

230427 勉強会 ～半教師あり学習～

M1 上野

半教師あり学習とは

従来のDLは教師あり学習

- すべてのデータにアノテーションするのは大変
- アノテーション作業を減らしたい
 - データの一部にアノテーションしよう = 半教師あり学習

教師あり学習



半教師あり学習

- ラベル付きデータとラベルなしデータを用いて学習
- 教師あり学習と教師なし学習の間

半教師あり学習



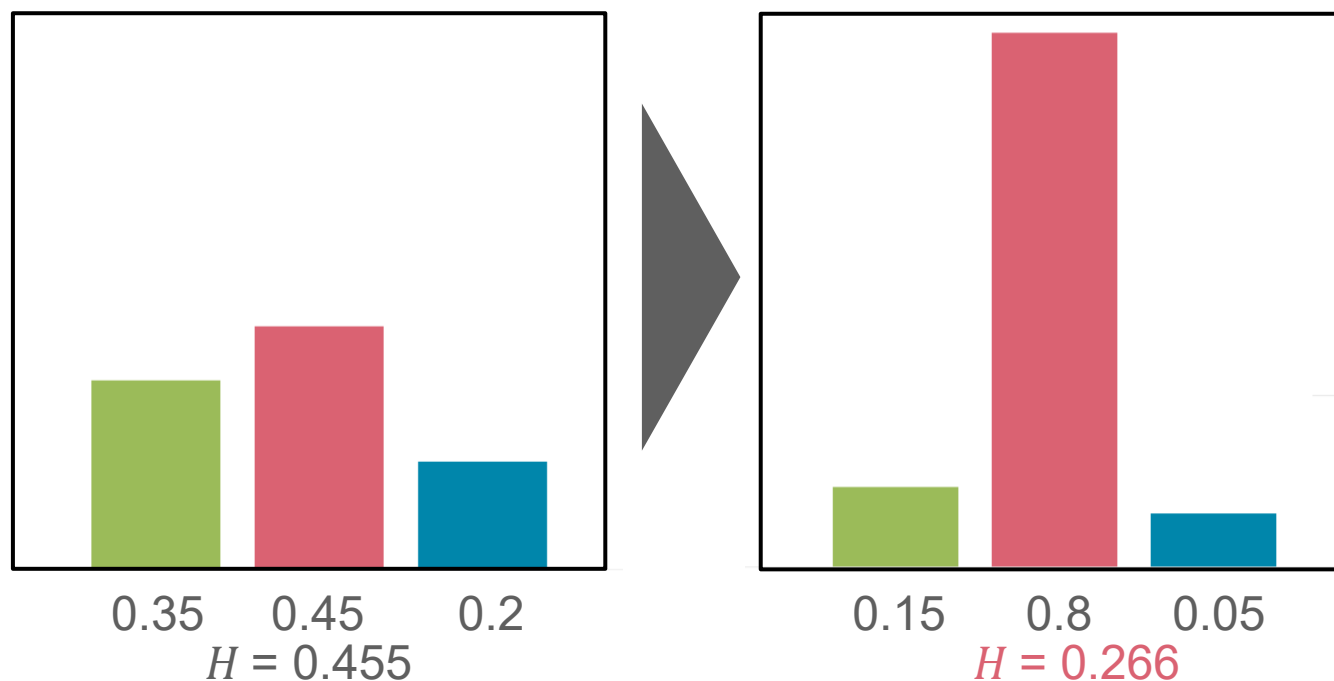
主なアプローチ

- Entropy Minimization
- Consistency Regularization
 - MixMatchベース
- 自己教師あり学習→Finetuning

Entropy Minimization

- モデルの予測 $p(y|x)$ の**エントロピー H を小さくする**ように学習
 - 主なアプローチ | 疑似ラベル
 - エントロピーが小さい = 予測の信頼度が高い

$$H = \operatorname{argmax}_x \left(- \sum_y p(y|x) \log p(y|x) \right)$$



Consistency Regularization

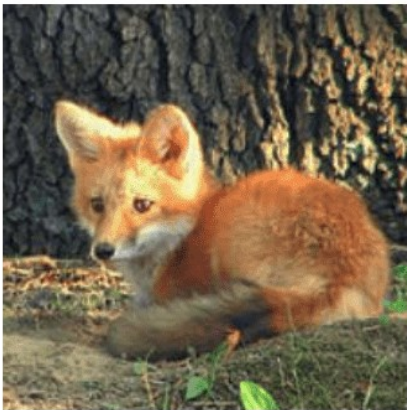
- データにノイズが混ざっても出力が変わらないように学習
 - ノイズはデータ拡張で再現

弱いデータ拡張

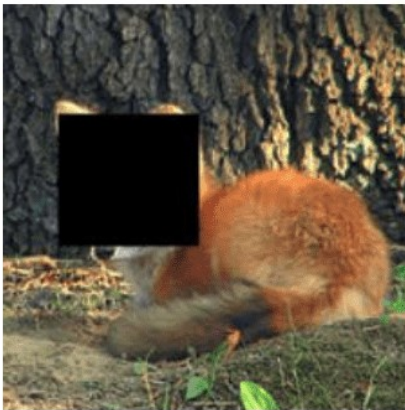
- 回転やクロップなど、拡張後もデータの認識が容易なもの

強いデータ拡張

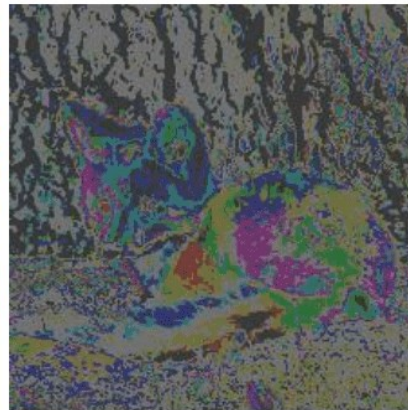
- CutoutやRandAugmentなど、拡張後のデータの認識が困難なもの



(a1) Red fox



(a2) Cutout



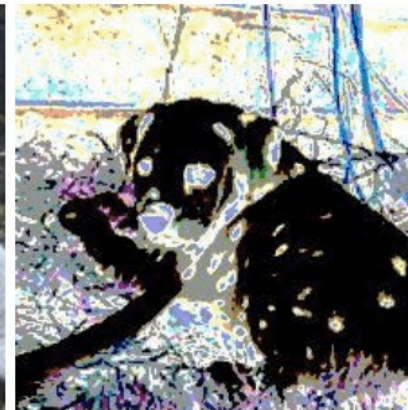
(a3) RandAugment



(b1) Dog



(b2) Cutout

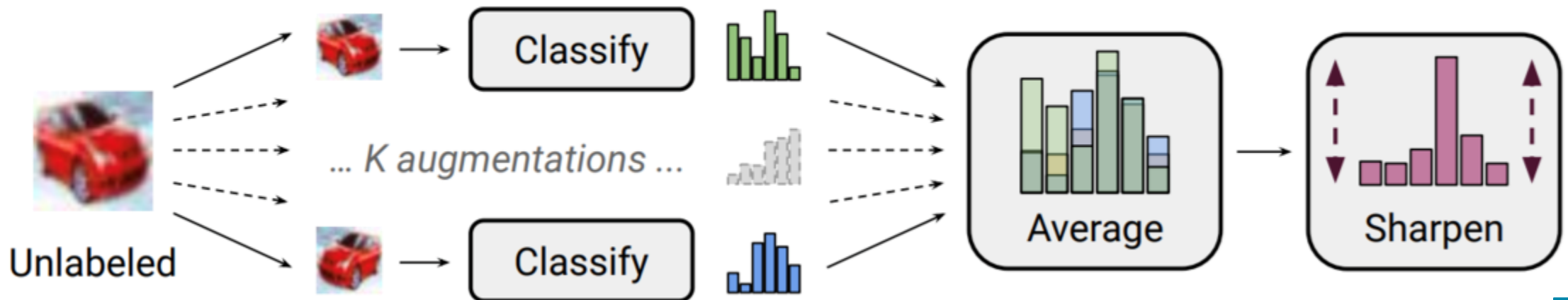


(b3) RandAugment

MixMatch: A Holistic Approach to Semi- Supervised Learning, NIPS '19

- Entropy Minimization, Consistency Regularizationを**同時**に行う
手順

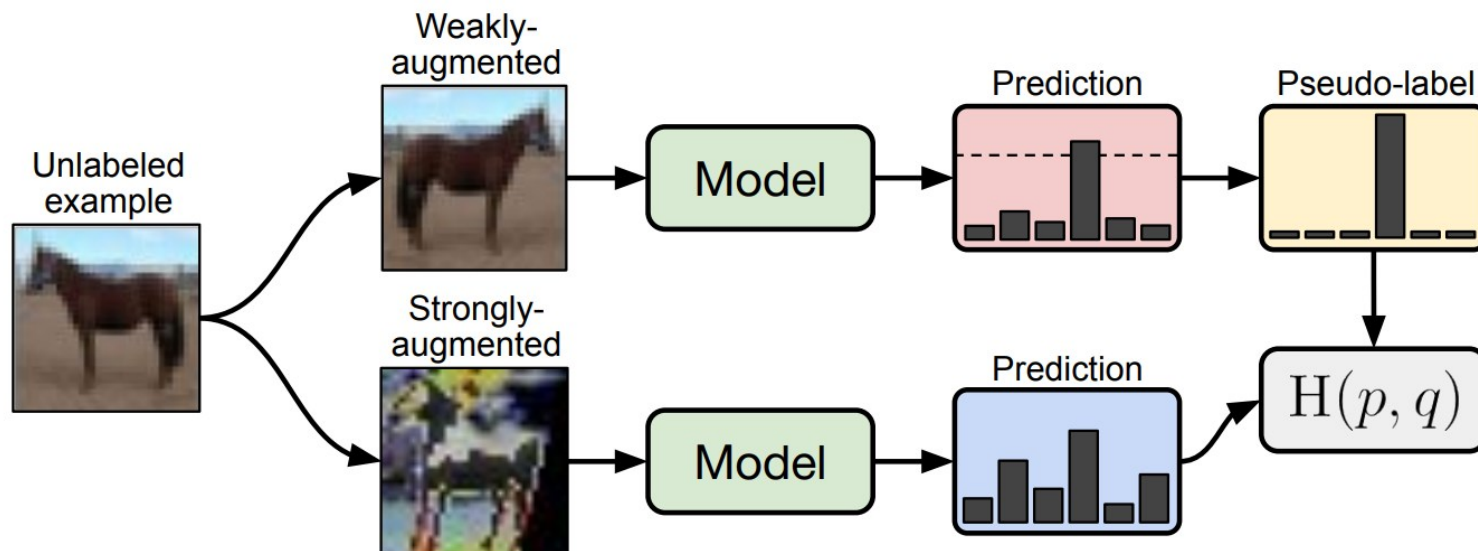
1. ラベル付きデータから拡張データ X を生成
 2. ラベルなしデータから K 種類の拡張データ U を生成
 3. ラベルなしデータの出力平均から疑似ラベルを作成 ← Entropy Minimization
 4. X , U とそのラベルを混合 (MixUp) して学習
- ← Consistency Regularization



FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, NIPS '20

- 強い/弱い データ拡張を用いてMixMatchを改良, 簡素化
手順

- 各データに強い/弱い データ拡張を行い2つの拡張データを生成
- 強いデータ拡張を与えた拡張データに対する予測から疑似ラベルを生成 ← Entropy Minimization
- 弱いデータ拡張を与えた拡張データと2の疑似ラベルを最適化 ← Consistency Regularization



半教師あり学習の性能

- MixMatch, FixMatchベース手法が高性能
 - ― 設定次第では教師あり学習を上回る

Dataset	CIFAR-10				CIFAR-100			SVHN			STL-10	
# Label	10	40	250	4000	400	2500	10000	40	250	1000	40	1000
II Model (Rasmus et al., 2015)	79.18±1.11	74.34±1.76	46.24±1.29	13.13±0.59	86.96±0.80	58.80±0.66	36.65±0.00	67.48±0.95	13.30±1.12	7.16±0.11	74.31±0.85	32.78±0.40
Pseudo Label (Lee et al., 2013)	80.21±0.55	74.61±0.26	46.49±2.20	15.08±0.19	87.45±0.85	57.74±0.28	36.55±0.24	64.61±5.6	15.59±0.95	9.40±0.32	74.68±0.99	32.64±0.71
VAT (Miyato et al., 2018)	79.81±1.17	74.66±2.12	41.03±1.79	10.51±0.12	85.20±1.40	46.84±0.79	32.14±0.19	74.75±3.38	4.33±0.12	4.11±0.20	74.74±0.38	37.95±1.12
MeanTeacher (Tarvainen & Valpola, 2017)	76.37±0.44	70.09±1.60	37.46±3.30	8.10±0.21	81.11±1.44	45.17±1.06	31.75±0.23	36.09±3.98	3.45±0.03	3.27±0.05	71.72±1.45	33.90±1.37
MixMatch (Berthelot et al., 2019b)	65.76±7.06	36.19±6.48	13.63±0.59	6.66±0.26	67.59±0.66	39.76±0.48	27.78±0.29	30.60±8.39	4.56±0.32	3.69±0.37	54.93±0.96	21.70±0.68
ReMixMatch (Berthelot et al., 2019a)	20.77±7.48	9.88±1.03	6.30±0.05	4.84±0.01	42.75±1.05	26.03 ±0.35	20.02 ±0.27	24.04±9.13	6.36±0.22	5.16±0.31	32.12±6.24	6.74±0.14
UDA (Xie et al., 2020a)	34.53±10.69	10.62±3.75	5.16±0.06	4.29±0.07	46.39±1.59	27.73±0.21	22.49±0.23	5.12±4.27	1.92 ±0.05	1.89 ±0.01	37.42±8.44	6.64±0.17
FixMatch (Sohn et al., 2020)	24.79±7.65	7.47±0.28	4.86 ±0.05	4.21±0.08	46.42±0.82	28.03±0.16	22.20±0.12	3.81±1.18	2.02±0.02	<u>1.96</u> ±0.03	35.97±4.14	6.25±0.33
Dash (Xu et al., 2021)	27.28±14.09	8.93±3.11	5.16±0.23	4.36±0.11	44.82±0.96	27.15±0.22	21.88±0.07	<u>2.19</u> ±0.18	2.04±0.02	1.97±0.01	34.52±4.30	6.39±0.56
MPL (Pham et al., 2021)	23.55±6.01	6.62±0.91	5.76±0.24	4.55±0.04	46.26±1.84	27.71±0.19	21.74±0.09	9.33±8.02	2.29±0.04	2.28±0.02	35.76±4.83	6.66±0.00
FlexMatch (Zhang et al., 2021)	<u>13.85</u> ±12.04	<u>4.97</u> ±0.06	4.98±0.09	<u>4.19</u> ±0.01	<u>39.94</u> ±1.62	26.49±0.20	21.90±0.15	8.19±3.20	6.59±2.29	6.72±0.30	<u>29.15</u> ±4.16	<u>5.77</u> ±0.18
FreeMatch	8.07 ±4.24	4.90 ±0.04	<u>4.88</u> ±0.18	4.10 ±0.02	37.98 ±0.42	<u>26.47</u> ±0.20	<u>21.68</u> ±0.03	1.97 ±0.02	<u>1.97</u> ±0.01	<u>1.96</u> ±0.03	15.56 ±0.55	5.63 ±0.15
Fully-Supervised	4.62±0.05				19.30±0.09			2.13±0.01			-	

※強いデータ拡張は自然画像以外では機能しない（FixMatchより）

まとめ

- 半教師あり学習 | 一部のデータにアノテーションし, 学習する枠組み
 - アノテーション作業を削減
 - 設定によっては教師あり学習の性能を上回る
- 近年のトレンド | MixMatch, FixMatchベース
 - 問題設定が類似しており, ALとの組み合わせもチラホラ
 - ActiveMatch
- 利用可能なタスクは限られる
 - 強いデータ拡張が機能するもの限定