

# Repondération Préférentielle pour l'Apprentissage Biqualité

Pierre Nodet<sup>\*,\*\*</sup>, Vincent Lemaire<sup>\*</sup> Alexis Bondu<sup>\*</sup> Antoine Cornuéjols<sup>\*\*</sup>

<sup>\*</sup> Orange Labs, Paris & Lannion, France

<sup>\*\*</sup> AgroParisTech, Paris, France

**Résumé.** Cet article propose une vision originale et globale de l'Apprentissage Faiblement Supervisé, menant à la conception d'approches génériques capable de traiter tout type de faiblesses en supervision. Un nouveau cadre appelé "Données Biqualité" est introduit, qui suppose qu'un petit jeu de données fiable d'exemples correctement étiquetés est disponible, en plus d'un jeu de données non fiable comprenant un grand nombre d'exemples potentiellement corrompus. Dans ce cadre nous proposons un nouveau schéma de repondération capable de détecter les exemples non corrompus du jeu de données non fiable. Cet algorithme permet d'apprendre des classifieurs sur les deux jeux de données. Nos expériences simulant plusieurs types de bruits d'étiquetage démontrent empiriquement que l'algorithme proposé sur-performe l'état de l'art.

## 1 Introduction

La classification supervisée a pour objectif d'apprendre un classifieur à partir d'exemples étiquetés qui est capable de prédire la classe d'exemples non vus pendant l'apprentissage. En pratique, les algorithmes classiques de classification se heurtent aux imperfections des jeux de données réels. Ainsi l'apprentissage faiblement supervisé à récemment vu un regain en popularité avec de nombreux articles traitant de plusieurs types de "*faiblesses en supervision*". Tous ces types de faiblesses en supervision sont traités séparément dans la littérature conduisant à des approches hautement spécialisées. En pratique il est pourtant très difficile d'identifier précisément les types de faiblesses en supervision que comporte un jeu de données. Cette article présente un nouveau cadre appelé données biqualités qui couvre un large spectre de faiblesses en supervision et permet une unification des méthodes d'apprentissage faiblement supervisé. Pour une lecture plus approfondie, se référer à une introduction sur l'apprentissage faiblement supervisé et sur les liens existants avec l'apprentissage biqualité dans (Nodet et al., 2021).

L'apprentissage sur des données biqualités a récemment été mis en lumière dans (Charikar et al., 2017; Hendrycks et al., 2018; Hataya et Nakayama, 2019), cela consiste à apprendre un classifieur à partir de deux jeux de données différents, l'un fiable et l'autre non fiable. Le but originel était d'unifier l'apprentissage semi-supervisé et robuste au bruit d'annotation par la combinaison des deux. Néanmoins ce scénario n'est pas limité à cette unification mais il peut couvrir un large spectre de faiblesses en supervision tel que démontré avec l'algorithme proposé et les résultats associés obtenus.

Le jeu de données fiable  $D_T$  ('T' pour « trusted ») est composé d'individus  $x_i$  associés aux étiquettes  $y_i$  formant des paires  $(x_i, y_i)$  où toutes les étiquettes  $y_i \in \mathcal{Y}$  sont supposées correctes