Predicting Label Distribution from Ternary Labels

Yunan Lu, Xiuyi Jia

luyn@njust.edu.cn, jiaxy@njust.edu.cn

Nanjing University of Science and Technology

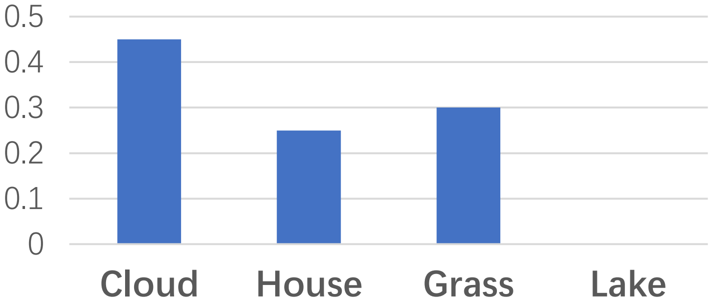
October 2024

Motivation

- LDL (Label Distribution Learning): An effective learning paradigm for addressing label polysemy.

Instance

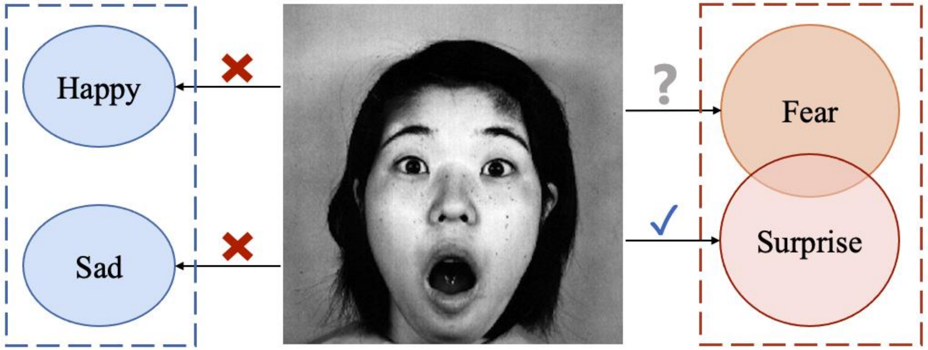
Label Distribution



Bottleneck → Annotating the ground-truth label distributions.

- LE (Label Enhancement): Infer label distributions from the more easily accessible multi-label data.

- Multi-label data relies on binary annotations, i.e., utilizing binary values ±1 to annotate each label.



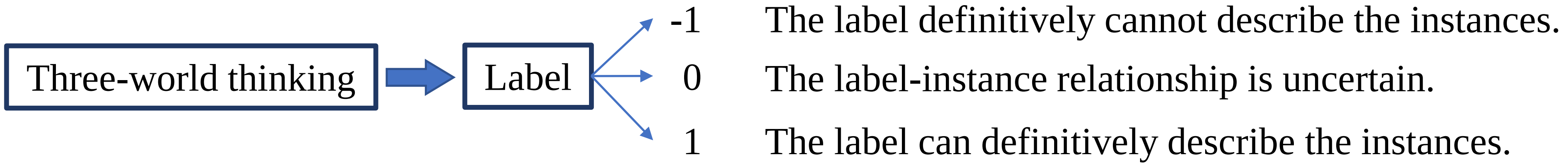
Label	Annotation
Happy	irrelevant (−1)
Sad	irrelevant (−1)
Surprise	relevant (+1)
Fear	uncertain (?)

Challenge →

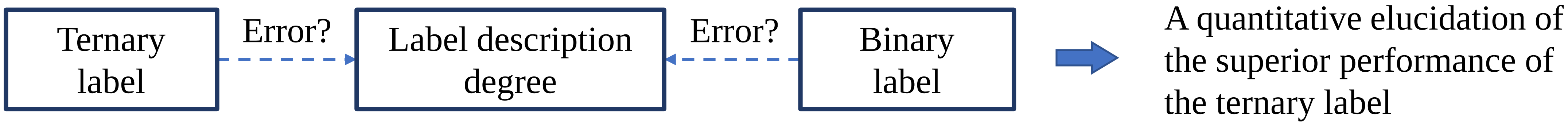
Accurate identification is challenging in real-world tasks due to the prevalence of ambiguous instances and labels that provoke uncertain binary associations.

Contribution

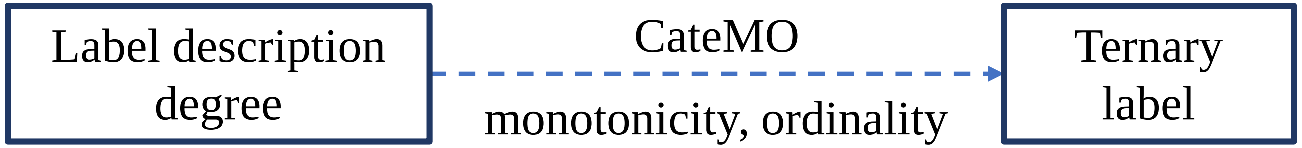
- We propose to predict label distribution from ternary labels: Enhanced annotation accuracy and reduced annotating cost.



- We rigorously analyze the error of approximating the ground-truth label description degrees by ternary and binary labels.



- We propose the CateMO distribution specifically designed to capture the mapping from label description degrees to ternary labels, which is theoretically constructed to maintain the monotonicity and ordinality of the probabilities associated with ternary labels.



Quantitative Analysis

- We suppose that the relationship between ternary/binary label s/b and label description degree z as follows.

Ternary Label	Binary Label
$s=-1 \rightarrow z \in [0,\tau)$	$b=-1 \rightarrow z \in [0,\xi)$
$s=0 \rightarrow z \in [\tau,\kappa]$	
$s=+1 \rightarrow z \in (\kappa,1]$	$b=+1 \rightarrow z \in [\xi,1]$

We use $\hat{\tau},\hat{\kappa},\hat{\xi}$ as the estimation of τ,κ,ξ.

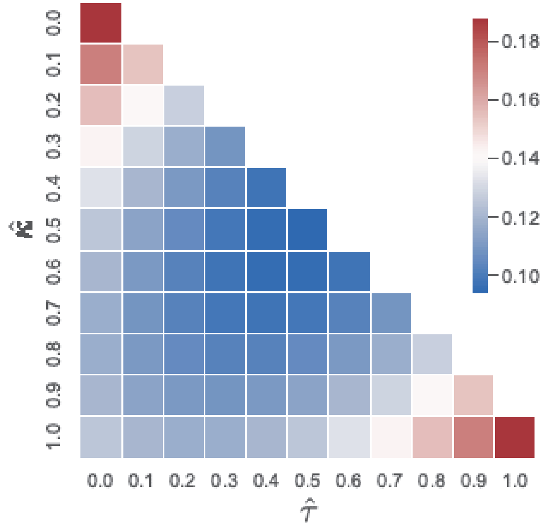
- We rigorously analyze the error of approximating the ground-truth label description degrees by ternary and binary labels.

$$\mathbb{E}_{\hat{s},s}[\psi(\mathcal{I}_{\hat{s}},\mathcal{I}_s)] = \frac{2}{9}(\tau+\kappa)^2 + \frac{2}{9}(\hat{\tau}+\hat{\kappa})^2 - \frac{1}{6}(\hat{\tau}\kappa+\hat{\kappa}\tau) - \frac{1}{3}(\hat{\tau}+\kappa)(\hat{\kappa}+\tau) + \frac{1}{18}(1-\kappa-\hat{\kappa})$$

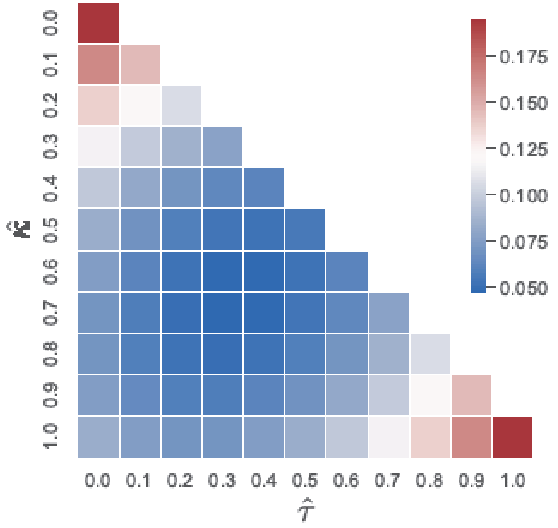
$$\mathbb{E}_{\hat{b},s}[\psi(\mathcal{I}_{\hat{b}},\mathcal{I}_s)] = \rho\left(\frac{\tau+\kappa}{6} - \frac{1+\hat{\xi}}{9}\right) + \frac{2(\tau+\kappa)^2}{9} + \frac{\hat{\xi}^2-\hat{\xi}\tau-\hat{\xi}\kappa-\tau\kappa}{3} - \frac{3\tau+4\kappa-\hat{\xi}-3}{18}$$

Quantitative Analysis

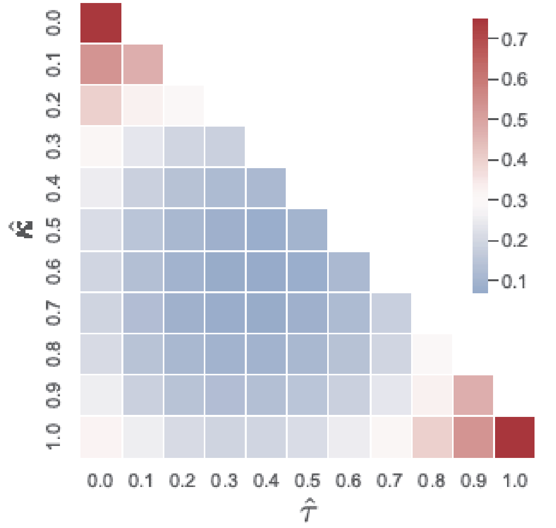
- Comparison of the approximation error of ternary and binary labels on the label description degree.



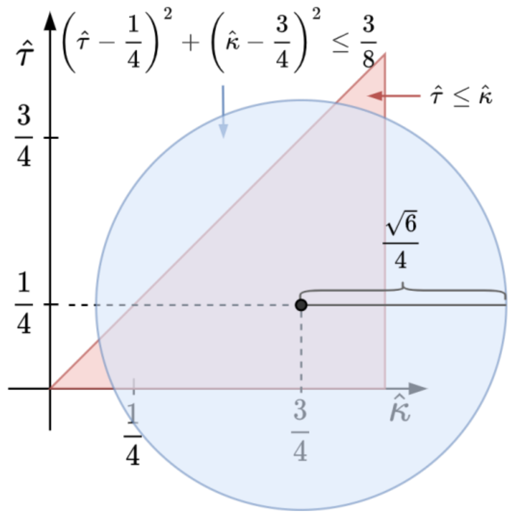
Approximation error of binary labels with random ξ



Approximation error of ternary labels with random κ, τ



Frequency with which the approximation error of ternary labels exceeds that of binary labels under random κ, τ

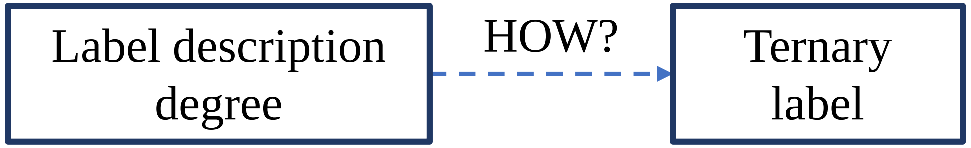


Conditions under which ternary label is superior to binary label on average (overlapping regions)

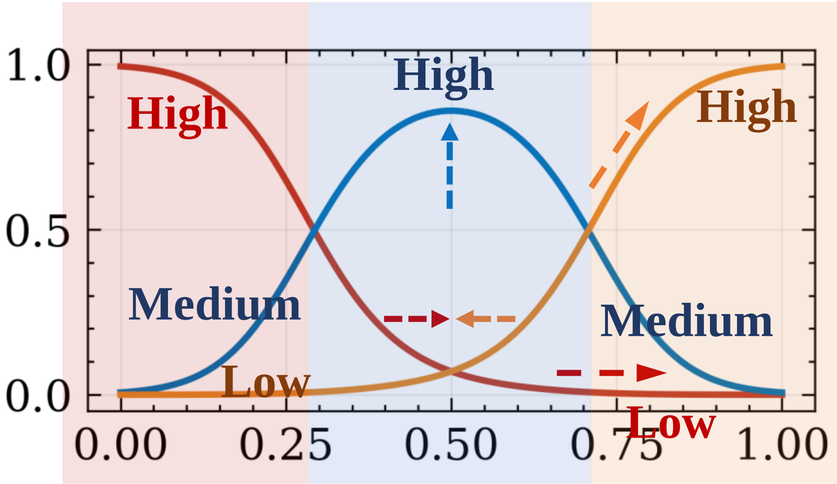
- Observation 1: The ternary label outperforms the binary label in most cases. Specifically, the binary label shows superiority only in the extreme cases where both the binary and ternary labels exhibit very high approximation error.
- Observation 2: Overlapping area is essentially consistent to the blue area.
- Conclusion: The ternary label is superior to the binary label w.r.t. approximating the ground-truth label description degrees.

Methodology

- We propose CateMO (Categorical distribution with Monotonicity and Orderliness) to model the conditional probability of ternary label given the LDD (label description degree).



- CateMO should maintain the monotonicity and orderliness of the probabilities of ternary labels.



- **Probability Monotonicity**
  - The larger the LDD, the higher the probability of positive label.
  - The larger the LDD, the lower the probability of negative label.
  - The closer the probabilities of negative and positive labels, the higher the probability of uncertain label.
- **Probability Orderliness**
  - Large LDD $\Rightarrow P(positive) > P(uncertain) > P(negative)$
  - Small LDD $\Rightarrow P(positive) < P(uncertain) < P(negative)$
  - Medium LDD $\Rightarrow P(positive), P(negative) < P(uncertain)$

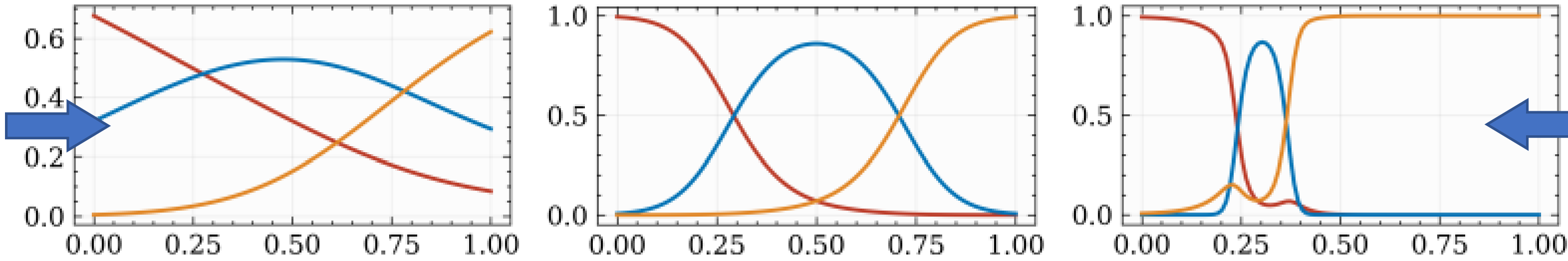
CateMO: Categorical distribution with Monotonicity and Orderliness

- The formula of CateMO: Extending the softmax function.

$$p(s=-1|z) \longrightarrow \underline{\varphi}(z) = \frac{1}{Z}e^{-\underline{\lambda} z^2} \qquad p(s=0|z) \longrightarrow \varphi(z) = \frac{1}{Z}e^{-\lambda(z-\hat{z})^2} \qquad p(s=1|z) \longrightarrow \overline{\varphi}(z) = \frac{1}{Z}e^{-\overline{\lambda}(z-1)^2}$$

- Shape of CateMO with different parameters.

Violate the Monotonicity Assumption



(a) $[\underline{\lambda}, \overline{\lambda}, \lambda, \hat{z}] = [2, 5, 3, \frac{1}{2}]$ (b) $[\underline{\lambda}, \overline{\lambda}, \lambda, \hat{z}] = [10, 10, 20, \frac{1}{2}]$ (c) $[\underline{\lambda}, \overline{\lambda}, \lambda, \hat{z}] = [30, 5, 500, \frac{3}{10}]$

Violate the Orderliness Assumption

- The condition that enables CateMO to satisfy the monotonicity and orderliness assumptions.

$$\lambda \neq -\underline{\lambda}\overline{\lambda}(\hat{z}\overline{\lambda} - \hat{z}\underline{\lambda} - \overline{\lambda})^{-1}, \hat{z} = (2\lambda\sqrt{\overline{\lambda}} + 2\lambda\sqrt{\underline{\lambda}})^{-1}(2\lambda\sqrt{\overline{\lambda}} - \underline{\lambda}\sqrt{\overline{\lambda}} + \overline{\lambda}\sqrt{\underline{\lambda}}),$$
$$\max\{(\hat{z} + \hat{z}e^{\overline{\lambda}})^{-1}\overline{\lambda}, ((1+e^{\underline{\lambda}})(1-\hat{z}))^{-1}\underline{\lambda}\} < \lambda < \min\{\underline{\lambda}(1-\hat{z})^{-1}, \overline{\lambda}\hat{z}^{-1}\}.$$

- Usage of CateMO.

$$\mathrm{Dist}(\boldsymbol{s}, \boldsymbol{z}) = -\sum_{m=1}^{M} \log \mathrm{CateMO}(s_m \mid z_m), \quad p(\boldsymbol{s}|\boldsymbol{z}) = \prod_{m=1}^{M} \mathrm{CateMO}(s_m \mid z_m).$$

Experiments

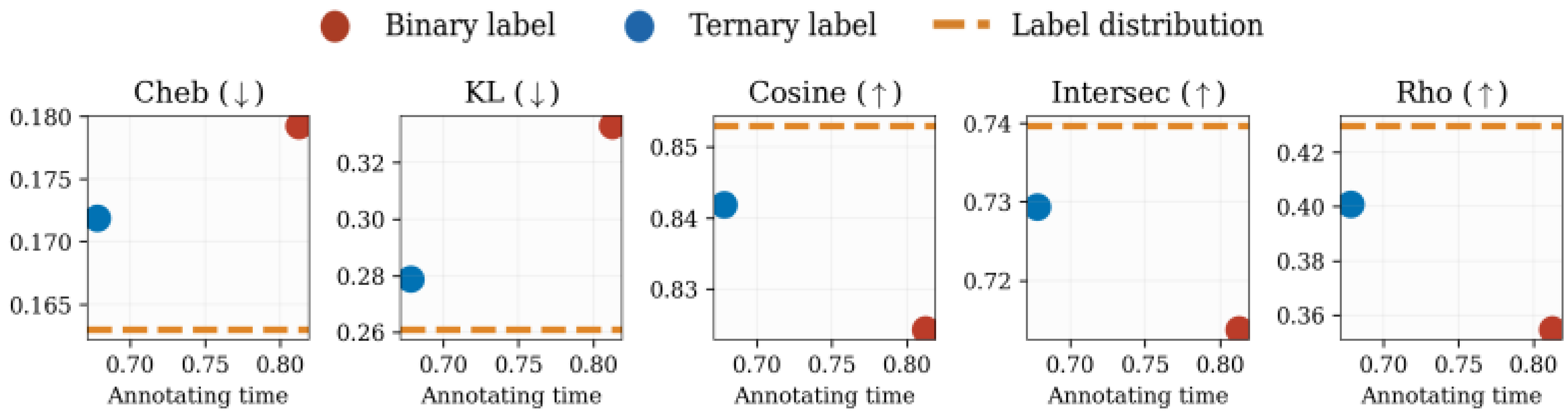
- Performance of predicting label distributions. Results are shown as "mean±std", where **bold** and *italics* denote the 1st and 2nd, respectively

Ground-Truth	0.145 ± 0.015	0.897 ± 0.009	0.535 ± 0.017	0.737 ± 0.007	0.103 ± 0.006	0.925 ± 0.004
MR-CateMO	**0.154 ± 0.015**	**0.892 ± 0.009**	**0.561 ± 0.022**	**0.723 ± 0.008**	**0.107 ± 0.007**	**0.921 ± 0.005**
MR-MSE	0.166 ± 0.017	0.884 ± 0.010	0.564 ± 0.029	*0.718 ± 0.013*	*0.108 ± 0.006*	*0.920 ± 0.004*
MR-DT	*0.163 ± 0.012*	*0.886 ± 0.006*	*0.562 ± 0.019*	*0.718 ± 0.009*	0.118 ± 0.009	0.910 ± 0.008
MR-LL	0.171 ± 0.014	0.883 ± 0.009	0.566 ± 0.019	0.715 ± 0.008	0.115 ± 0.009	*0.912 ± 0.008*
LR-CateMO	**0.156 ± 0.015**	**0.889 ± 0.010**	**0.545 ± 0.021**	**0.731 ± 0.008**	**0.103 ± 0.006**	**0.924 ± 0.005**
LR-MSE	*0.161 ± 0.017*	*0.884 ± 0.011*	0.575 ± 0.025	0.712 ± 0.011	*0.104 ± 0.006*	**0.924 ± 0.005**
LR-DT	*0.161 ± 0.017*	*0.884 ± 0.011*	*0.549 ± 0.025*	*0.730 ± 0.009*	*0.104 ± 0.007*	**0.924 ± 0.005**
LR-LL	0.320 ± 0.145	0.847 ± 0.023	0.617 ± 0.044	0.685 ± 0.018	0.150 ± 0.033	0.909 ± 0.008
GL-CateMO	**0.158 ± 0.016**	**0.887 ± 0.011**	**0.542 ± 0.022**	**0.734 ± 0.008**	**0.103 ± 0.007**	**0.924 ± 0.005**
GL-MSE	*0.162 ± 0.016*	*0.884 ± 0.011*	*0.571 ± 0.027*	0.715 ± 0.012	*0.104 ± 0.007*	**0.924 ± 0.005**
GL-DT	0.210 ± 0.023	0.857 ± 0.013	0.590 ± 0.026	*0.719 ± 0.011*	0.126 ± 0.007	0.912 ± 0.004
GL-LL	0.311 ± 0.156	0.852 ± 0.024	0.610 ± 0.055	0.710 ± 0.008	0.137 ± 0.009	0.906 ± 0.006

- The suffix "-LL" denotes that these algorithms run on binary labels directly.
- The suffixes "-DT", "-MSE" and "-CateMO" denote that these algorithms run on ternary labels by the DT method, MSE method and our proposed CateMO distribution, respectively.

Experiments

- Cost-benefit analysis of different forms of labels. The horizontal and vertical axes denote the average annotating time (in seconds) and performance, respectively



Thank you for your attention!

If you have any questions about our research, please contact us by sending an email to luyn@njust.edu.cn or jiaxy@njust.edu.cn.