# A DISCRIMINATIVE MAXIMUM ENTROPY MODEL FOR RELIABILITY-AWARE CLASSIFICATION

Ming Hsieh Department of Electrical Engineering

Naveen Kumar, Shrikanth Narayanan
Signal Analysis and Interpretation Laboratory (SAIL), USC
Department of Electrical and Computer Engineering

USC Viterbi School of Engineering

## Motivation

In real-world classification tasks we encounter problems like

- Noisy observations
- Presence of outliers
- Feature and sample *reliabilities* vary

**How to define and model this reliability?**

## Reliability Assumption

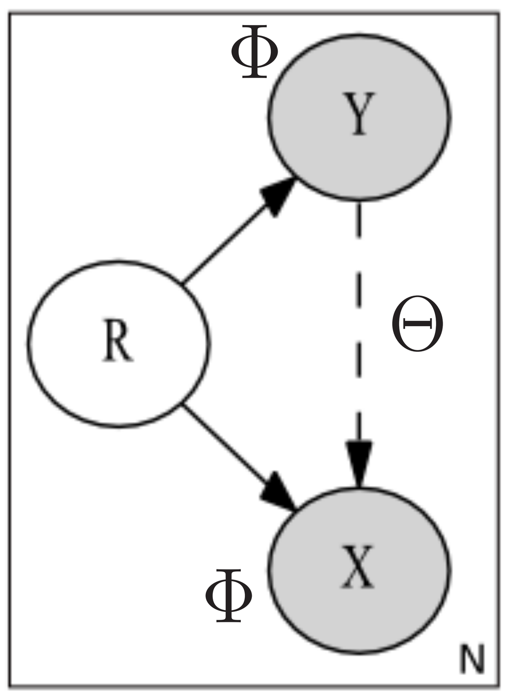
**Classification Task**

- Predict labels $Y$ using features $X$
- $X \in \mathbb{R}^D; Y \in 1 \dots K$

**Reliability-aware framework**

- $\mathbf{\Theta}$: Feature-dependent *reliable* model
- $\mathbf{\Phi}$: Feature-independent *unreliable* model
- $R = \{0, 1\}$ : latent reliability variable for the sample

$$\boxed{Pr(X, Y|R) = Pr(X, Y; \mathbf{\Theta})^R (Pr(Y; \mathbf{\Phi}) Pr(X; \mathbf{\Phi}))^{1-R}}$$




**Dotted line** : $X$-$Y$ dependence regulated by $R$
**Shaded:** Observed

**A Generative Model**

$$\boxed{Pr(X|Y, R) = Pr(X|Y; \mathbf{\Theta})^R Pr(X; \mathbf{\Phi})^{1-R}}$$

*How are the features generated?*

## Discriminative reliability-aware model

- Only learn models for $Pr(Y|X)$
- *How was the label generated?*
- **Reliable:** Annotator consults the data before labeling
- **Unreliable:** Annotator tosses a die to assign label

$$Pr(Y, R|X) = \underbrace{Pr(Y|X; \mathbf{\Theta})^R}_{\text{reliable}} \underbrace{Pr(Y; \mathbf{\Phi})^{1-R}}_{\text{unreliable}} Pr(R|X)$$

$R = 1$

Data-dependent

$R = 0$

Data-independent

## Algorithm

**Reliable model** (MaxEnt)

$$Pr(Y_{ik} = 1|X_i; \mathbf{\Theta}) = \frac{e^{W_k{}^T X_i}}{\sum_{j=1}^K e^{W_j{}^T X_i}}$$

**Unreliable model**( Categorical)

$$Pr(Y_{ik}; \mathbf{\Phi}) = \eta_k$$

**Reliability model** (Logistic)

$$Pr(R_i = 1|X_i) = \sigma(r^T X_i) = 1/(1 + e^{-r^T X_i})$$

**Maximize the total data-likelihood using EM**

$$\begin{aligned} Pr[\mathcal{D}|W, \eta, r] &= \prod_{i=1}^N Pr(Y_i, R_i|X_i; W, \eta, r) \\ &= \prod_{i=1}^N Pr(Y_i|X_i; W)^{R_i} Pr(Y_i; \eta)^{1-R_i} Pr(R_i|X; r) \end{aligned}$$

## Intelligibility Classification Experiment

**Pathological Speech**

- Atypicality resulting from disease or surgery of the vocal tract
- Reduced speech intelligibility
- Intelligibility depends on many diverse factors

**Dataset**

- NKI CCRT Speech Corpus
- 2385 sentence level utterances labelled as **I** or **NI**
- 13 features based on pronunciation, voice quality, prosody

## Results

- Chance accuracy : 50.3%
- Baseline model: ordinary logistic regression
- Results on a 5-fold cross validation

**Baseline and proposed methods on different feature sets**

| Feature | Baseline | Proposed |
|---|---|---|
| voice quality | **58.0** | **60.3** |
| prosody | 55.8 | 55.8 |
| pronunciation | 67.1 | 66.8 |
| Feature fusion | 67.6 | 67.9 |
| Score fusion | 67.9 | 67.9 |

## Conclusion

- Helps when features are unreliable
- Reliabilities different for feature set
- Similar to a mixture-of-experts model
- Improvement not signficant over feature fusion - different reliabilities