ULF: Unsupervised Labeling Function Correction using Cross-Validation for Weak Supervision

Anastasiia Sedova, Benjamin Roth

University of Vienna, Austria

Noise in Weakly Supervised Data

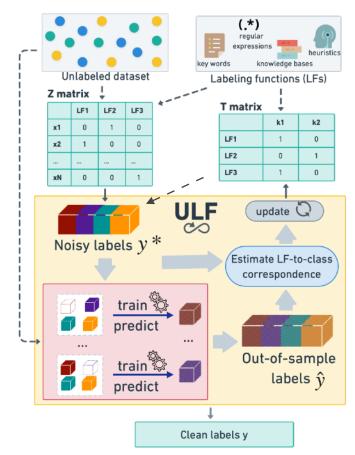| Sample	Matched Labeling Functions	Assigned Label	Gold Label
CHECK MY CHANNEL **OUT PLEASE**. I DO SINGING COVERS	keyword_my, keyword_please regex_check_out	SPAM	SPAM
It looks so real and **my** daughter is a big fan and she likes a lot of your **songs**.	keyword_my short_songs	SPAM/ HAM	HAM

Examples from the YouTube spam comments detection dataset (Alberto et al. 2015)

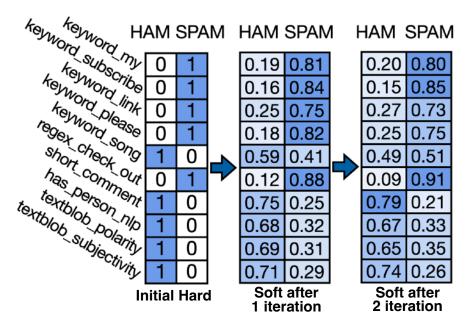- **Our goal:** change labeling functions (LFs) to class assignment from hard to soft

ULF - Unsupervised Labeling Function Correction

- ULF re-estimates LFs to class correspondance.

- Re-estimation is performed based on the out-of-sample predictions (without additional manual supervision).

- The out-of-sample predictions are calculated using cross-validation with splitting the data according to matched labeling functions

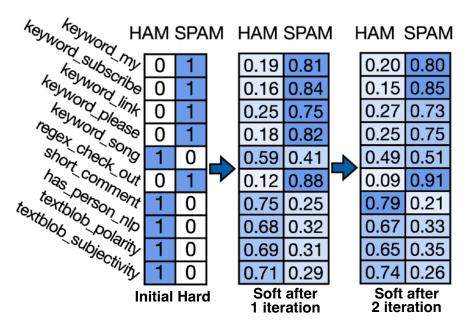- Only confident out-of-sample predictions are used for re-estimation.

ULF Results

	YouTube	Spouse	TREC	SMS	Yorùbá	Hausa	Avg
Gold	98.8	-	96.6	97.7	67.3	83.5	88.8
Majority Vote (MV)	93.2	21.3	68.6	93.0	48.1	43.9	61.4
MeTaL (Ratner et al., 2019)	96.0	19.6	55.8	89.2	58.6	41.6	60.1
Snorkel-DP (Ratner et al., 2020)	95.6	32.6	61.8	94.6	58.7	45.7	64.8
FlyingSquid (Fu et al., 2020)	94.0	14.9	35.8	23.7	32.4	45.1	41.0
WeaSEL (Cachay et al., 2021)	96.0	14.9	64.4	23.6	49.6	43.2	48.6
FABLE (Zhang et al., 2023)	94.8	27.8	54.6	91.1	23.2	18.6	51.7
MV + Cosine (Yu et al., 2021)	96.4	33.3	65.8	93.6	52.6	45.4	64.5
MeTaL + Cosine	95.6	26.9	67.4	80.7	**62.0**	45.5	63.0
Snorkel-DP + Cosine	96.0	28.1	73.8	96.1	55.0	46.5	65.9
FlyingSquid + Cosine	95.6	24.9	38.6	90.1	33.3	41.5	54.0
FABLE + Cosine	94.0	33.9	70.6	**97.7**	60.1	44.7	66.8
ULF (Ours)	**96.8**	**36.9**	**76.8**	96.2	55.8	**48.2**	**68.4**

ULF Results - Case Study

	HAM SPAM	HAM SPAM	HAM SPAM
keyword_my	0 | 1	0.19 | 0.81	0.20 | 0.80
keyword_subscribe	0 | 1	0.16 | 0.84	0.15 | 0.85
keyword_link	0 | 1	0.25 | 0.75	0.27 | 0.73
keyword_please	0 | 1	0.18 | 0.82	0.25 | 0.75
keyword_song	1 | 0	0.59 | 0.41	0.49 | 0.51
regex_check_out	0 | 1	0.12 | 0.88	0.09 | 0.91
short_comment	1 | 0	0.75 | 0.25	0.79 | 0.21
has_person_nlp	1 | 0	0.68 | 0.32	0.67 | 0.33
textblob_polarity	1 | 0	0.69 | 0.31	0.65 | 0.35
textblob_subjectivity	1 | 0	0.71 | 0.29	0.74 | 0.26
	Initial Hard	**Soft after 1 iteration**	**Soft after 2 iteration**

Sample	LFs matched	Noisy Label	Corrected Label	Gold Label
Hello! I'm Marian, I'm a singer from Venezuela! ... I wanted to play my own pop and pop-rock **songs**. It would mean a lot if you could have a look at **my** channel to check my music! and if u like, **subscribe** to it! XOXO THANKS!!	keyword_my keyword_subscribe keyword_song textblob_subjectivity	HAM	SPAM	SPAM
Nice **song** .See **my** new track.	keyword_my keyword_song textblob_subjectivity	HAM	SPAM	SPAM
Follow me on Twitter @mscalifornia95	—	HAM	SPAM	SPAM

ULF: Unsupervised Labeling Function Correction using Cross-Validation for Weak Supervision

Anastasiia Sedova, Benjamin Roth

University of Vienna, Austria