NTT

22-27 MAY 2022

ICASSP 2022
SINGAPORE

Customer Satisfaction Estimation using Unsupervised Representation Learning with Multi-Format Prediction Loss

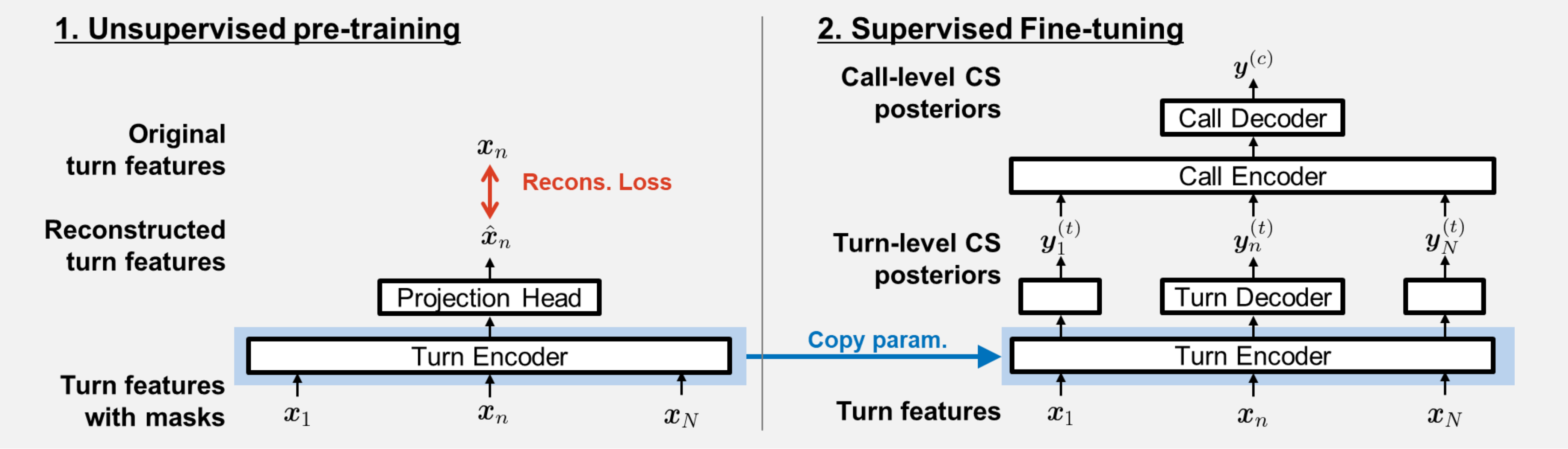
Atsushi Ando, Yumiko Murata, Ryo Masumura, Satoshi Suzuki,
Naoki Makishima, Takafumi Moriya, Takanori Ashihara, Hiroshi Sato

NTT Corporation, Japan

Summary

- Task: Call- / turn-level customer satisfaction estimations (3-class; *pos, neu, neg*)
- Contributions

1. **Introduce unsupervised representation learning** by a large amount of unlabeled data

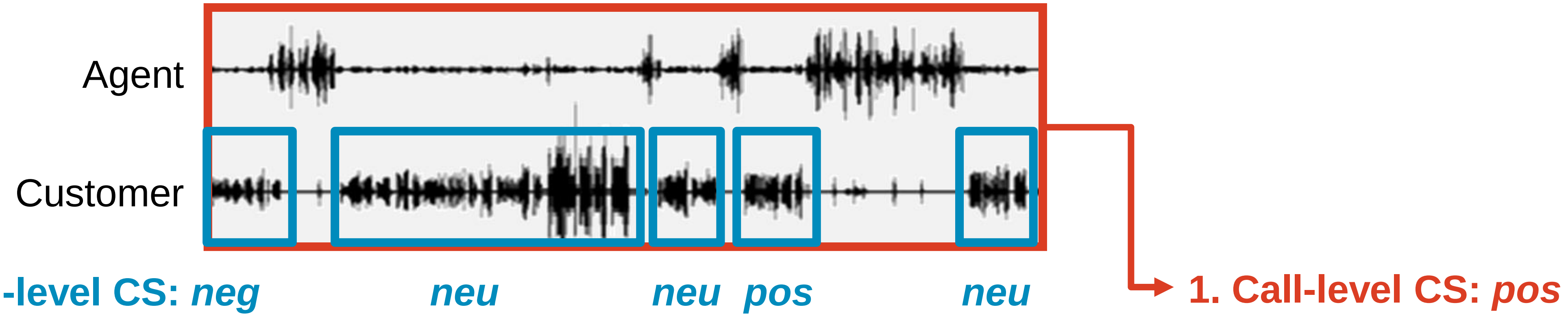


2. Propose a new loss function called **Multi-Format Prediction (MFP) loss** to improve the reconstruction of both continuous and discrete features in unsupervised pre-training

- Results
  - Both Call- / Turn-level estimations improved on real English contact center calls

Task Descriptions

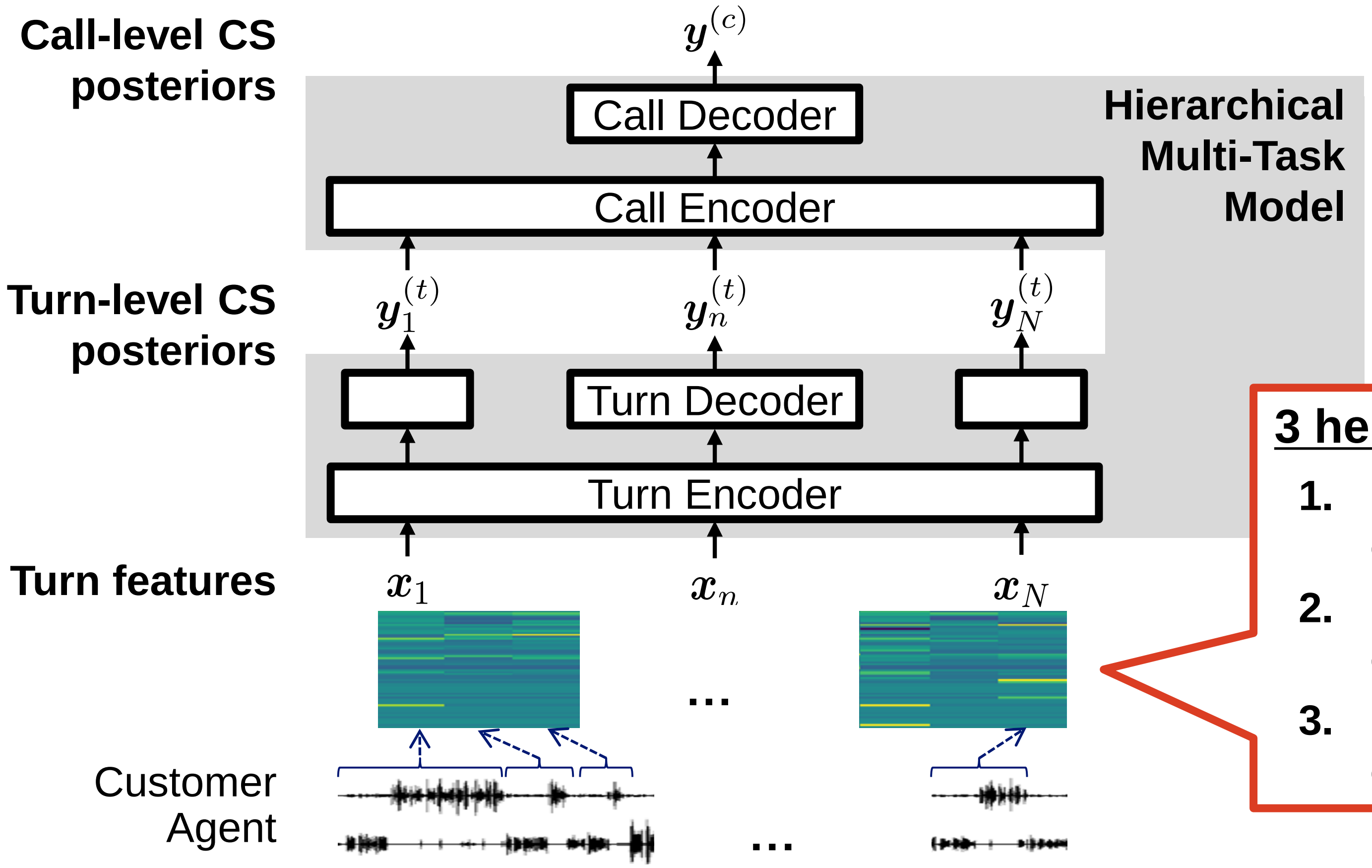
- **Estimate 2-levels of customer satisfaction (CS) degrees**
  1. **Call-level CS : CS with an overall call** **(*pos, neu, neg*)**
  2. **Turn-level CS : CS of each customer turn during a call** **(*pos, neu, neg*)**



  - Customer and agent turns are automatically detected by Voice-Activity-Detection (VAD)
  - Ground truths of call- / turn-level CS are determined by human annotators

Conventional Method

- **Hierarchical Multi-Task (HMT) model [Ando+, 20]**
  - Utilize the relationship between turn-level and call-level CS degrees



3 heuristic features

1. **Prosodic features**
   ex) F0/Power statistics of customer turn
2. **Interactive features**
   ex) length, rate of customer/agent turn
3. **Lexical features**
   ex) Bag-of-Words(BoW) in customer turn

- **Problem: performance is insufficient in limited labeled training data**

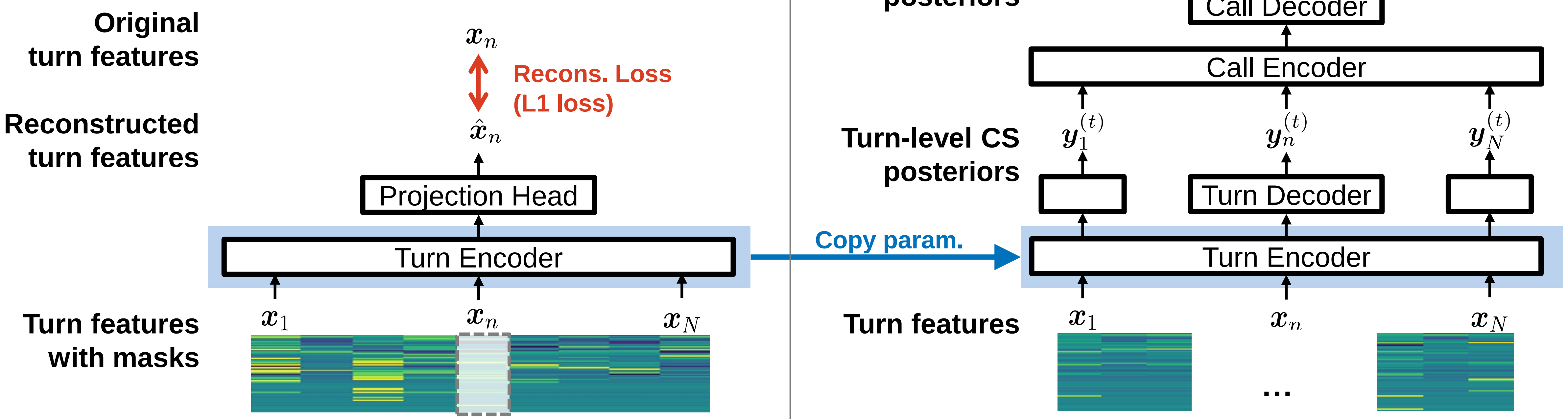
Approach

- **Utilize large amounts of unlabeled data**
- **2-step training:**
  1. **Unsupervised pre-training with unlabeled data**
  2. **Supervised fine-tuning with labeled data**

1. Unsupervised pre-training

… Masked feature model (Mockingjay) [Liu+,20]

2. Supervised Fine-tuning

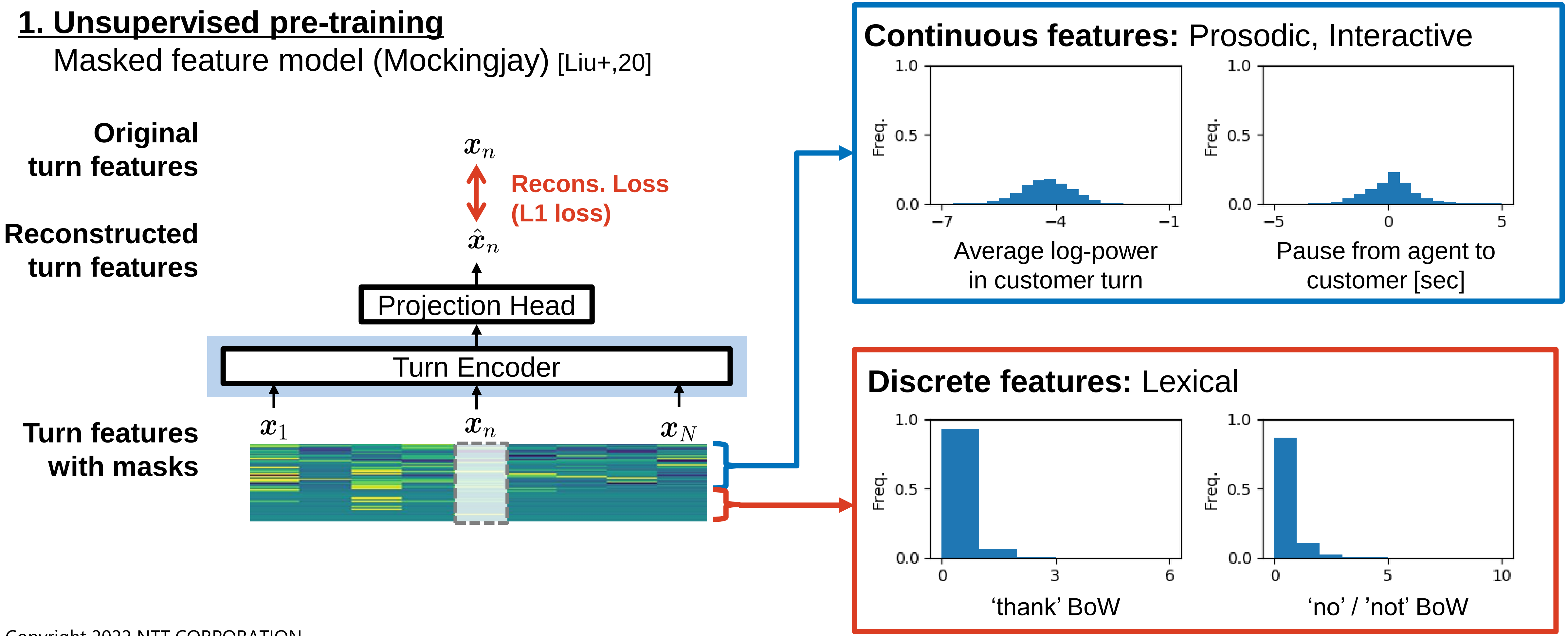


Problem in unsupervised pre-training

- **Discrete features (ex. BoW) are difficult to reconstruct by L1 loss**
  - Lead to outputting 0 values on all turns

1. Unsupervised pre-training

Masked feature model (Mockingjay) [Liu+,20]



Proposed Method

- **Introduce a new loss function to improve the reconstruction of both continuous and discrete features in unsupervised pre-training:**
  Multi-Format Prediction (MFP) Loss
  - **Continuous features … L1 Loss**

$$\mathcal{L}_{\mathrm{L1}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{i\in I_c}|x_{n,i} - \hat{x}_{n,i}|$$

Continuous feat. indices

  - **Discrete features … Weighted BCE (0 or >0)**

Discrete feat. indices

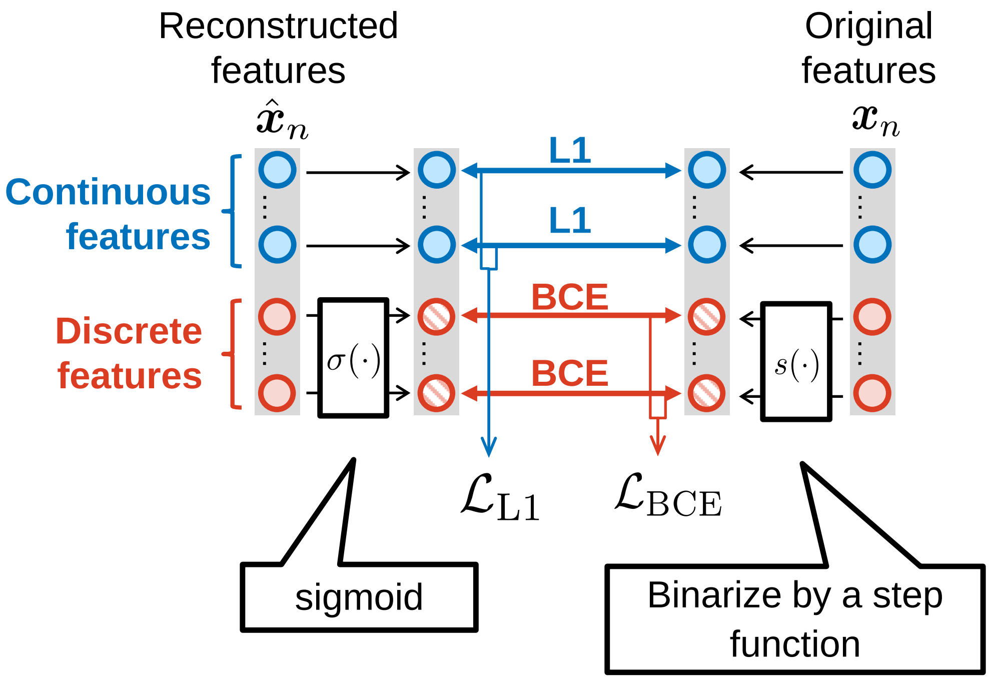
Value >0 weight
(Inverse frequencies for unlabeled data)

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i\in I_d}\{w_{i,1}s(x_{n,i})\log(\sigma(\hat{x}_{n,i}))$$
$$+w_{i,0}(1-s(x_{n,i}))\log(1-\sigma(\hat{x}_{n,i}))\}$$

Value =0 weight
(Inverse frequencies for unlabeled data)

  - **Total MFP loss**

$$\mathcal{L}_{\mathrm{MFP}} = \beta\mathcal{L}_{\mathrm{L1}} + (1-\beta)\mathcal{L}_{\mathrm{BCE}}$$



Experiments

- Dataset: real English contact center calls
  - Unlabeled: 14782 calls, 388411 turns (approx. 2500 hours)
  - Labeled: 170 calls, 4466 turns (28 hours)
    - 5-fold cross validation
    - Turn- / Call-level labels were determined by well-trained 3 annotators

	pos	*neu*	*neg*
Call	47	97	26
Turn	200	4096	170

- Setups
  - Turn features: 60 dim.
    - Prosodic: 20dim, Interactive: 11dim, Lexical: 29dim (BoW 25dim)
  - Methods
    - Baseline: w/o unsupervised pre-training
    - Proposed:
      - w/ unsupervised pre-training by L1 loss (same as Mockingjay [Liu+,20])
      - w/ unsupervised pre-training by MFP loss (Proposed)
  - Metrics: Accuracy (Acc.) / macro-averaged F-measures of all classes (macroF1)

Results

- **Unsupervised pre-training by MFP loss improved both turn-level and call-level CS estimations**
  - The L1 loss-based method showed no improvements in call-level estimation

	Pre-training loss function	Turn-level estim.		Call-level estim.	
		Acc.	macroF1	Acc.	macroF1
w/o unsupervised pre-training (Baseline)		.857	.525	.571	.522
w/ unsupervised pre-training	L1 loss (Mockingjay[Liu+,20])	**.878**	**.546**	.571	.492
	MFP loss (Proposed)	.875	.543	**.647**	**.600**

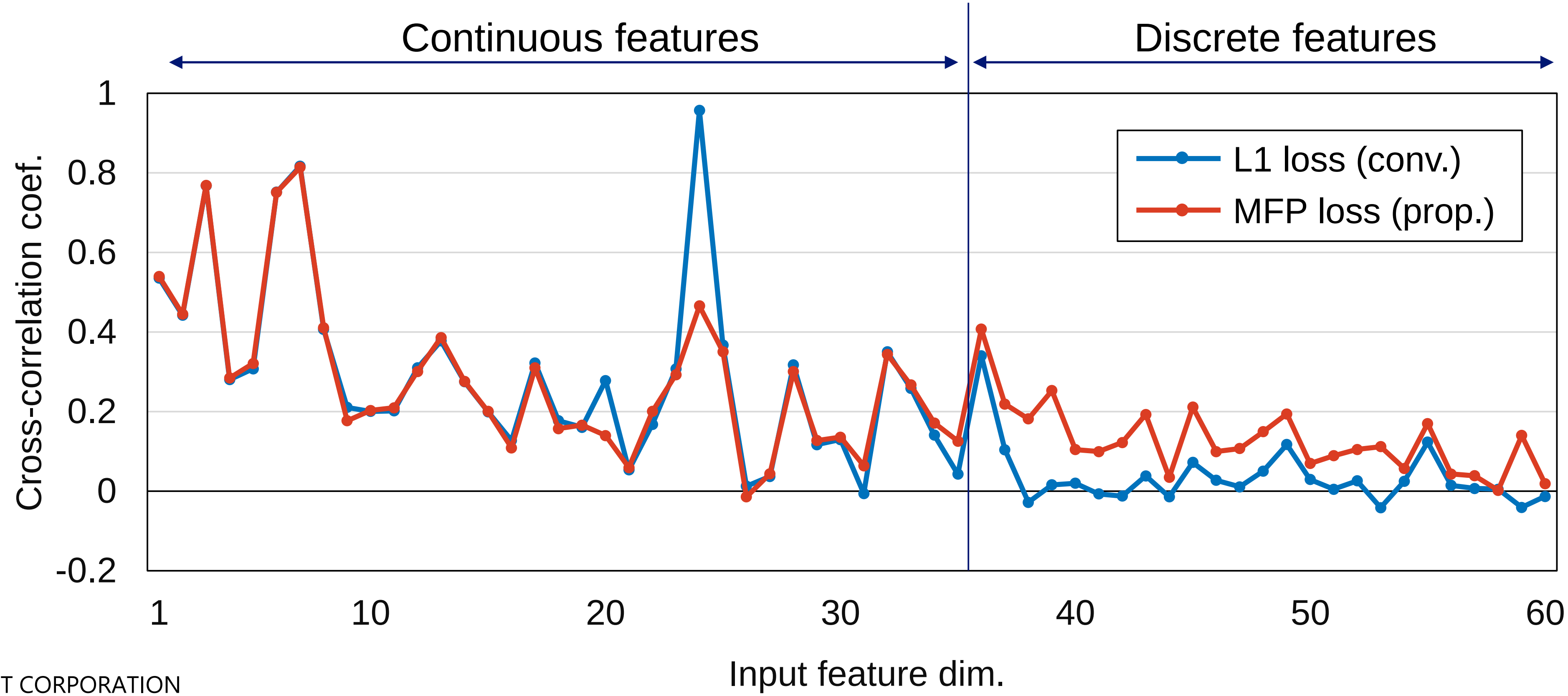
Results

- **Unsupervised pre-training by MFP loss improved both turn-level and call-level CS estimations**
  - The L1 loss-based method showed no improvements in call-level estimation

	Pre-training loss function	Turn-level estim.		Call-level estim.	
		Acc.	macroF1	Acc.	macroF1
w/o unsupervised pre-training (Baseline)		.857	.525	.571	.522
w/ unsupervised pre-training	L1 loss (Mockingjay[Liu+,20])	**.878**	**.546**	.571	.492
	MFP loss (Proposed)	.875	.543	**.647**	**.600**

Discussions: Feature Reconstruction Performances

- **Evaluated the cross-correlation coefficients between the original and the reconstructed features in the unsupervised pre-training**
- **MFP loss-based model showed higher correlation values than L1-model for the discrete features**



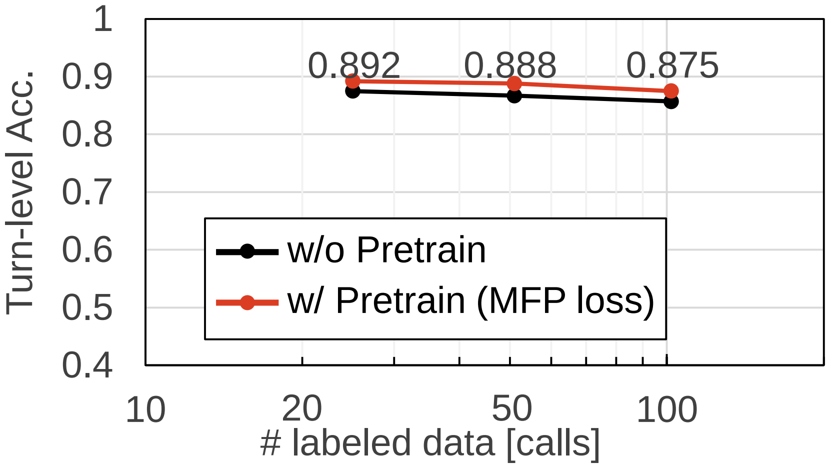
Discussions: Training Curve

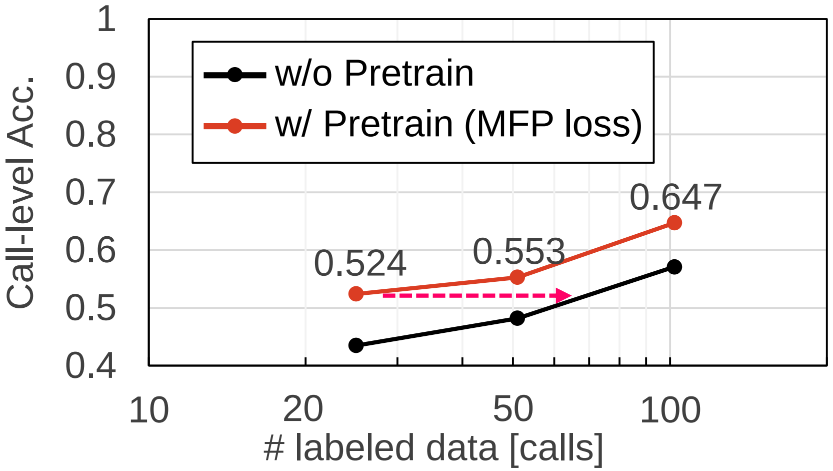
- **Unsupervised pre-training with x100 unlabeled data is equivalent to supervised training with x2 labeled data**

Turn-level estim.

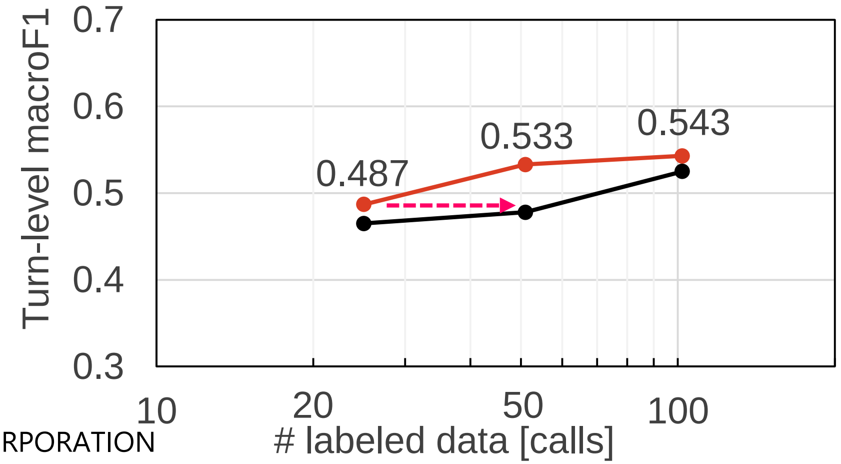
Call-level estim.

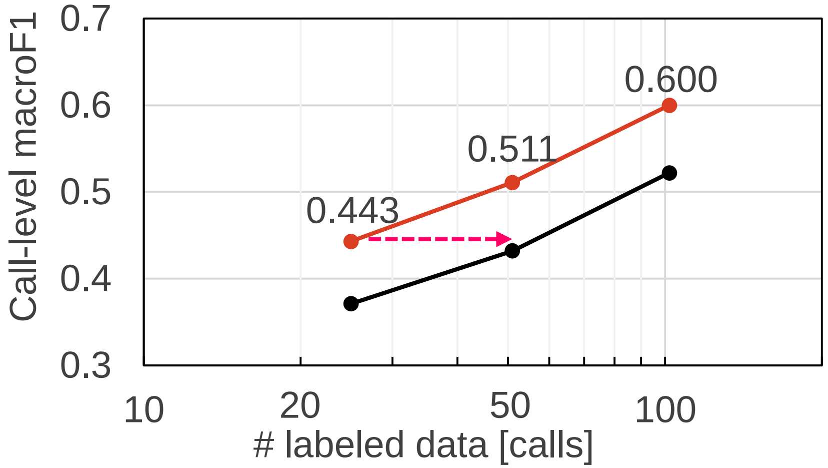
Accuracy





macroF1





Conclusions

- Summary
  - Task:
    Call- / Turn-level customer satisfaction estimation (3-class; *pos, neu, neg*)
  - Approach:
    unsupervised pre-training with large amounts of unlabeled data
  - Contribution:
    Introduce **a new loss function called Multi-Format Prediction (MFP) loss** to improve the reconstruction of both continuous and discrete features in unsupervised pre-training
  - Results:
    Both Call- / Turn-level estimation performances improved on real English contact center calls, and MFP loss showed better reconstructions for discrete features

- Future work
  - Evaluations of other contact center calls/languages