. Papers since 2007 on "multilabel classification" and "multilabel learning" according to Google Scholar.

approaches, such as binary relevance, achieve excellent results in terms of prediction accuracy, yet they require a large amount of computational resources. Sparse learning, which builds models with minimum computational demands, is a solution to this issue. Other state-of-the-art approaches include ensemble-based methods that partition the label set, enabling faster computational times; label embedding methods, which assume that the embedding matrix will have a low enough rank despite the large label set; and, finally, deep learning architectures, which provide embeddings for the input and output spaces.

As a final reason for the perpetuity of multilabel learning, we would mention its horizontal nature, as it just defines a special type of data and corresponding supervised learning tasks. It can, therefore, be addressed by different families of algorithms, such as trees, probabilistic models, and neural networks. Moreover, it can also be considered in the context of special supervised learning tasks, such as active learning, semisupervised learning, multi-instance learning, and zero-shot learning. Much of the huge amount of work in multilabel learning is related to adaptations of existing learning algorithms and tasks from the binary or multiclass classification scenario to the multilabel one.

RECENT TRENDS

Class Imbalance

A recently identified important challenge in multilabel learning is that of class imbalance. Two different types of class imbalance are present in multilabel data. First, the number of instances relevant to each label is much less than the number of irrelevant ones, resulting in *imbalance within labels*. Second, the frequencies of the labels are typically significantly different from each other, which leads to *imbalance between labels*.

Approaches for dealing with class imbalance in multilabel data can be divided into two main groups: algorithm-based methods and multilabel sampling methods. Algorithm-based methods focus on developing multilabel learning approaches that are able to handle the class imbalance problem directly. They introduce cost-sensitive learning or weighting strategies to existing multilabel learning models or employ problem transformation techniques to reduce the imbalance within and between labels simultaneously. On the other hand, multilabel sampling methods alleviate the imbalance level of the whole dataset by manipulating the training instances in a preprocessing step. Sampling methods that leverage the imbalance within a local region as the criterion for removing or creating instances outperform those that depend on the global imbalance level of the whole dataset.[4]

Concept Evolution

An issue that arises in real-world multilabel data streams is that the target variables evolve over time. The ancient Greek philosopher Heraclitus has said that "the only thing constant is change itself." The Medical Subject Headings (MeSH) ontology, used for indexing the references of PubMed, is updated weekly, with major updates released annually.[b] Such updates involve simple changes, such as the introduction of new headings and the retirement of existing ones, but also complex changes, such as merging/splitting headings, changes in the hierarchical structure of the headings, and changes involving the concepts associated with a heading or the entry terms associated with a concept. IEEE Thesaurus[c] is a controlled vocabulary of more than 11,500 descriptive terms that is updated on a biannual basis and is useful for the indexing and retrieval of articles and other material from IEEE publications. The types of changes that occur in the IEEE Taxonomy are similar to those of MeSH.

When new labels are introduced in a multilabel data stream, the current model cannot deliver predictions for them, as they typically lack ground-truth annotations. One way to address this is via zero-shot learning. In the case of textual data, for example, a semantic embedding of both the new labels and the document in the same space can be computed, and predictions can be given based on their similarity in this space.[5] Another way to address the lack of

[b]What's new in MeSH: https://www.nlm.nih.gov/mesh/whatsnew.html.
[c]IEEE Thesaurus: https://www.ieee.org/publications/services/thesaurus.html.

ground-truth data is via weakly supervised learning, in particular by trying to obtain inaccurate supervision cheaply. In the case of textual data again, a mention of the label inside the document could hint that the document is a positive example for the label. For the MeSH ontology in particular, provenance information has been successfully used as a source of inaccurate supervision to address the lack of ground-truth data.[6]

Deep Learning

Multilabel learning was no exception to the areas that were enriched by the recent advances in deep learning. Deep learning architectures can be used to obtain embedding representations for both the input (feature) and the output (label) space. The powerful learning capabilities of deep neural networks are typically used in multilabel classification tasks involving image and textual data, with convolutional, recurrent, and transformer architectures being frequently used. Deep learning is commonly used in multilabel classification to generate label embeddings using autoencoders that retain label dependencies. This is especially prevalent in extreme multilabel classification approaches, where most of the available solutions demand high computational resources. In addition, label dependencies have been effectively modeled by graph-based deep neural networks, such as graph convolutional neural networks, where each label is represented as a node in the graph.[7]

Toward producing more robust multilabel classifiers that are less prone to label noise, adversarial learning is utilized. Generative adversarial network models, which consist of a generator and a discriminator, can also be used in partial multilabel learning, where each instance is assigned multiple labels, but only some of them are relevant to the instance. The most common application of adversarial learning is in multilabel learning applications for image classification. Another type of generative model, the deep sequential generative model, has been used for weakly supervised multilabel classification, allowing the utilization of knowledge from unlabeled data or data with partially missing labels.

Interpretability

Interpretability is a hot topic in AI that concerns the ability of intelligent systems to justify their decisions. In domains where human lives are at stake, such as hate speech detection and health care, this ability is critical. Interpretability techniques can be either local or global, depending on the scope of the provided interpretation.

Local interpretations concern a particular prediction, while global interpretations present a general overview of the model. In multilabel learning, the explanation for a particular output of a model could concern either a single label or the full set of predicted labels.

Another distinction between interpretability techniques is their applicability, with model-agnostic techniques being equally usable in a wide range of models and model-specific techniques being designed for a particular type of model. One way to produce model-agnostic interpretations is to follow the surrogate model scheme and train a transparent model, either on the whole training set (global) or a group of neighbors to an examined input (local). Experimental results of a model-specific technique, focused on transformer models for multilabel text classification, suggest that providing a separate local interpretation per label yields more faithful interpretations.[8]

Other

There are several other interesting recent trends in multilabel learning that we cannot cover in this article. Based on a topic modeling of the titles of multilabel learning articles from the last three years, we would like to at least mention them. One is hierarchical multilabel classification. As mentioned in the concept evolution section, labels are often hierarchically related. This calls for methods that respect this knowledge. Several papers and articles discuss semisupervised multilabel learning methods to address the lack of ground truth as well as active multilabel learning methods, used when there is a limited budget to obtain ground-truth annotations. Finally, a number of papers and articles have addressed online learning and concept drift from multilabel data streams.

OPEN ISSUES

The horizontal nature of multilabel learning also comes with downsides. A lot of work that is related to this field has been conducted in isolation within different communities, such as computer vision, natural language processing, recommender systems, and bioinformatics, as well as in the separate context of different tasks, such as multivariate regression, multilabel classification, matrix completion, network inference, and dyadic prediction. Recent work on multitarget prediction—a generalization of multilabel learning where the target variables no longer need to be binary but can instead be numeric, ordinal, or even a mixture of types—has offered a unifying view of several different tasks that fall under its umbrella.[9] Multitarget data extend the already rich portfolio of application domains of multilabel data. Multitarget numeric data, for example, are commonly found in domains like ecology, energy, and sales, mainly in the context of time series forecasting. We believe there is room for additional transfer of knowledge between communities and tasks, which could be enabled by interdisciplinary scientific workshops as part of AI conferences.

A related general open issue in AI is the development of hybrid approaches combining both knowledge-based symbolic techniques together with nonsymbolic, typically neural, techniques. In the context of multilabel learning, it would be interesting to have a principled way to combine background knowledge in terms of deterministic relationships among an ontology of labels and data with instances of particular relationships among labels and among input variables and labels.[10] This will require the combination of expertise in several AI areas like knowledge graphs, machine learning, and constrained programming.

ACKNOWLEDGMENTS

This work was partially supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project 514).

REFERENCES

1. F. Markatopoulou, G. Tsoumakas, and I. Vlahavas, "Dynamic ensemble pruning based on multi-label classification," *Neurocomputing*, vol. 150, pp. 501–512, Feb. 2015, doi: 10.1016/j.neucom.2014.07.063.
2. J. Read, "From multi-label learning to cross-domain transfer: A model-agnostic approach," 2022, *arXiv:2207.11742*.
3. Y. Prabhu and M. Varma, "FastXML: A fast, accurate and stable tree-classifier for extreme multi-label learning," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2014, pp. 263–272, doi: 10.1145/2623330.2623651.
4. B. Liu, K. Blekas, and G. Tsoumakas, "Multi-label sampling based on local label imbalance," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108294, doi: 10.1016/j.patcog.2021.108294.
5. S. P. Veeranna, J. Nam, E. L. Mencía, and J. Fürnkranz, "Using semantic similarity for multi-label zero-shot classification of text documents," in *Proc. 24th Eur. Symp. Artif. Neural Netw. (ESANN)*, 2016, pp. 423–428.
6. N. Mylonas, S. Karlos, and G. Tsoumakas, "A multi-instance multi-label weakly supervised approach for dealing with emerging MeSH descriptors," in *Proc. 19th Int. Conf. Artif. Intell. Med. (AIME)*, 2021, pp. 397–407, doi: 10.1007/978-3-030-77211-6_47.
7. Z.-M. Chen, X.-S. Wei, P. Wang, and Y. Guo, "Multi-label image recognition with graph convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 5177–5186, doi: 10.1109/CVPR.2019.00532.
8. N. Mylonas, I. Mollas, and G. Tsoumakas, "An attention matrix for every decision: Faithfulness-based arbitration among multiple attention-based interpretations of transformers in text classification," 2022, *arXiv:2209.10876*.
9. W. Waegeman, K. Dembczyński, and E. Hüllermeier, "Multi-target prediction: A unifying view on problems and methods," *Data Mining Knowl. Discovery*, vol. 33, no. 2, pp. 293–324, Mar. 2019, doi: 10.1007/s10618-018-0595-5.
10. C. Papagiannopoulou, G. Tsoumakas, and I. Tsamardinos, "Discovering and exploiting deterministic label relationships in multi-label learning," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2015, pp. 915–924, doi: 10.1145/2783258.2783302.

NIKOLAOS MYLONAS is a Ph.D. student at Aristotle University of Thessaloniki (AUTH), Thessaloniki, 54124, Greece. His research interests include zero-shot and weakly supervised learning as well as interpretability in the context of multilabel learning. Mylonas received his M.Sc. degree in artificial intelligence from AUTH. Contact him at myloniko@csd.auth.gr.

IOANNIS MOLLAS is a Ph.D. student at Aristotle University of Thessaloniki (AUTH), Thessaloniki, 54124, Greece. His research interests include explainable artificial intelligence and, specifically, interpretable machine learning. Mollas received his M.Sc. degree in artificial intelligence from AUTH. Contact him at iamollas@csd.auth.gr.

BIN LIU is a lecturer at Chongqing University of Posts and Telecommunications, Chongqing, 400065, China. His research interests include multilabel learning, class imbalance, and biological data mining. Liu received his Ph.D. degree in computer science from Aristotle University of Thessaloniki. Contact him at liubin@cqupt.edu.cn.

YANNIS MANOLOPOULOS is a professor at the Open University of Cyprus, Nicosia, 2220, Cyprus, and professor emeritus of Aristotle University of Thessaloniki (AUTH). His research interests include data management and related issues. Manolopoulos received his Ph.D. degree in computer engineering from AUTH. He is a member of Academia Europaea. Contact him at yannis.manolopoulos@ouc.ac.cy.

GRIGORIOS TSOUMAKAS is associate professor at Aristotle University of Thessaloniki (AUTH), Thessaloniki, 54124, Greece. His research interests include multitarget prediction, interpretability, and summarization. He received his Ph.D. in computer science from AUTH. He is a Senior Member of IEEE and ACM. Contact him at greg@csd.auth.gr.