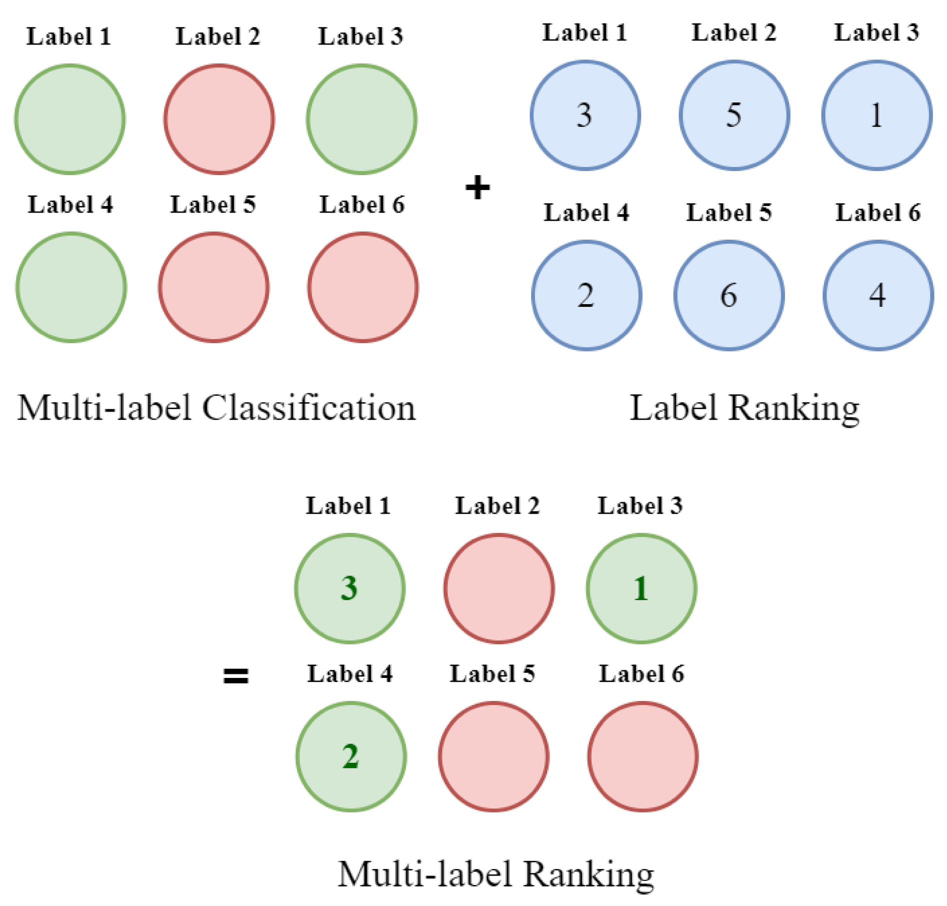
LEARNING LABEL RANKS FOR MULTI-LABEL CLASSIFICATION

Emine Darı 150180034

Introduction

Multi-label ranking (MLR) [1] is a supervised learning problem where the objective is to not only classify the labels by their relevancy to the instance, but also to predict a ranking that represents the instance's preferences over chosen relevant labels. MLR can be considered as generalization of the two sub-problems: multi-label classification and label ranking, as visualized in the diagram below.



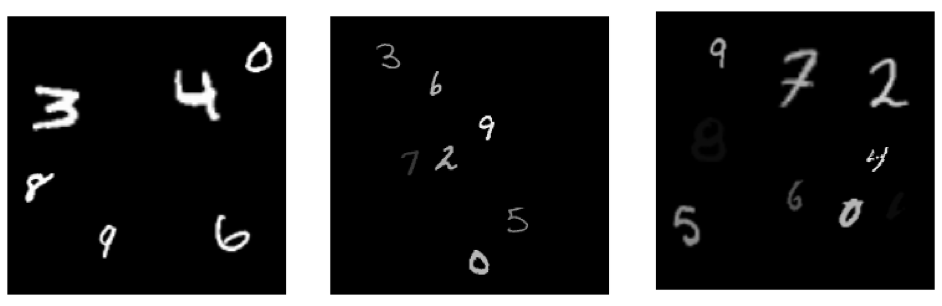
Motivation

What GaussianMLR effectively does is decomposing the input data into its components with their respective importance . When the suitable dataset is available, it can be used to extract rich semantic information, which is not feasible otherwise.

Contributions

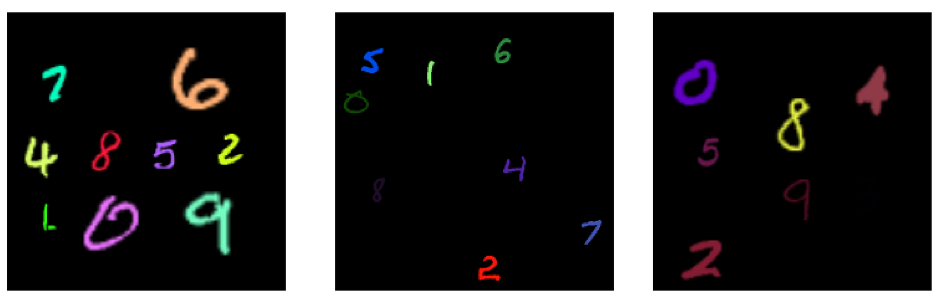
- Instead of assuming the positive and negative labels of equal importance, we establish the paradigm of exploiting the ranking information between positive labels of an instance.
- We model MLR as a distribution learning problem based on a probabilistic approach that unifies the bi-partition and ranking of the labels in the same space, and incorporate positive label pairs into the optimization.
- We introduce 8 ranked image datasets, with varying importance factors for label ranks, which create a controllable environment to test new approaches.

Ranked MNIST Gray



Scale Brightness Mix

Ranked MNIST Color



Scale Brightness Mix

- We compare our novel framework with different related methods, interpret the outcomes empirically, and set up a clear set of baselines for the MLR problem.

Problem Definition

We construct the MLR problem assuming that the two sub-problems are independent, i.e., $\mathcal{Y}^{(i)}$ and $\mathcal{B}^{(i)}$ are independent given an instance $\boldsymbol{x}^{(i)}$:

$$P_{\mathcal{Y},\mathcal{B}}(\mathcal{Y}^{(i)}, \mathcal{B}^{(i)} | \boldsymbol{x}^{(i)}; \theta) =$$

$$P_{\mathcal{Y}}(\mathcal{Y}^{(i)} | \boldsymbol{x}^{(i)} \theta) P_{\mathcal{B}}(\mathcal{B}^{(i)} | \boldsymbol{x}^{(i)}; \phi).$$

Then, we formally define the following likelihood optimization problem:

$$\max_{\theta, \phi} P_{\mathcal{Y}}(\mathcal{Y}^{(i)} | \boldsymbol{x}^{(i)}; \theta) P_{\mathcal{B}}(\mathcal{B}^{(i)} | \boldsymbol{x}^{(i)}; \phi).$$

$P_{\mathcal{Y}}$: parameterized family of probability mass functions of possible positive class sets

$\mathcal{Y}^{(i)}$: set of positive classes for an instance

$\mathcal{B}^{(i)}$: bucket order, a class of partial orders allowing ties.

**Please see the paper for futher notation details.*

Theory of GaussianMLR

The objective function of our proposed method GaussianMLR is given by:

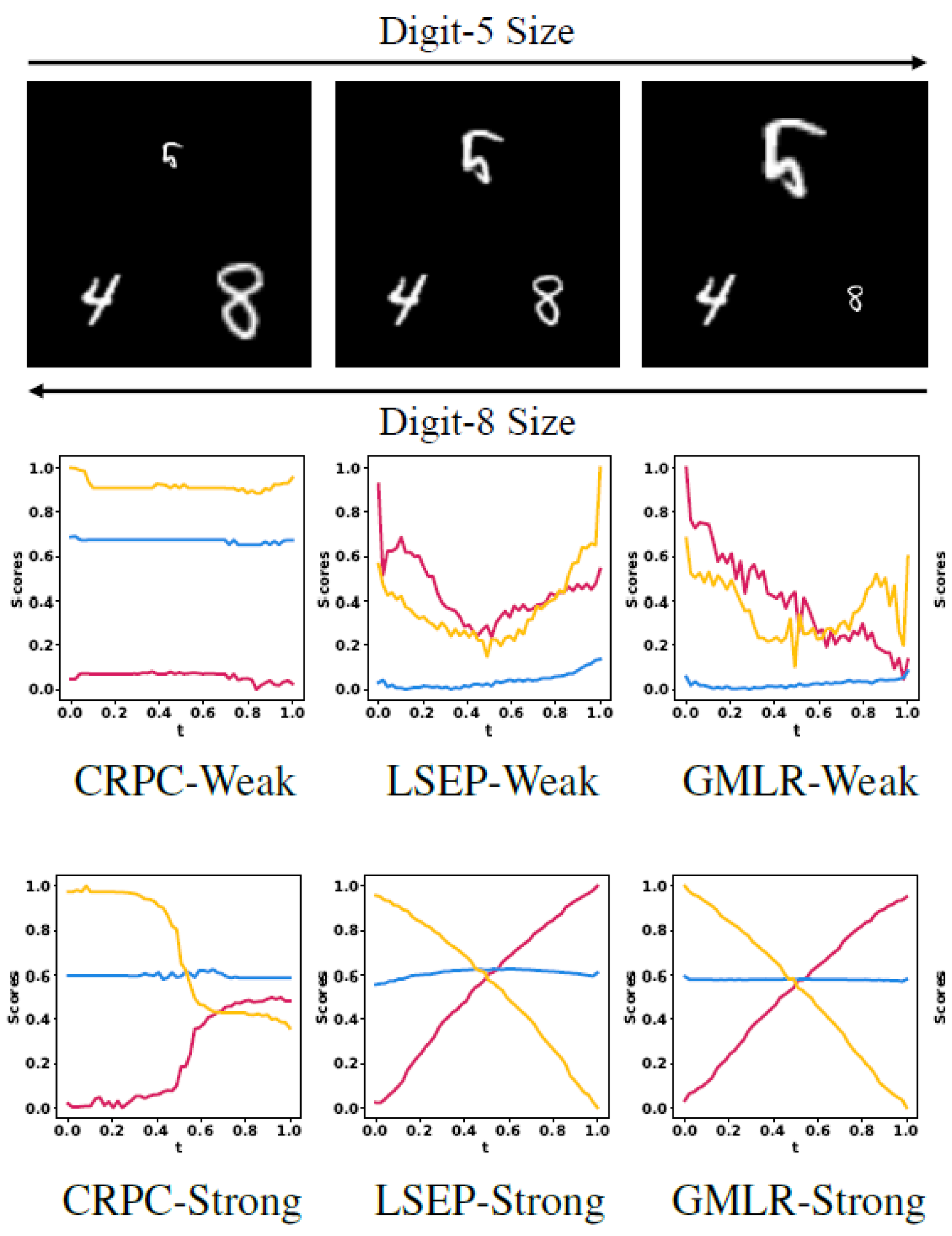
$$\min_{\zeta} \frac{1}{N} \sum_{i=1}^{N} L_c(\hat{\mu}^{(i)}, \hat{\sigma}^{(i)}, \mathcal{Y}^{(i)}) + L_r(\hat{\mu}^{(i)}, \hat{\sigma}^{(i)}, \mathcal{B}^{(i)}),$$

where L_c is the loss function for classification and L_r is for ranking. We assume all positive classes for an instance has an underlying significance value, modeled by Gaussian distributions, such that: $s_j^{(i)} \sim \mathcal{N}(\mu_j^{(i)}, \sigma_j^{2(i)})$.

Our goal is to obtain $f(\boldsymbol{x}^{(i)}; \zeta) = [\hat{\mu}^{(i)} \; \hat{\sigma}^{2(i)}]$ such that it produces significance values matching with the ranking information in the ground truth.

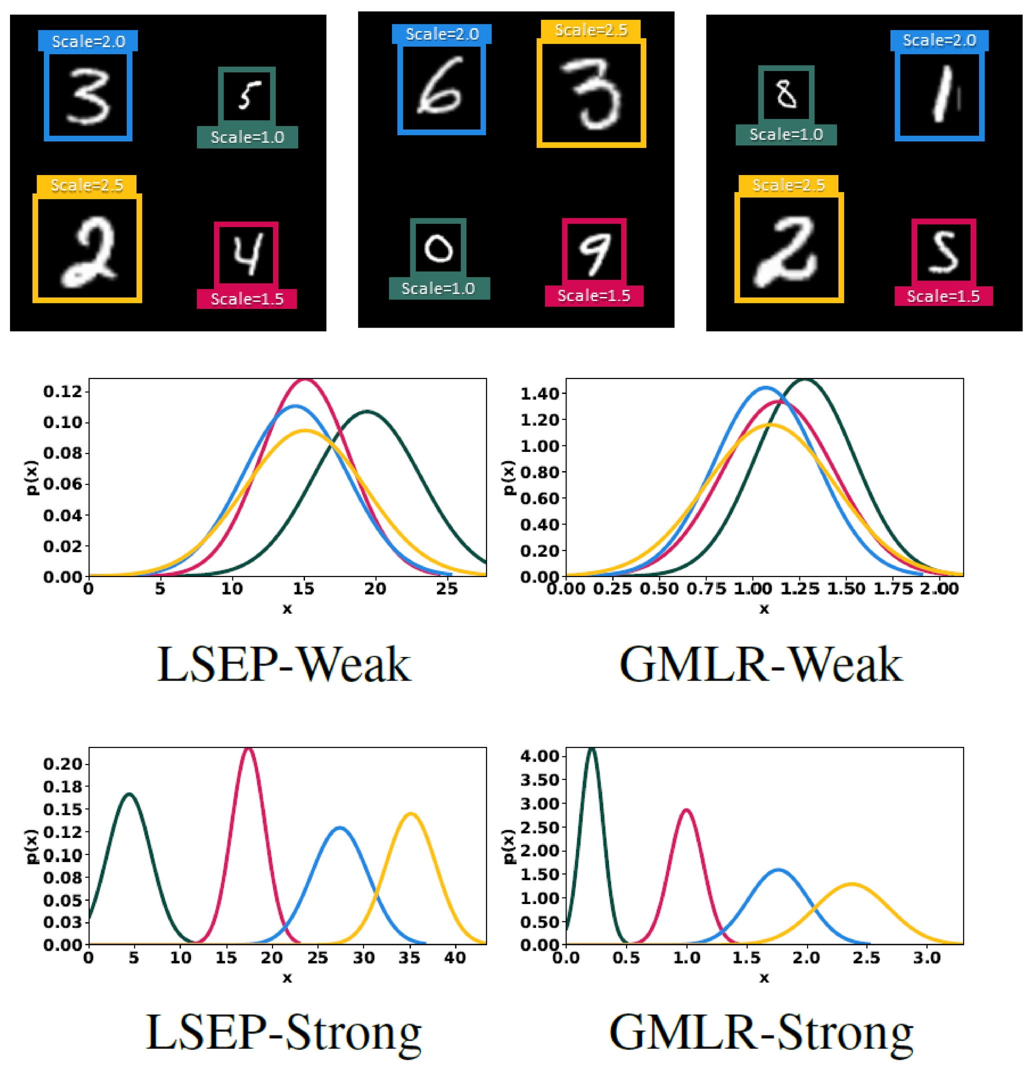
Experiments

A. Adjusting Significance Effects



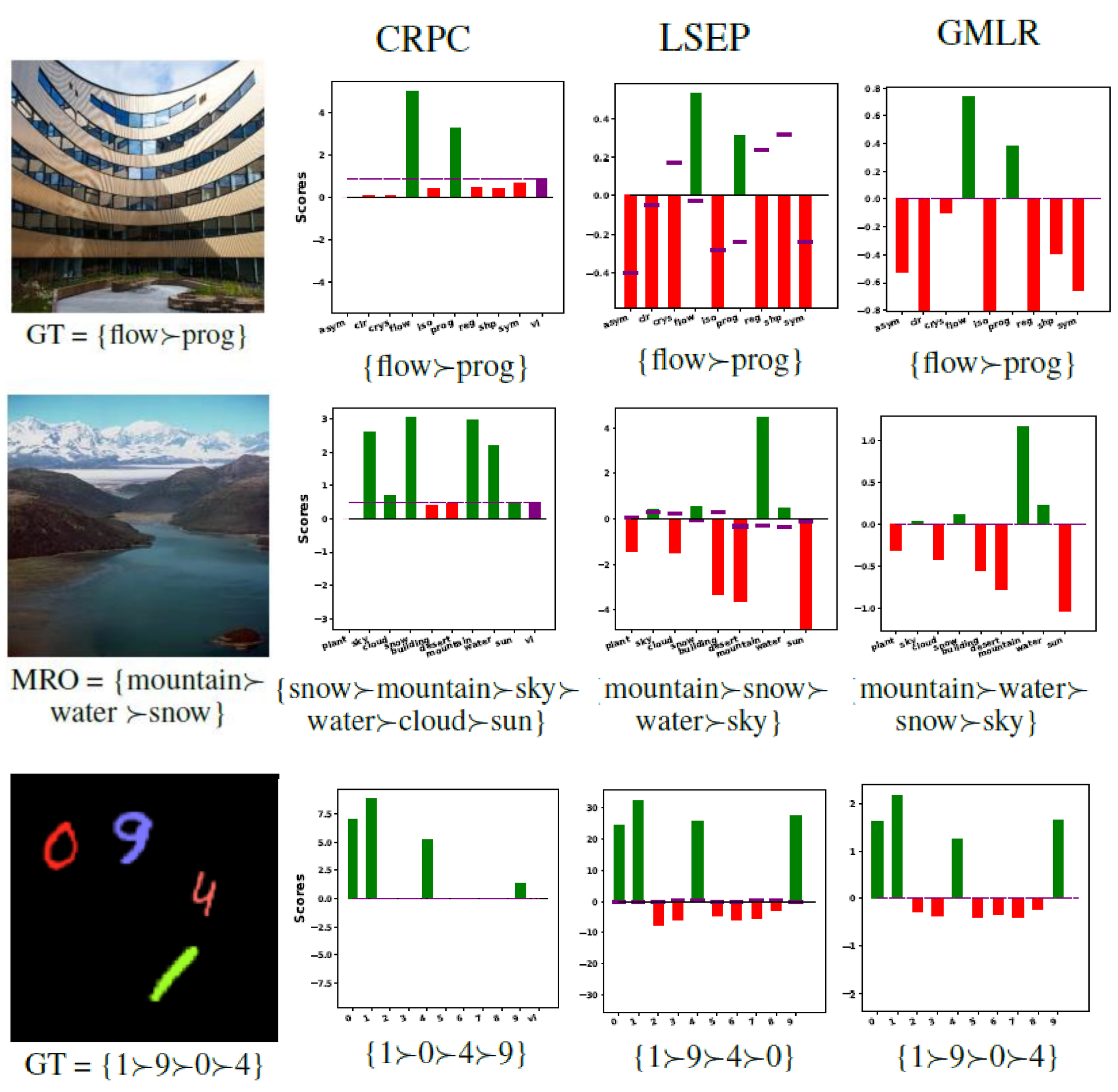
Gradually changing the size of the digits, where lines show changes in scores of ⟨1st, 2nd, 3rd⟩ digits, which are ⟨5,4,8⟩.

B. Calibration



Bounding boxes show the scale factors of **Scale=1.0**, **Scale=1.5**, **Scale=2.0**, **Scale=2.5** for each digit. Set of scores for each significance value (scale) is fit to a Gaussian distribution in the plots for the 4-labeled images.

C. Bar Graphs of Predicted Scores



Bar graphs of the predictions of strong versions of CRPC [3], LSEP [5] and GMLR are given for AVDP [28], NSID [3] and Ranked MNIST Color, to also visualize the working principles of each baseline.

Conclusion

While previously studied weak MLR methods learn almost no useful information about the underlying significance values of the positive classes, the strong MLR paradigm of GMLR yields remarkably calibrated significance values. GMLR compares favorably to the competing baselines as demonstrated by the experiments, where it is indicated that the underlying appearance and geometric characteristics pertinent to ranking are learned by GMLR.

References

Please see the paper via the QR Code for the references, which are numbered accordingly in this poster, as well as for details of notations, other experiments and further explanation.

Faculty of Computer and Informatics Engineering
İTÜ Ayazaga Campus, 34469, Maslak/Istanbul
bbf@itu.edu.tr
http://bm.itu.edu.tr