Multi-Label Classification: Status and Challenges

Chih-Jen Lin

National Taiwan Univ.

MBZUAI

November, 2023

Outline

Outline

Multi-label Classification

- Multi-label classification is an important technique that has been used in many Artificial Intelligence applications.
- Different from binary/multi-class classification, now an instance may be associated with multiple (e.g., zero to many) labels

Multi-label Classification (Cont'd)

- Applications include text classification, node classification, and many others
- The following example, from the well known RCV1 data set (Lewis et al., 2004), shows Reuters news documents and their labels.

```
C24 CCAT<TAB>uruguay uruguay compan ...
C151 C15 CCAT<TAB>spun stak compan ...
```

- Here <TAB> separates labels and texts. The first document is associated with two labels C24 and CCAT, while its text includes a sequence of words.

Application: ICD (International Classification of Diseases) Prediction

ICD-9 codes	ICD-10 codes
530.81: Esophageal reflux 244.9: Unspecified acquired hypothyroidism 272.4: Other and unspecified hyperlipidemia 300.00: Anxiety state, unspecified 401.9: Unspecified essential hypertension 278.00: Obesity, unspecified	A084: Viral intestinal infection, unspecified E271: Primary adrenocortical insufficiency E860: Dehydration E039: Hypothyroidism, unspecified F329: Major depressive disorder, single episode, unspecified
… with trauma to left upper limb past medical history pmh htn diverticulitis mitral regurgitation reflux fatty liver and anxiety psh total abdominal hysterectomy open cholecystectomy open appendectomy umbilical hernia and repair of incarcerated incisional hernia social history ___ family history mother stomach cancer physical exam heent wnl pulse regular lungs clear mental status wnl left upper limb demonstrated swelling purpura tenderness pain with motion and gross dysfunction pertinent results xray positive for comminuted proximal humerus fracture left brief hospital course admitted to ___after surgey no post operative problems dc to home with husband ___ on ___ nexium statin levothyroxine amlodipine discharge medications same plus oxycodone discharge disposition home discharge diagnosis comminuted fracture of proximal humerus into glenohumeral joint discharge condition stable both physically and …	fter the patient came to the floor she had no further episodes of vomiting or diarrhea she slept well overnight and by next morning she felt well was able eat both breakfast and lunch without issue her vs remained stable and she was discharged in good condition we attributed this to a viral gastreoenteritis that led to inability to take hydrocortisone and adrenal crisis we discussed follow up with her endocrinologist to discuss what she might do in another situation given this is the second time this has happened within the year medications on admission the preadmission medication list is accurate and complete bupropion mg po daily fludrocortisone acetate mg po daily hydrocortisone mg po daily levothyroxine sodium mcg po daily discharge medications hydrocortisone mg po q8h please higher dose tid for three days and then go back to your home dose bupropion mg po daily fludrocortisone acetate mg po daily levothyroxine sodium mcg po daily …

source: `https://arxiv.org/pdf/2304.13998.pdf`

Status of Multi-Label Classification

- Multi-label classification has been well studied in the past several decades
- However, in the big-data and deep-learning era, we find that users still face many difficulties to apply this technique.
- An issue is that tools to easily and conveniently solve users' problems are somewhat lacking

Status of Multi-Label Classification (Cont'd)

- Multi-label classification is complicated due to the many possible scenarios
- For example, situations for data with few labels (e.g., $\leq 1,000$) may be very different from those with millions of labels
- There are no definitive recipes to guide users to handle all different scenarios

The Project LibMultiLabel

- This is an on-going development for multi-label classification, with a focus on text problems

  `https://www.csie.ntu.edu.tw/~cjlin/libmultilabel`

- It is a simple tool with the following functionalities.
  - end-to-end services from raw texts to final evaluation/analysis
  - support for common neural network architectures and linear classifiers
  - easy hyper-parameter selection

Technical Challenges

Many technical challenges remain to researched for multi-label classification. For example,

- The label set may be huge. This causes lengthy training/prediction time, and huge model size. This topic is called extreme multi-label learning
- Some labels rarely appear or never appear in the training sets. Building a model to predict them is challenging

Outline

Extreme Multi-Label Learning

- Recently in many application domains, the number of labels is huge
- The situation is referred to as extreme multi-label learning

Extreme Multi-Label Learning (Cont'd)

- When number of labels is small, multi-label classification is usually considered together with binary/multi-class classification
- When number of labels is huge, the situation becomes very different

Memory and Time Difficulties

- If k is the number of labels, typically a multi-label method gets k decision functions:

$$f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \dots, f_k(\boldsymbol{x}))$$

- In prediction, we have

$$f_j(\boldsymbol{x}) = \begin{cases} \geq 0 & \text{has label } j \\ < 0 & \text{has not} \end{cases}$$

- But we cannot afford the time and memory cost as k goes large

Memory and Time Difficulties (Cont'd)

- Many effective methods have been developed. In general, we may need to (recursively) cluster labels to different groups
- I will not discuss details
- Instead, in the rest of this talk I will instead discuss the connection to similarity learning
- From the connection we may borrow ideas to get effective approaches

Issues about Zero Shots

- With many labels, easily some labels may appear only in the test set
- These unknown labels are often called zero-shots
- If we know nothing about these unseen labels, there is no way to build a model for predicting them
- Therefore, for applications of extreme multi-label learning, usually we need that labels come with some information (called label features or descriptions)

Issues about Zero Shots (Cont'd)

- For the earlier example,

  `C24 CCAT<TAB>uruguay uruguay compan ...`
  `C151 C15 CCAT<TAB>spun stak compan ...`

  we need that some descriptions are available for labels `C24`, `CCAT`, etc.
- With label descriptions, we show that extreme multi-label learning becomes a special case of similarity learning

Outline

Similarity Learning

- This task learns a similarity function from pairs of two entities, together with binary labels in $\{1, 0\}$ that determine if these pairs are similar or not
- Similarity learning is the key task for many applications
- An important example is the recommender systems
  - The two entities are a user and an item
  - The function value is if the user likes the item

Similarity Learning (Cont'd)

- Another important example is the question/answering systems
- Even large language models can be considered as similarity learning

Partially Labeled Data Sets

- With the huge number of pairs, in general only a subset of pairs are observed
- That is, between m left entities and n right entities, most are "?"
- Here we have that
  - 1: similar
  - 0: dissimilar
  - ?: unlabeled

	R_1	R_2	$\cdots$	$\cdots$	R_n
L_1	1	?	0	?	?
$\vdots$	?	?	?	1	1
L_m	?	0	?	?	1

Clicks/Non-clicks in Recommender Systems

- The advertising system displays a few items to each user and gets clicks/non-clicks
- However, user behaviors on non-displayed items are unknown
- Because a user is interested in only certain items, most "?" are indeed "0"

Clicks/Non-clicks in Recommender Systems (Cont'd)

- If we replace "?" and "0" with a constant $r \approx 0$, then we have

	R_1	R_2	$\cdots$	$\cdots$	R_n
L_1	1	r	r	r	r
$\vdots$	r	r	r	1	1
L_m	r	r	r	r	1

- This is usually called "recommender systems with implicit feedback"

Extreme Multi-Label Learning

- Now let's connect to extreme multi-label learning
- Consider

  left entities: instances,
  right entities: labels,
- Then extreme multi-label learning actually learns the

  similarity between an instance and a label

Extreme Multi-Label Learning (Cont'd)

- Thus it looks like we have the following situation

	R_1	R_2	$\cdots$	$\cdots$	R_n
L_1	1	0	0	0	0
$\vdots$	0	0	0	1	1
L_m	0	0	0	0	1

- A comparison with recommender systems seems to show that there is no "?"
- However, we argue that extreme multi-label learning is related to recommender systems with implicit feedback

Extreme Multi-Label Learning (Cont'd)

- First, with millions of labels, it is very likely that "1" entries are chosen only from labels that have been checked by or are familiar to human annotators
- Thus entries which are not "1" should be "0" or "?"
- That is, for each instance, it's possible that only some but not all relevant class labels have been identified

Extreme Multi-Label Learning (Cont'd)

- Second, in extreme similarity learning, left/right entities may change
- The same situation occurs for multi-label classification, where some labels may appear only in the test set
- Then we do have "?" entries

Extreme Multi-Label Learning (Cont'd)

- Then we have the following situation

	R_1	R_2	$\cdots$	$\cdots$	R_n	R_{n+1}
L_1	1	0	0	0	0	?
$\vdots$	0	0	0	1	1	?
L_m	0	0	0	0	1	?

Similarity Learning and Large Language Models

- Recently LLMs (Large Language Models) draw lots of attention
- Interestingly, we can make connections to similarity learning
- Large language models like GPT-1 (Radford et al., 2018) are trained on trillions of tokens by estimating

$$\Pr(\text{next token} \mid \text{pre-context})$$

- This is a process of learning one part of the input from another part of input

Discussion

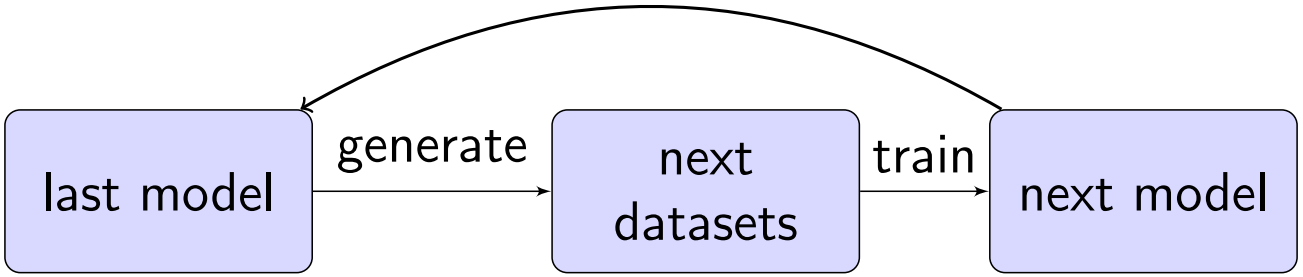
- When number of labels is small, multi-label classification is closely related to binary/multi-class classification
- But when number of labels is huge, the label set may change. The situation becomes closer to recommender or question/answering systems, which can all be explained under the framework of extreme similarity learning

Discussion (Cont'd)

- Interestingly, these areas, though all are special case of similarity learning, were not well linked so far
- An interesting research direction is to check if techniques for one area can be useful for another

Discussion (Cont'd)

- We demonstrate an example by showing that issues faced in recommender systems may also occur to LLMs
- For recommender systems, a serious issue is the feedback loop due to using recommendations from the last model as data to train the next model



Discussion (Cont'd)

- This issue leads the recommendation model to focus on only a few popular items while ignores the others gradually
- Large language models rely on huge data, but as time goes by, a similar issue may occur
- Data used for training large language models is mainly crawled from the Internet
- With the wide spread of these models, more data on Internet will be generated by LLMs but not human

Discussion (Cont'd)

- Unlike GPT-1 trained only on human-generated data, GPT-$\{n\}$ will be trained with datasets polluted by model-generated data
- A recent work (Shumailov et al., 2023) has shown that iterating models in this way results in a progressive loss on the knowledge of the original distribution
- Techniques to deal with the feedback loop in recommender systems might be utilized to check the new concern in the area of large language models

Outline

Conclusions

- Multi-label text classification is complicated due to the many possible scenarios
- It is essential to develop easy-to-use tools for users
- When number of labels is large, the problem is related to many areas under the framework of similarity learning. Techniques may be shared between these areas

We are Hiring at MBZUAI

- Research assistant: anyone without a Ph.D. degree is eligible
  - Minimum duration: six months
  - Example: a Ph.D. student who wants to do a six-month intern
  - Example: one has a B.S./M.S. degree but needs a short-term position before moving forward

We are Hiring at MBZUAI (Con'd)

- M.S./Ph.D. students: You can apply now. Talk to me or other faculty members if you are interested in working with us