

Supplementary material for:
“On the relative value of weak information of
supervision: An empirical study”

Jerónimo Hernández-González^{a,*}, Aritz Pérez^b

^a*Departament de Matemàtiques i Informàtica, Universitat de Barcelona (UB), Gran Via de
les Corts Catalanes 585, Barcelona, Spain*

^b*Basque Center for Applied Mathematics, Al. Mazarredo 14, Bilbao, Spain*

1. Additional figures

Following the same experimental setup than the one explained in the paper, we show more experimental results where the “correct model” assumption does not hold. For the following figures, data was synthetically generated from a generative model of large complexity ($K = 4$), whereas the learned model was simpler ($K = 1$):

- Figure 1 shows the contribution of weak supervision as the proportion of weakly labeled examples increases, and also as the size of the candidate sets increases.
- Figure 2 shows the contribution of weak supervision as the size of the labeled subset and that of the weakly supervised subset are increased.
- Figure 3 shows the contribution of weak supervision as the level of co-occurrence and the size of the candidate sets increase.
- Figure 4 shows the contribution of weak supervision as the level of co-occurrence and proportion of weakly labeled examples increase.

*Corresponding author

Email addresses: jeronimo.hernandez@ub.edu (Jerónimo Hernández-González),
aperez@bcamath.org (Aritz Pérez)

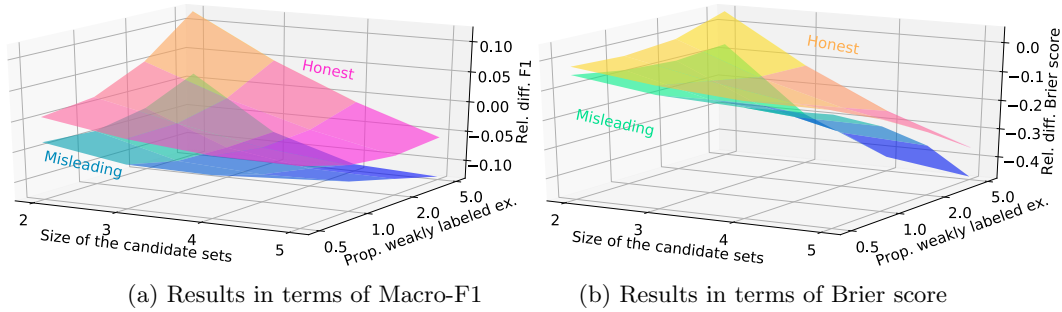


Figure 1: Graphical description of the value of weak supervision as the size of the candidate sets and the proportion of weakly labeled examples increase on synthetic data. In both figures, two surfaces display results with misleading (labels in S are really improbable) and honest (labels in S are probable) labels, respectively. All the surfaces show the difference of performance of a classifier learned with weak supervision minus that of a similar classifier learned only with the labeled subset; in terms of Macro-F1 in a), of Brier score in b), relative to the performance of the real model. Other parameters are fixed to a small fully labeled subset ($N_f = 33$), complex generative ($K = 4$) and simple learned ($K = 1$) models, and no induced co-occurrence ($s = 0$).

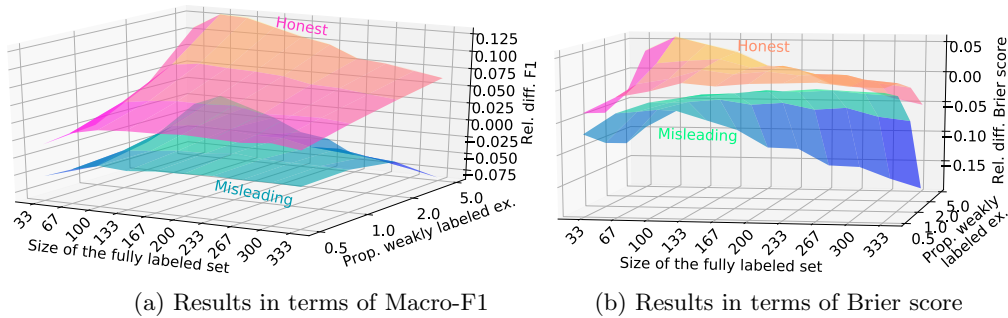


Figure 2: Graphical description of the value of weak supervision as the amount of fully labeled data and the proportion of weakly labeled examples increase on synthetic data. In both figures, two surfaces display results with misleading (labels in S are really improbable) and honest (labels in S are probable) labels, respectively. All the surfaces show the difference of performance of a classifier learned with weak supervision minus that of a similar classifier learned only with the labeled subset; in terms of Macro-F1 in a), of Brier score in b), relative to the performance of the real model. Other parameters are fixed to small candidate sets ($|S| = 2$), simple generative and learned models ($K = 1$), and no induced co-occurrence ($s = 0$).

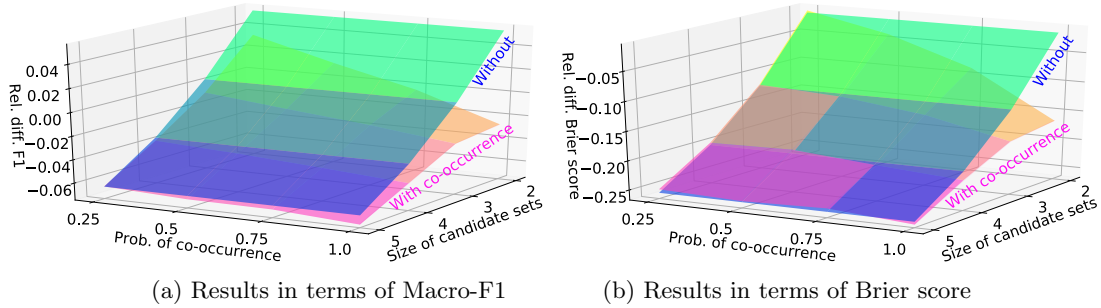


Figure 3: Graphical description of the value of weak supervision as the probability of co-occurrence in the candidate sets and the size of the candidate sets increase on synthetic data. In all figures, a surface shows results of experiments without inducing co-occurrence, whereas another surface shows results when inducing co-occurrence with increasing probability. All the surfaces show the difference of performance of a classifier learned with weak supervision minus that of a similar classifier learned only with the labeled subset; in terms of Macro-F1 in a), of Brier score in b), relative to the performance of the real model. Other parameters are fixed to a small fully labeled subset ($N_f = 33$), twice as weakly labeled samples ($r = 2$), and simple generative and learned models ($K = 1$).

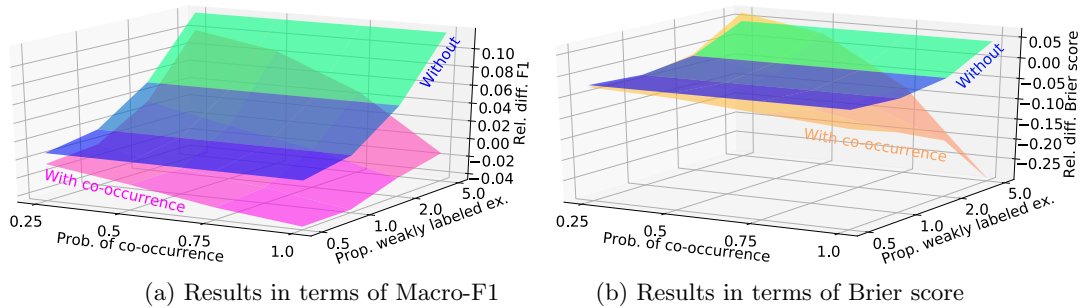


Figure 4: Graphical description of the value of weak supervision as as the probability of co-occurrence in the candidate sets and the proportion of weakly labeled examples increase on synthetic data. In all figures, a surface shows results of experiments without inducing co-occurrence, whereas another surface shows results when inducing consistent labels with increasing probability of co-occurrence. All the surfaces show the difference of performance of a classifier learned with weak supervision minus that of a similar classifier learned only with the labeled subset; in terms of Macro-F1 in a), of Brier score in b), relative to the performance of the real model. Other parameters are fixed to a small fully labeled subset ($N_f = 33$), candidate set size $|S| = 2$, and simple generative and learned models ($K = 1$).