Table 1: The results of Example-F1 on the EMOTIONS, SCENE and MEDICAL data sets.

DATA SET	LIMO	CML+GAU
EMOTIONS	0.5216	0.5215
SCENE	0.5376	0.5378
MEDICAL	0.1769	0.1767

Response to Reviewer #1

Q1: What is the performance of the proposed method on the image datasets? A: Thanks for your comments. We have conducted the experiment on the image dataset (scene). Moreover, we also test our methods on other data sets with different domains, such as text (medical, enron), biology (yeast) and music (emotions). The experiment results have shown the improved performance of our proposed methods on various domains.

Response to Reviewer #2

Q1: ...focusing on the marginals...which in theory (but in practice is) is not different from BR ...? A: Thanks for your comments. Theorem 1 indicates that copula allows a complete separation of dependence modeling from the marginal distributions and by specifying a copula one can summarize all the dependencies between margins.

Therefore, based on Theorem 1, we can use a $(p+q)$-copula function to summarize all the dependencies between labels and features. Based on $(p+q)$-copula function, we are able to derive marginal functions. Please note that the derived marginal functions have already inherited the dependence information from $(p+q)$-copula function. We have presented this point in lines 125-129.

Q2: ...loss functions...used in the experiments...comparison with F1-measure optimizers... A: Thanks for your comments. This paper optimizes the Hamming loss, and in the Supplementary Materials, Table 1 shows the results of Hamming loss on the various data sets.

LIMO (A Unified View of Multi-Label Performance Measures, ICML, 2017) is a state-of-the-art F1-measure optimizer. According to the reviewer's comments, we compare our proposed method with LIMO in terms of Example-F1 on the EMOTIONS, SCENE and MEDICAL data sets. The results are shown in Tabel 1. From Tabel 1, we can see that our proposed method is comparable to the F1-measure optimizer. Following the reviewer's comments, we will consider optimizing various loss functions in the future work.

Q3: whether a simple BR estimator also does not enjoy such similar properties. A: Thanks for your comments. One of the most important insights of Sklar's Theorem (Theorem 1) is that the univariate margins and the multivariate dependence structure can be separated, and the dependence structure can be represented by a copula. Therefore, by specifying a copula one can summarize all the dependencies between margins. Inspired by Sklar's Theorem, we develop a framework of copula multi-label learning to model label and feature dependencies. The theoretical analysis in this paper makes no assumptions on the specific copula functions. We can derive the same statistical properties for our proposed estimator with any copula functions. If there is a copula contain the independent information between the labels, then our theoretical results also hold in this special case (BR).

Response to Reviewer #3

Q1: This paper uses normal copula and student's copula, is it possible to use other copula functions? A: Thanks for your comments. The theoretical analysis in this paper makes no assumptions on the specific copula functions. We can derive the same statistical properties for our proposed estimator with any copula functions. In the experiment, we use multivariate normal copula and multivariate student's t copula as two examples to show the performance of our proposed method.

Q2: The paper may need to add and discuss the three references. A: Thanks for your comments. These papers focus on the applications, such as image classification, text classification and health evaluation. We will cite these references in the revisions.