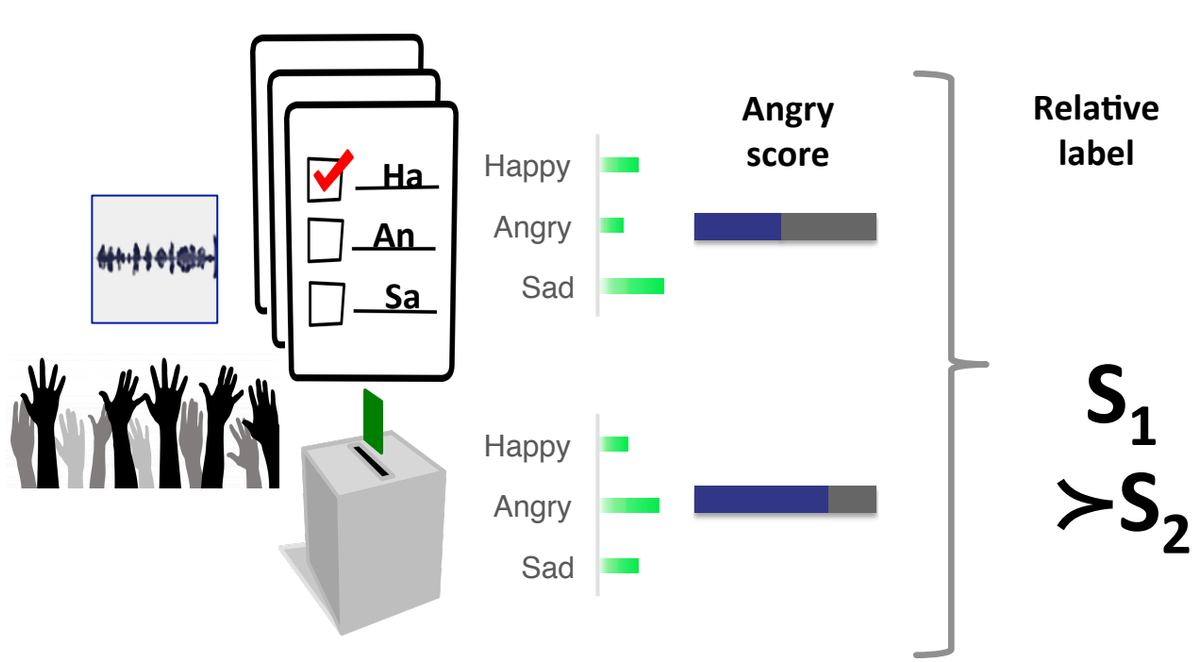
Retrieving Categorical Emotions using a Probabilistic Framework to Define Preference Learning Samples

Reza Lotfian and Carlos Busso

Multimodal Signal Processing (MSP) lab
The University of Texas at Dallas
Erik Jonsson School of Engineering and Computer Science



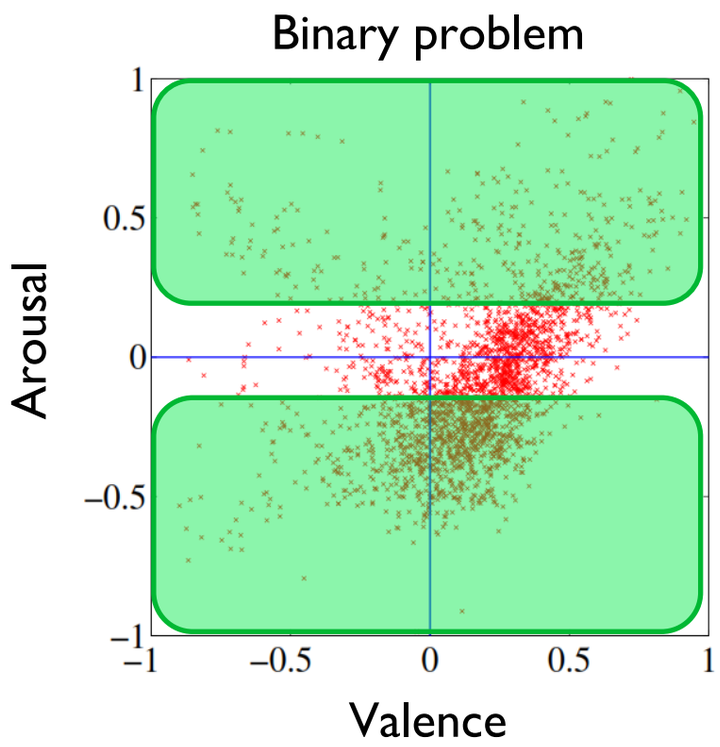
September. 9th, 2016

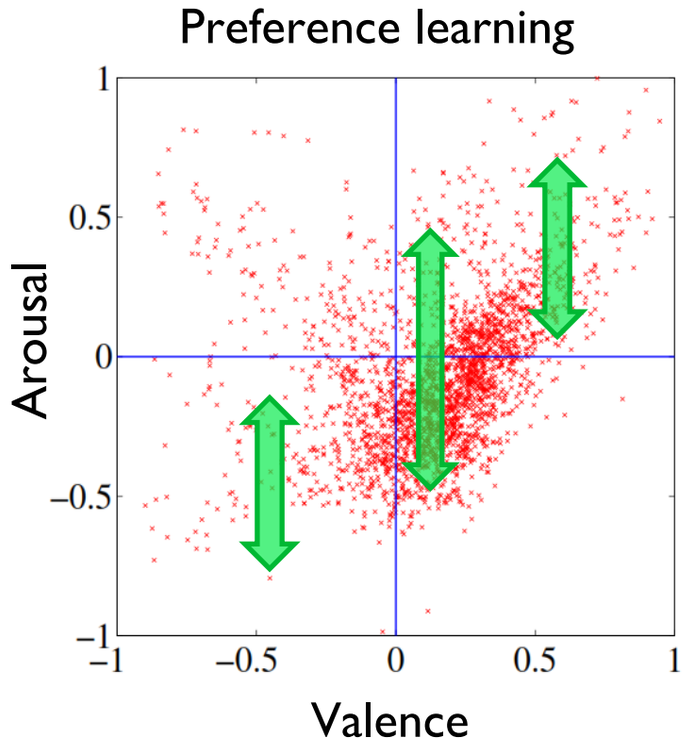
Motivation

- Retrieving speech conveying target emotion
  - Surveillance, call center
- Binary and multi-class
  - Focus of most previous studies
  - Few studies on preference learning on continuous attributions (arousal, valence,…) [Lotfian et.al, 2016]
- This study: Preference learning on categorical emotions (Happy, Sad,….)
  - E.g., “is sample A happier than sample B”

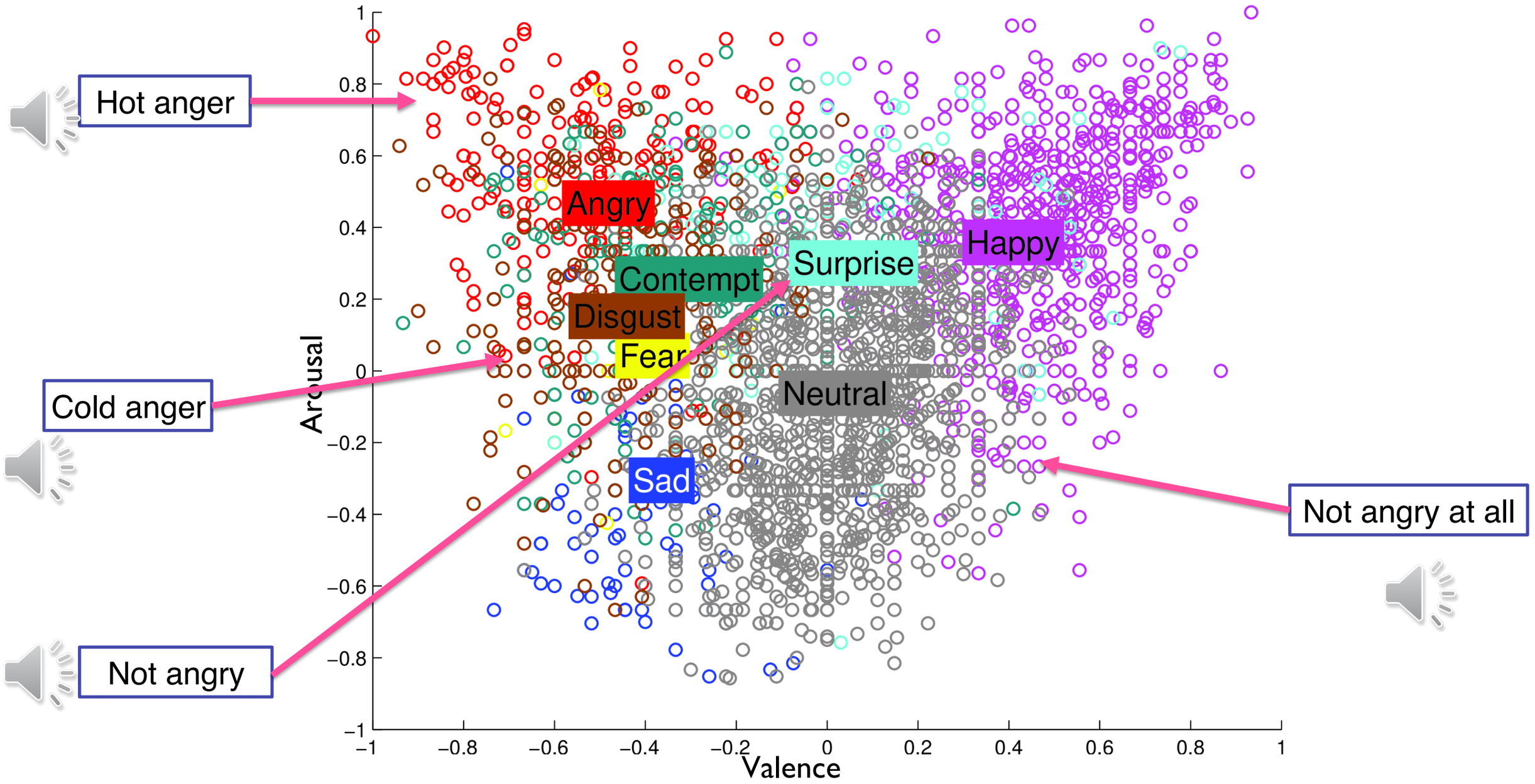
Preference Learning-Ranking

- Multiclass classification
  - Assign a class label to each sample (High versus low arousal)
- Preference learning
  - For each class rank samples based on relevance to the class (Intensity of emotion)
- More reliable training examples without sacrificing training size

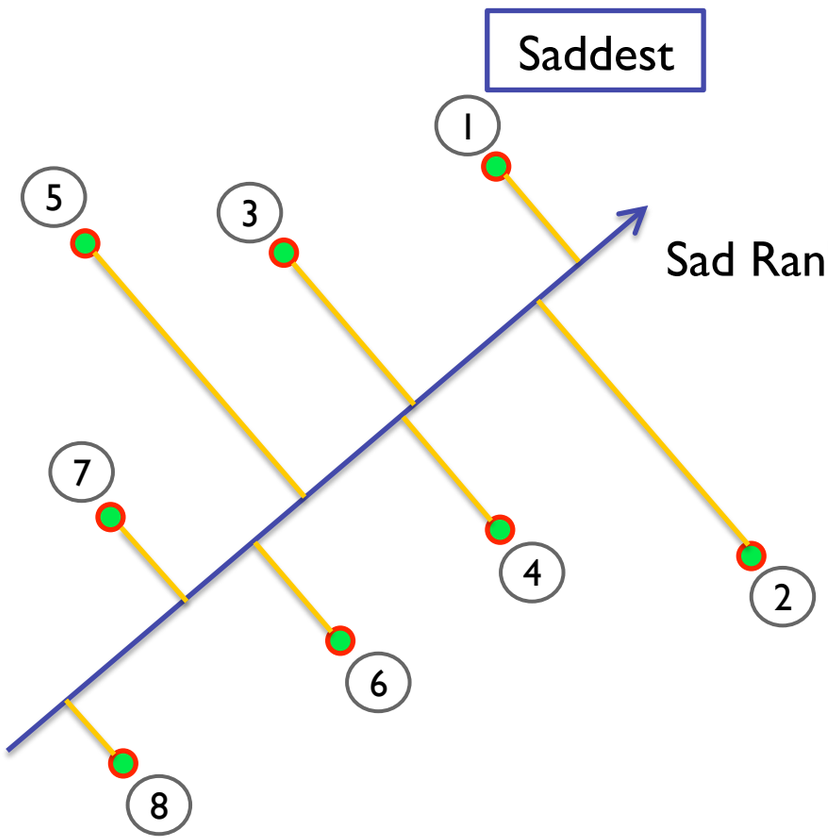


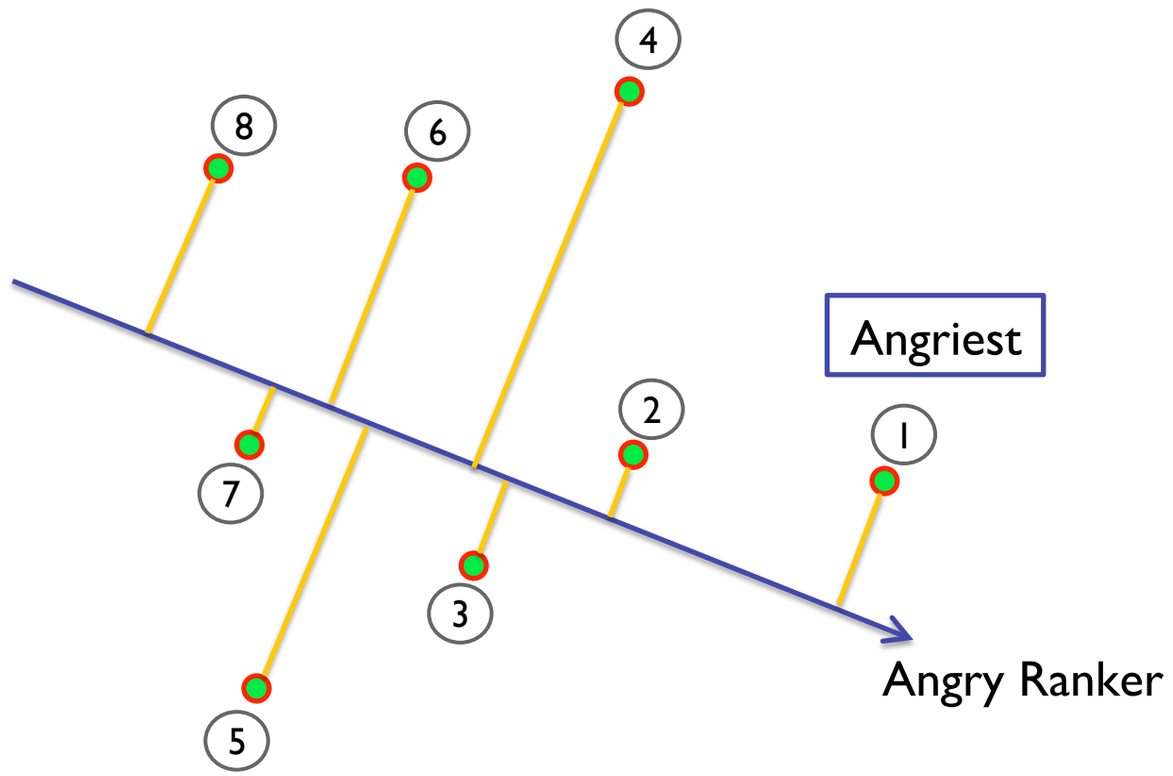


Categorical Emotion Ranker



Categorical Emotion Ranker





Training Preference

- Preference learning needs a set of ordered pairs of samples for training
  - Arousal, Valence, Dominance: interval variables

$$e^{s_1}_{arousal} - e^{s_2}_{arousal} > margin \longrightarrow s_1 \succ_{arousal} s_2$$

- Happiness, Sadness … binary (happy – not happy)

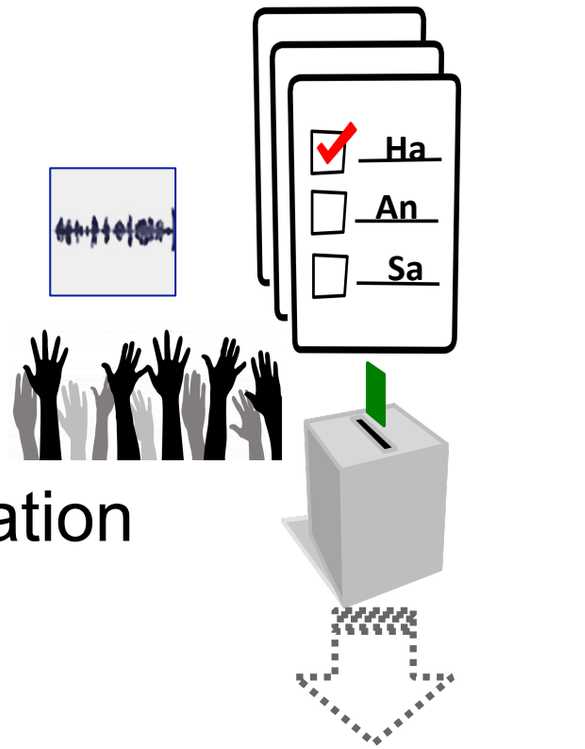
$happy \succ_{happiness} sad$ ✓ [Cao et.al, 2014]

$angry \succ_{happiness} sad$?

$happy_1 \succ_{happiness} happy_2$?

Definition of the Problem

- What we have for training:
  - Audio sample & set of subjective evaluations
    $s_1(happy, happy, happy)$
    $s_2(happy, happy, sad)$
- Hypothesis:
  - Individual annotations provides more information than majority vote



- s_1 is more likely to be happy than s_2

$$s_1 \succ_{happiness} s_2$$

- s_2 is more likely to be sad than s_1

$$s_2 \succ_{sadness} s_1$$

$$P(Happy|happy, happy, happy) > P(Happy|happy, happy, sad)$$

$$P(Sad|happy, happy, sad) > P(Sad|happy, happy, happy)$$

Quantifying Categorical Emotion Preference

- Set of R annotations $x_1, x_2, ..., x_R$
- Emotion class $w_j, j \in \{1, ..., E\}$
- Posteriori probability (choose the class that maximize)

$$P(w_k|x_1, x_2, x_3) = \frac{P(w_i)\prod_{i=1}^{R} p(x_i|w_k)}{\sum_{j=1}^{m} P(w_j)\prod_{i=1}^{R} p(x_i|w_k)}$$

- Sum rule [Kittler,1998]

$$\nu_j = (1-R)P(w_j) + \sum_{i=1}^{R} P(w_j|x_i)$$

Relevance s
core

Databases

- IEMOCAP (Feature selection and Training)
  - 5 sessions, 10 trained actors
  - Script and spontaneous improvisations
  - 12 hours of recording, manually segmented, transcribed and annotated by 3 independent evaluators

- MSP-IMPROV (Testing)
  - Collected under dyadic improvisation
  - Natural scenarios
  - 4 emotions: anger, sadness, happiness, neutral
  - 9 hours of recording, at least annotated by 5 independent evaluators through crowd-sourcing

Database	Angry	Happy	Sad	Neutral	Other	No Agreement	Total
IEMOCAP	289	284	608	1099	1704	800	4784
MSP-IMPROV	792	2644	885	3477	85	555	8438

Experimental Evaluation

- Learning preference from sum rule (relevance score)

$$\nu_j = (1 - R)P(w_j) + \sum_{i=1}^{R} P(w_j|x_i)$$

- Estimate $P(w_j)$ and $P(w_j|x_i)$ using training database

- Annotation labels: $x\downarrow i \in$ {Happy, Excited, Surprised, Fear, Angry, Frustrated, Disgusted, Sad, Neutral and Other}

- Target emotions: $\omega\downarrow j \in$ {Happiness, Anger, Sadness}

- Learning $P(w_j|x_i) \rightarrow$ assume majority vote consensus label is true label (w_j)

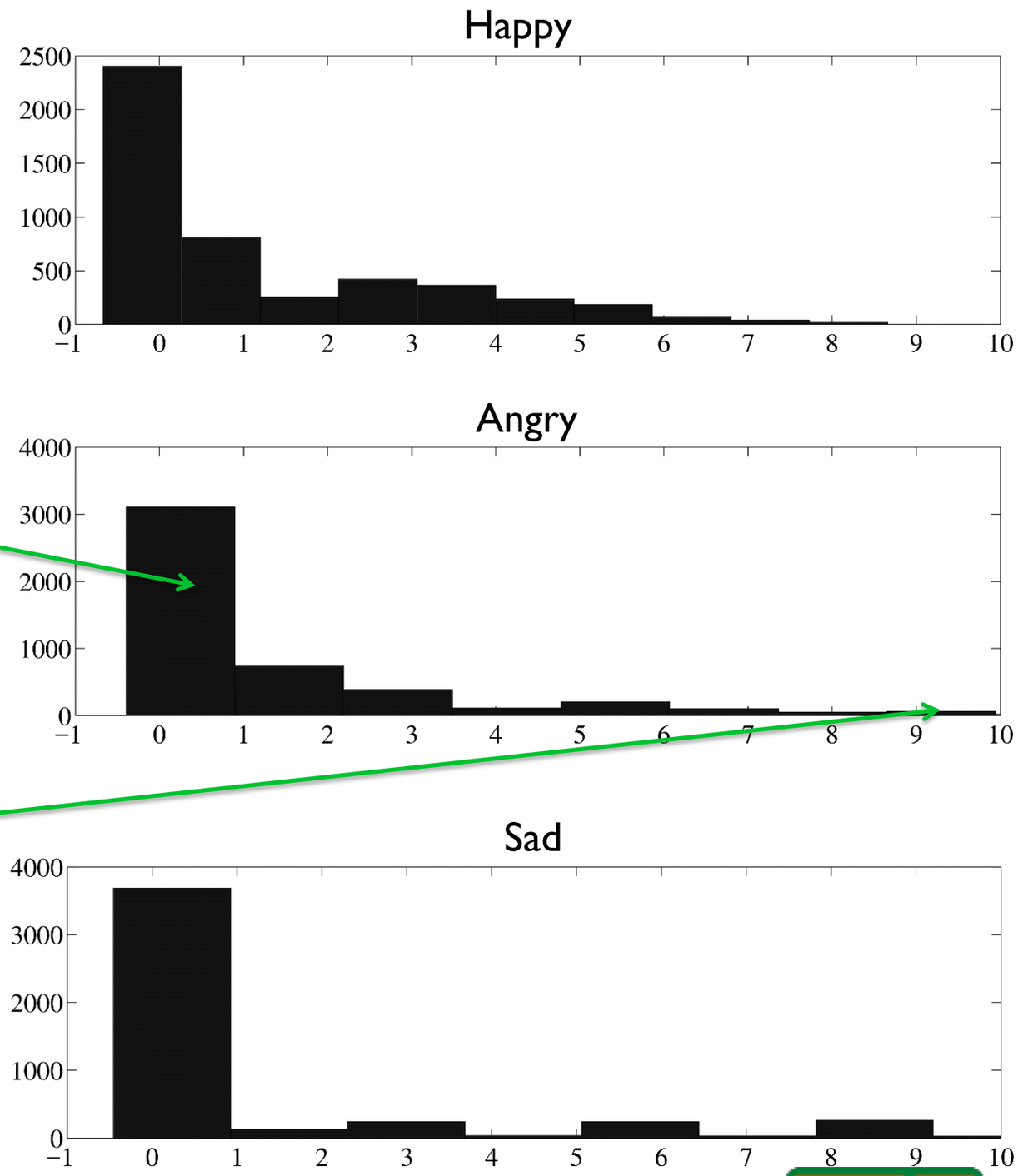
Experimental Evaluation

- Histogram of relevance score estimated on the training set

$$\nu_j = (1 - R)P(w_j) + \sum_{i=1}^{R} P(w_j|x_i)$$

Unlikely to be angry

Likely to be angry



IEMOCAP corpus

Acoustic features

- Speaker state challenge feature set at INTERSPEECH 2013
  - 6308 high level descriptors
  - OpenSMILE toolkit
- Feature selection (separate for each emotion)
  - Step 1: 6308→500
    - Information gain separating target emotion (e.g., Happy vs. other)
  - Step 2: 500→100
    - Floating forward feature selection
    - Maximizing the precision of retrieving top 10%

Preference Learning Methods

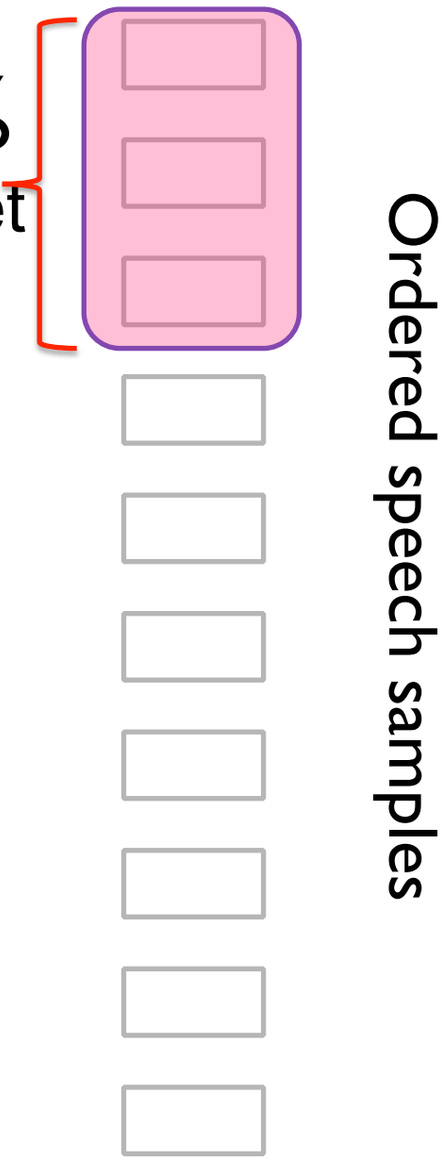
- Rank-SVM: LibSVM toolkit [Joachims, 2006]
- Gaussian Process (GP) preference learning toolkit [Chu & Ghahramani, 2005]
- Training: IEMOCAP database
- Testing: MSP-IMPROV database
- Relative labels:
  - Relevance score (proposed method)
  - Baseline:
    - Label based (Cao et al. [2014])

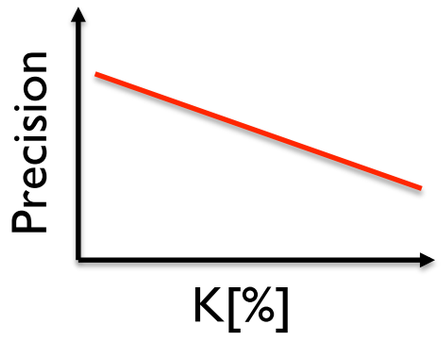
$$happy \succ_{happiness} sad$$

Measure of Retrieval Performance

Precision at K (P@K)

- Speech samples ordered by a rank method
- Select K that we know has target emotion in the test set
  - Example: P@30 → 2644 Happy samples in test set
    retrieve 2644*0.30=793 samples of highest rank
- Success if the sample is from target emotion class (Happy)
  - We can compare this approach to other machine learning algorithm





Comparing Precision @ K

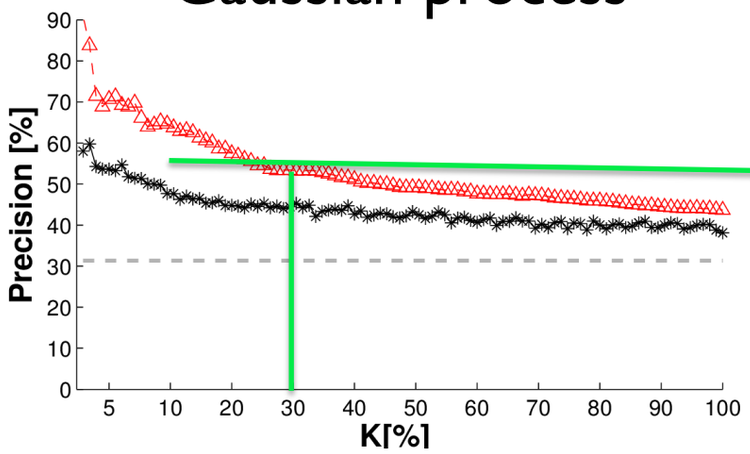
Relevance score

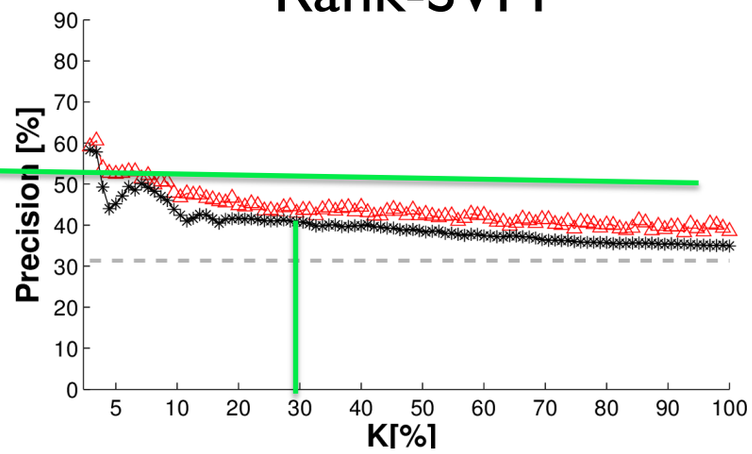
Cao et al

Gaussian process

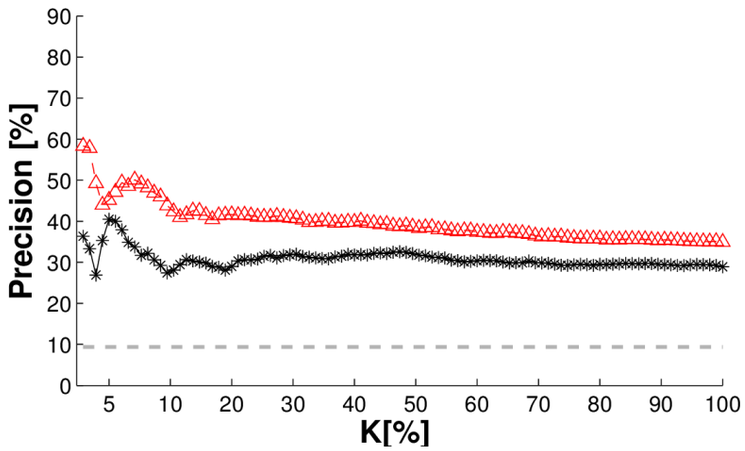
Rank-SVM

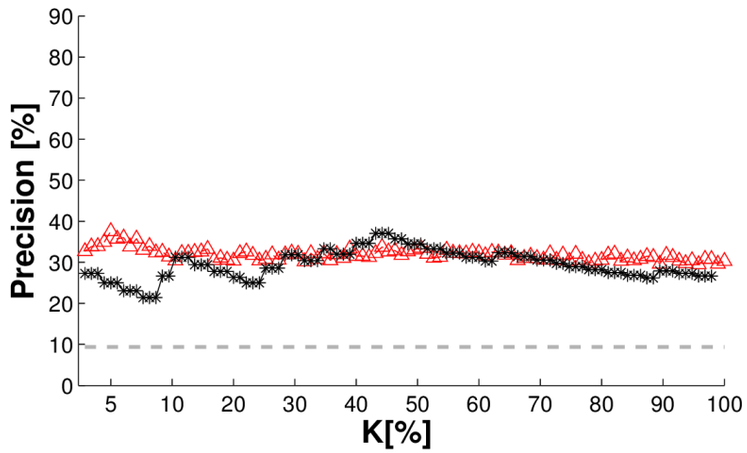
Happy



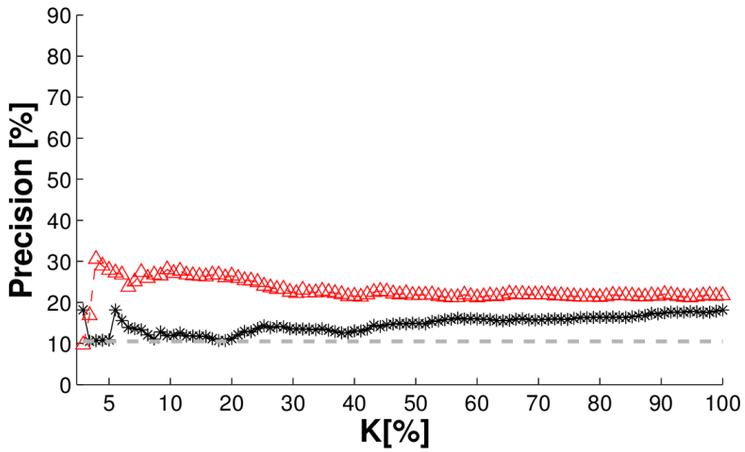


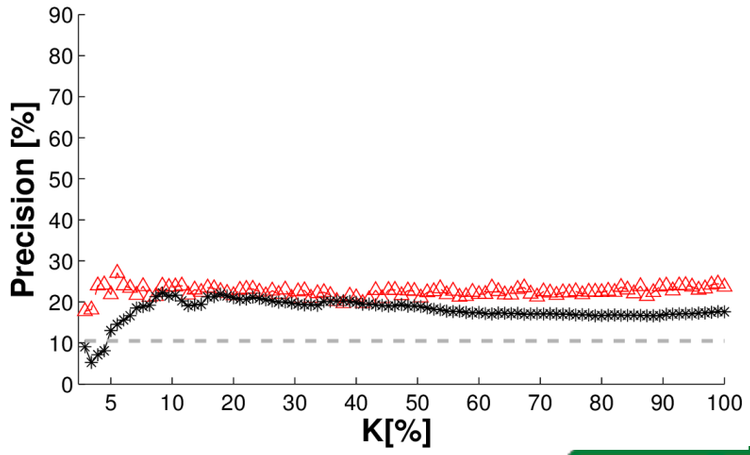
Angry



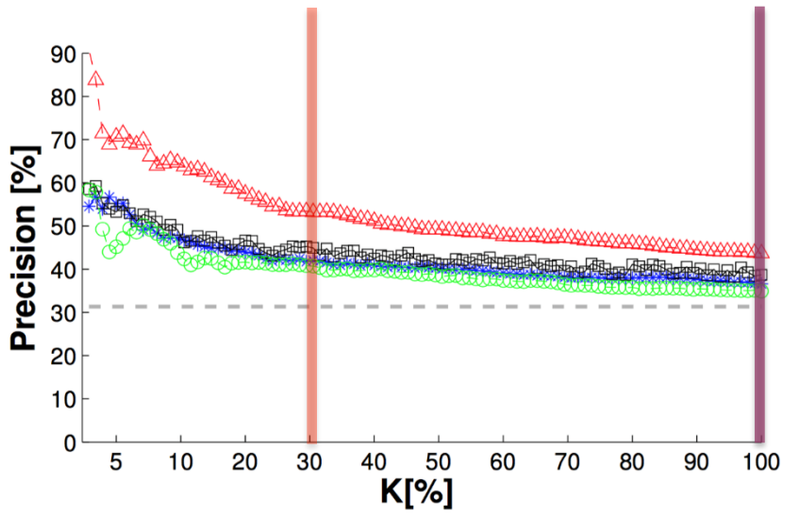


Sad

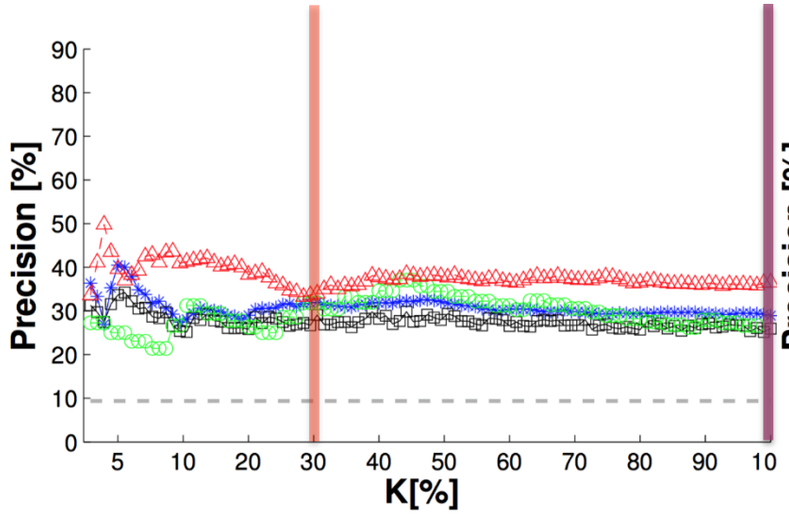




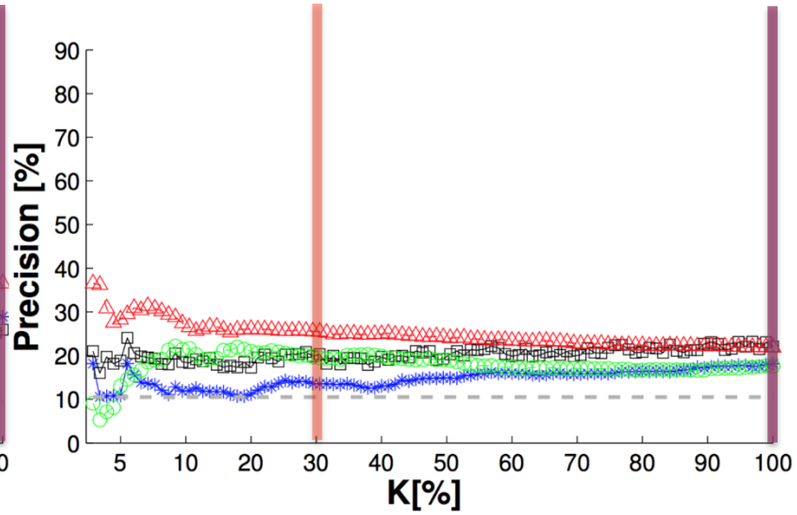
Accuracy of Retrieval



Happy



Angry



Sad

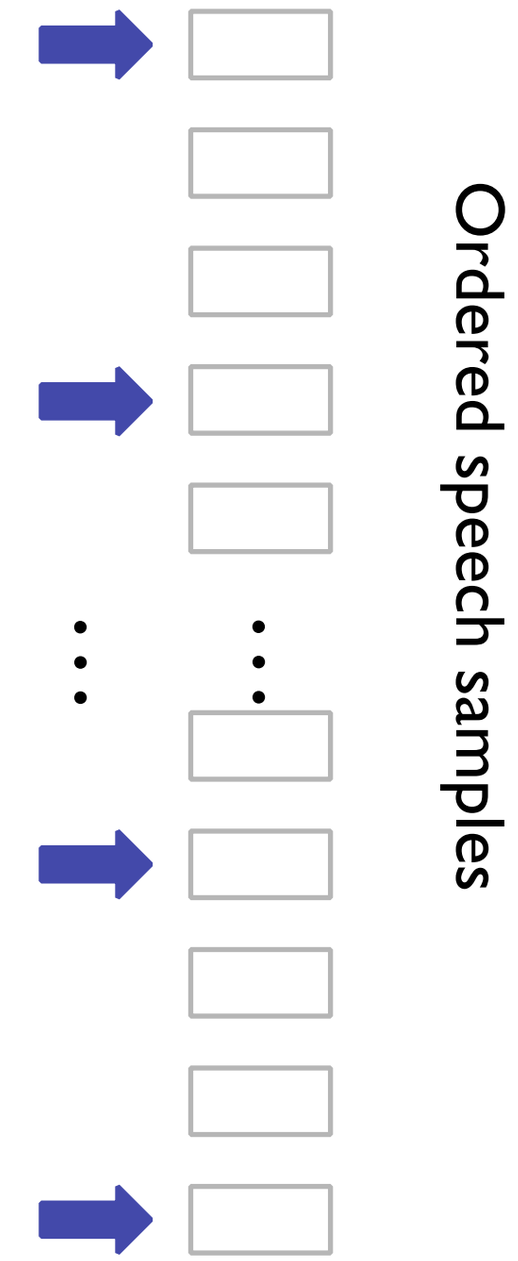
Cao et al. 2015(SVM) — Cao et al. 2015(GP) — Relevance score(SVM) — Relevance score(GP) — Random

- Label based ranking
- Relevance score ranking

Emotion Category	Cao et al. [2015] P@30	Cao et al. [2015] P@100	Relevance score P@30	Relevance score P@100
Happy	42.4	37.1	53.4*	43.7*
Angry	31.8	28.9	36.5	33.6*
Sad	20.1	22.0	25.6	21.9

Performance of Ranking

- Select 20 sample
- Find if the order is correct
  - Baseline: relevant score estimated for test set



Kendall rank correlation coefficient for Gaussian process preference learning

Emotion Category	Cao et al. [2015]	Relevance score
Happy	0.194	0.242
Angry	0.118	0.126
Sad	0.158	0.194

Conclusion and Future Work

- We proposed relative labels used in preference learning for categorical emotions
  - Using raw annotations instead of only consensus labels
- Higher precision than consensus label based ranking and binary classification

We have better relative labels!!!

- Future work:
  - Consider reliability of annotators in relevance score
  - Deep neural network based preference learning
  - Consider arousal and valence (Multi-task learning)

References

[1] Reza Lotfian and Carlos Busso, "Practical considerations on the use of preference learning for ranking emotional speech," ICASSP 2016, Shanghai, China, March 2016.

[2] H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech,” Computer Speech & Language, vol. 29, no. 1, pp. 186–202, January 2014.

[3] J. Kittler, “Combining classifiers: A theoretical framework,” Pattern Analysis & Applications, vol. 1, no. 1
, pp. 18–27, March 1998

[4] T. Joachims, “Training linear SVMs in linear time,” in ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, USA, August 2006, pp. 217–226.

[5] W. Chu and Z. Ghahramani, “Preference learning with gaussian processes,” in Proceedings of the 22nd international conference on Machine learning. ACM Press, 2005, pp. 137–144.

Thanks for your attention!

http://msp.utdallas.edu/