Verbal Behavior Event Detection Using Textual and Acoustic Semantics

Asif Salekin

Sarah Masud Preum

Motivation: Detecting Verbal Agitation

- Agitation affects people with dementia, autism, Alzheimer
- 65% demented elderly patients are hospitalized due to agitation
- **Manage cognitive disorder**
- Solution:
  – continuous monitoring by caregivers
  – automatic detection

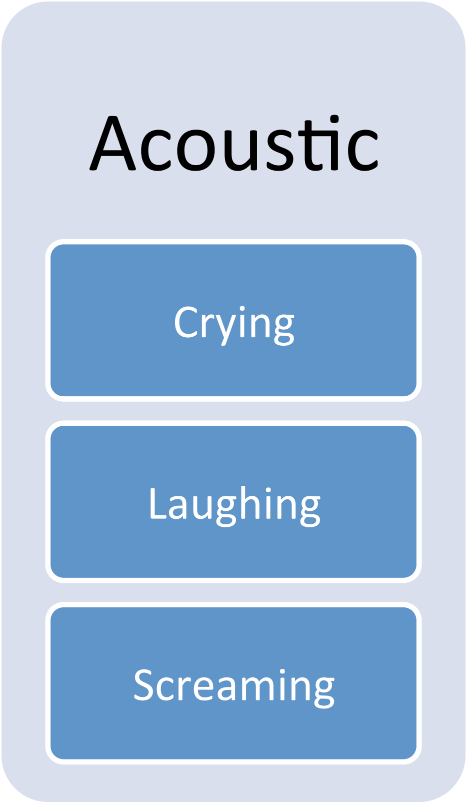
Motivation: Verbal Agitation Metrics

Cohen Mansfield inventory for verbal agitation

1. Crying
2. Laughing
3. Screaming
4. Negativism
5. Cursing
6. Saying repetitive sentence
7. Asking for help
8. Making sexual advance

Related Work: Detect Verbal Agitation Events Using **Acoustic** Features

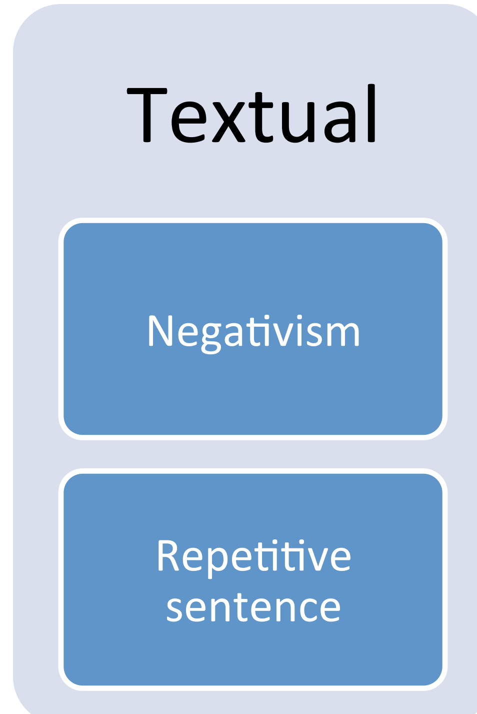
- [W. Huang, 2010], [L.S. Kennedy, 2004], [S. Petridis, 2008], [K. P. Truong, 2007]



Related work: Detect Verbal Agitation Events Using **Textual** Features

- Detecting negativism is equivalent to sentiment analysis [M. Weigand, 2010] [B. Pang, 2008], [T. Wilson 2005]

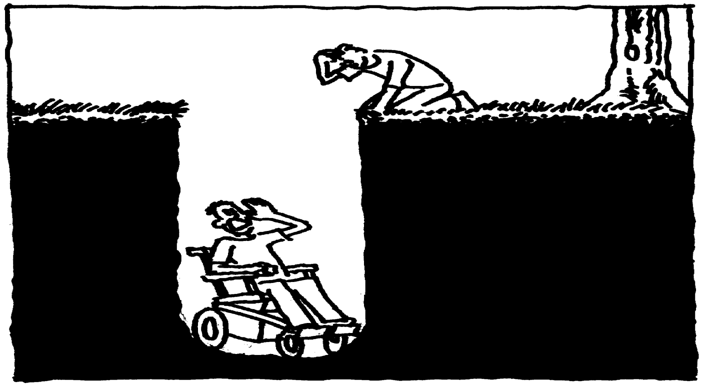
- Detecting repetitive sentence is a sequence mining problem: detect recurring subsequences [J. Pei, 2004]



Goals

- Detect **cursing**
- Detect **asking for help** *Patients in hospital*
- Detect **verbal sexual advances** *Detect sexual harassment in office environment*

Less explored but important

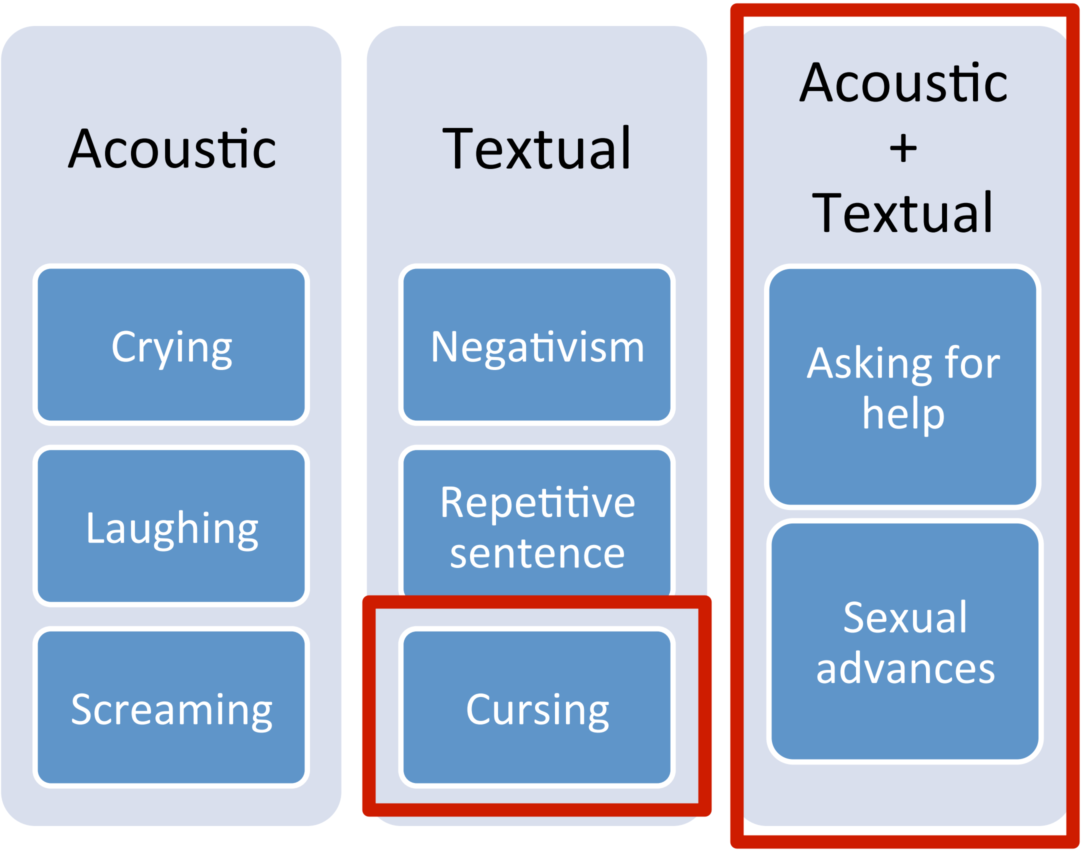




Challenges

- Cursing: Ambiguity in word with multiple senses
  - *"I rode my ass up the mountain"* vs *"Stop being an ass!"*
- Asking for help: Textual content can often be misleading, need acoustic semantics (tone of the speaker)
  - In a urging tone: "Please help me!"
  - In a neutral tone : "The man asked your help"
- Verbal sexual advances
  - Similar challenges

Goals and Challenges

- Detect three events from the Cohen Mansfield agitation inventory for verbal agitation

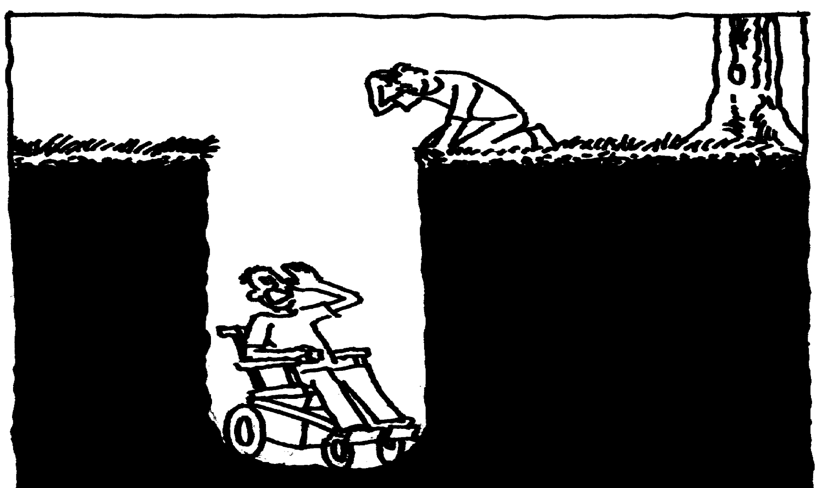


Detecting

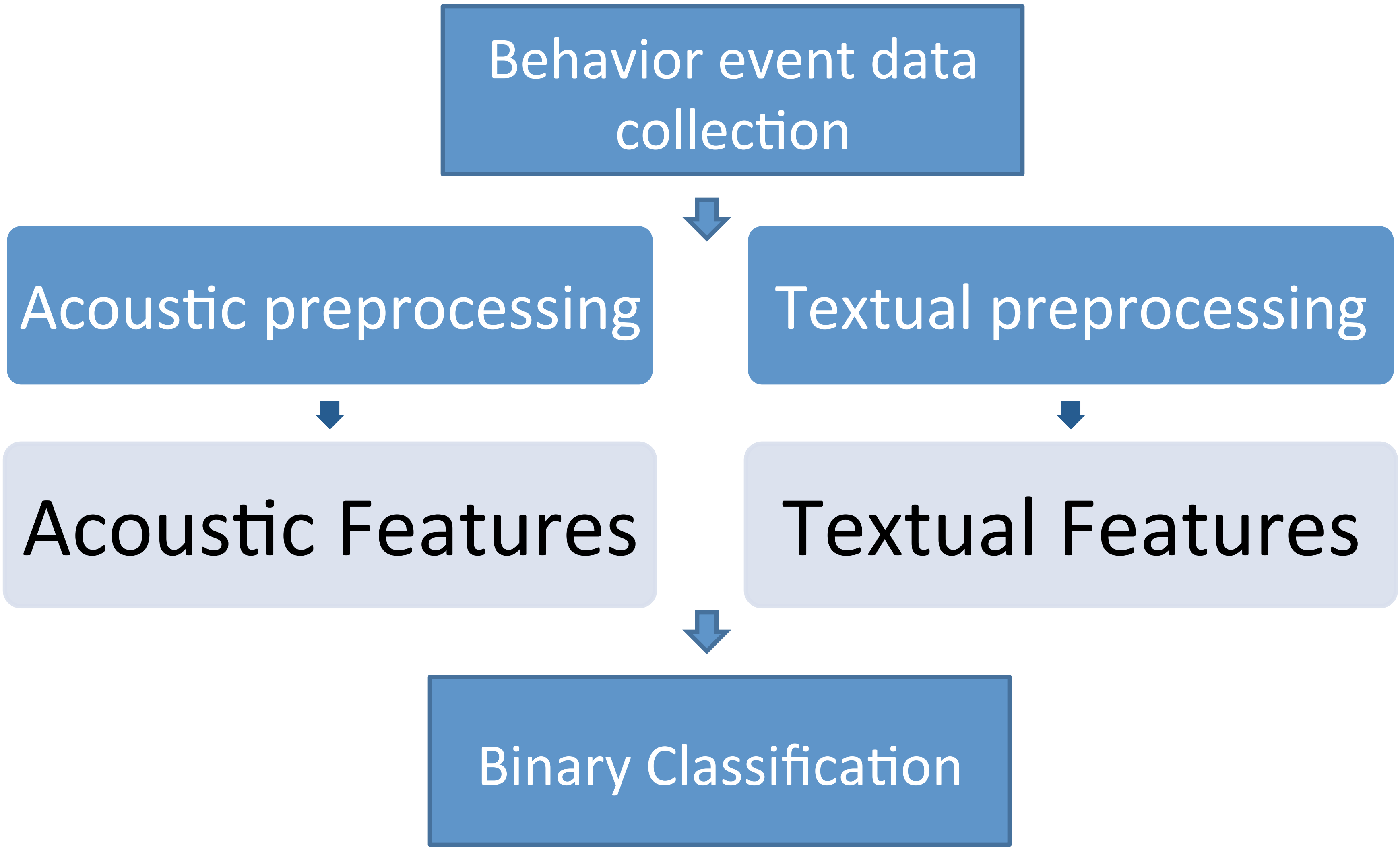
Asking For Help

&

Verbal Sexual Advances



Overview of Approach



Extracting Acoustic Features

- "human behaviors remain consistent with the specific emotion concepts" [Y. Zemack-Rugar, 2007]
  Verbal sexual advance → arousal

- Goal: **Represent the emotional concepts reflected in the tone of speech.**

<table>
<tr><th>Acoustic features</th><th>Role</th></tr>
<tr><td>Zero crossing rate</td><td>Detect voice vs non voice</td></tr>
<tr><td>Harmonic-to-noise ratio</td><td>Anger vs non anger</td></tr>
<tr><td>Energy</td><td rowspan="2">Arousal vs non arousal</td></tr>
<tr><td>Pitch</td></tr>
<tr><td>F0 fundamental frequency</td><td>Joy-surprise vs disgust-anger</td></tr>
</table>

Processing Text Data

- Speech to text conversion:
  Dragon NaturallySpeaking 12 (95%-99% accuracy)
- Stop word list reduction :
  some stop words are important features in the problem domain (e.g., help, please)
- Stemming: Porter stemmer
- Normalization: punctuation, case conversion

Extracting Textual Features

Text document: converted text from audio clips

Document 1: please please help me

....

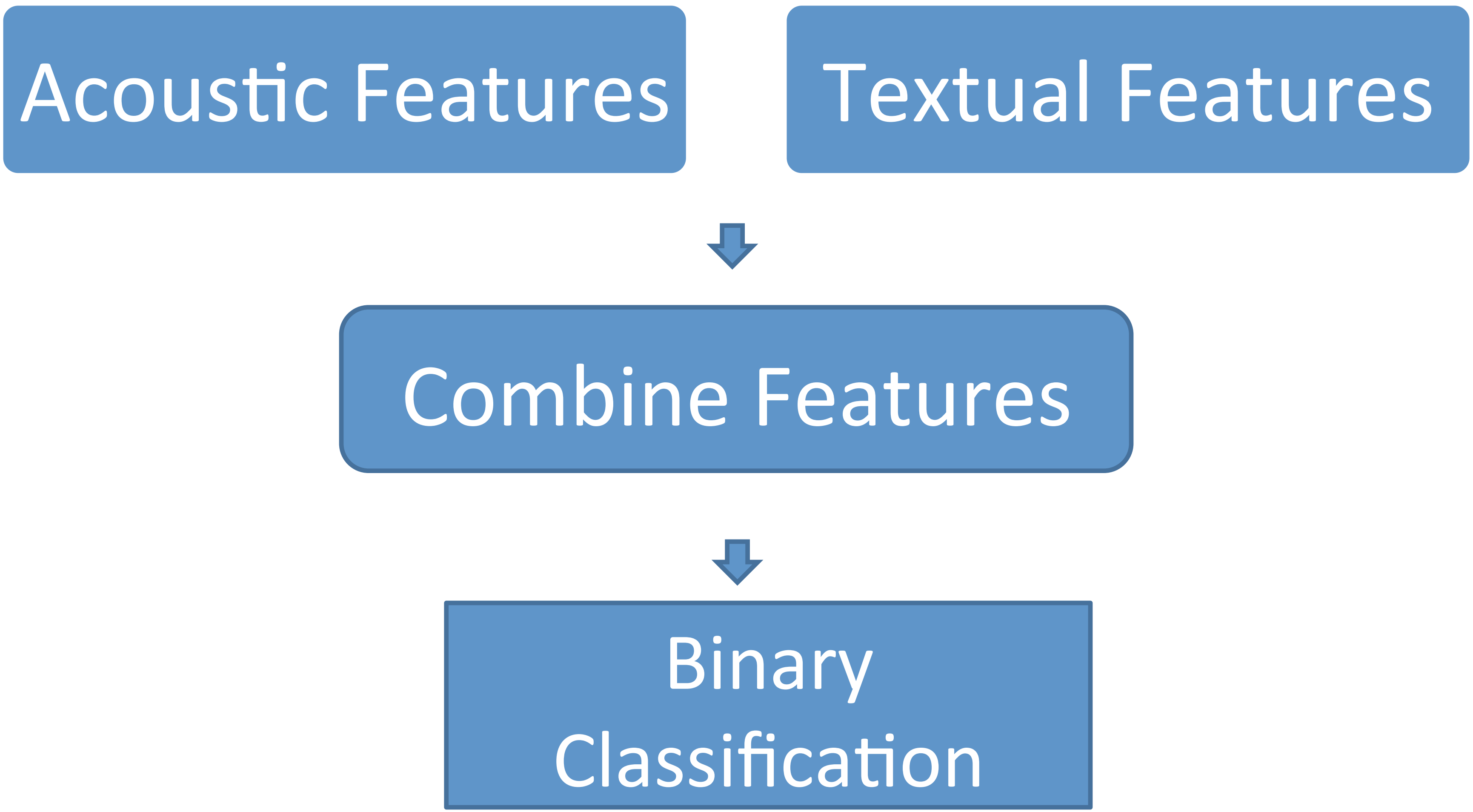
Bag of word representation: smoothed vector space of words

$$\hat{tf}(w,d) = \begin{cases} 1 + \log tf(w,d) & \text{if } tf(w,d) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$idf(w) = 1 + \log\left(\frac{N}{DF(w)}\right)$$

...	help		please	...	...	me
0	4.2	0	3	0	0	0.2

Combined Feature Space Formation



Detecting Cursing

Overview of Approach

Behavior Event Data Collection

↓

Speech to Text Conversion

↓

Textual preprocessing

↓

Extracting Textual Features

↓

Binary Classification

Resolving Ambiguity of Curse Detection

- Generate curse dictionary of 165 words:
  - http://www.noswearing.com/dictionary/d
  - http://en.wiktionary.org/wiki/Category:English_swear_words
- 36 Ambiguous words: dog, ass, etc.
- Word sense disambiguation
  - WordNet knowledge base
  - Modified Lesk algorithm

Resolving Ambiguity of Curse Detection: An Example

- *"I rode my ass up the mountain"* vs *"Stop being an ass!"*

		Word sense
Non curse	1	Hardy and sure footed animal smaller and with longer ears that horse
	2	The fleshy part of the human body that you sit on
Curse	3	A pompous fool
	4	Slang for sexual intercourse

- Adapted Lesk Algorithm -> modification
  - *Multiclass problem* *Binary class Problem*

Experiment Design

- Data
  - Controlled experiments with 4 volunteers
  - Movie clip extracts
- Ground truth: manual labeling
- Performance metrics: accuracy, F-1
- Analysis: kappa statistics for confidence analysis

Result: Detecting "Asking for Help"

- Using only acoustic features

Classifier	Accuracy	Kappa Statistics	F1-Measure
Naïve Bayes	75.8	0.42	0.75
K-nearest neighbor	**80.9**	**0.57**	**0.81**
Random forest	80.2	0.50	0.79

- Using only textual features performs even worse
- Using both acoustic features and textual features

Classifier	Accuracy	Kappa Statistics	F1-Measure
Naïve Bayes	84.7	0.67	0.86
K-nearest neighbor	83.5	0.62	0.84
Random forest	**89.6**	**0.75**	**0.89**

~11% increase ~10% increase

Result: Detecting "Verbal Sexual Advances"

- Using only acoustic features

Classifier	Accuracy	Kappa Statistics	F1-Measure
Naïve Bayes	71.4	0.42	0.72
K-nearest neighbor	80.3	0.61	0.81
Random forest	79.7	0.61	0.80

- Using only textual features performs even worse
- Using both acoustic features and textual features

Classifier	Accuracy	Kappa Statistics	F1-Measure
Naïve Bayes	73.2	0.45	0.72
K-nearest neighbor	76.4	0.53	0.74
Random forest	**86.6**	**0.72**	**0.87**

~8% increase

~7% increase

Result: Cursing

- Baseline 1: Use all words in curse dictionary:
- Baseline 2: Use only single meaning words in dictionary as curse

	Precision	**Recall**	**F-measure**
Baseline 1	0.73	1	0.84
Baseline 2	1	0.74	0.85
Our approach	**0.95**	**0.96**	**0.96**

Analyzing Results: Identifying New Challenge

- Random forest better than kNN, Naïve Bayes
  - Probably data not linearly separable
- Curse of Dimensionality??
  - Small dataset but large feature space
- Potential Solution:

  Evaluate feature reduction: PCA

Principle Component Analysis (PCA)

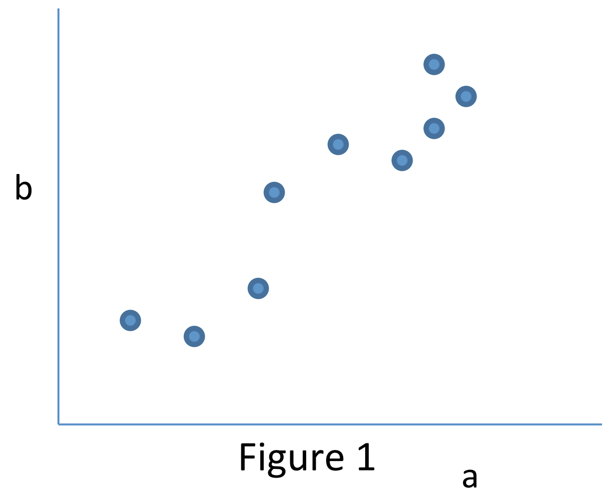


Figure 1

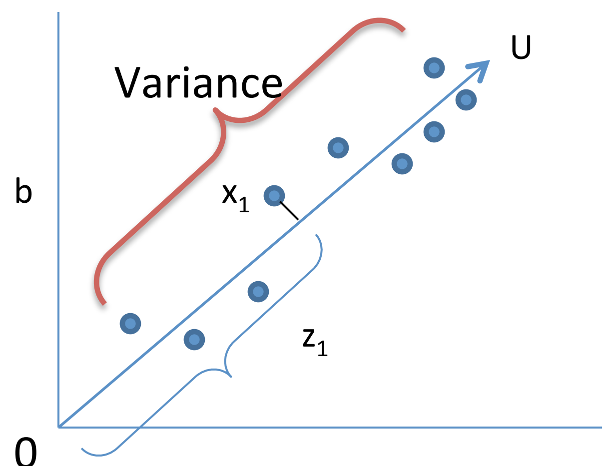


Figure 4

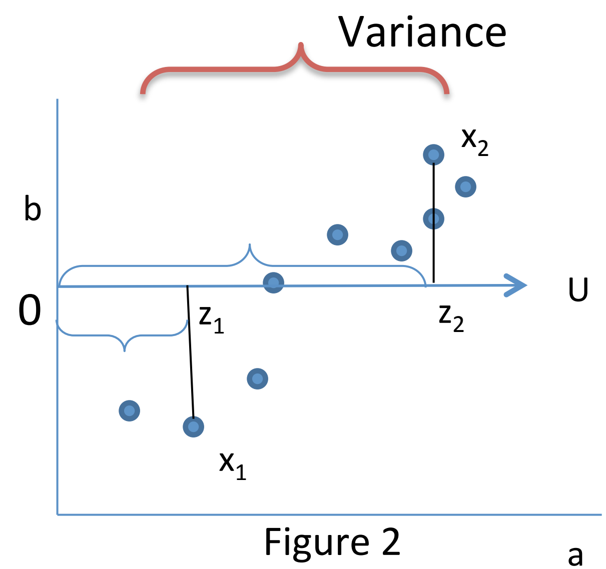


Figure 2

U

Problem: PCA

Results:

<u>Asking for help:</u>

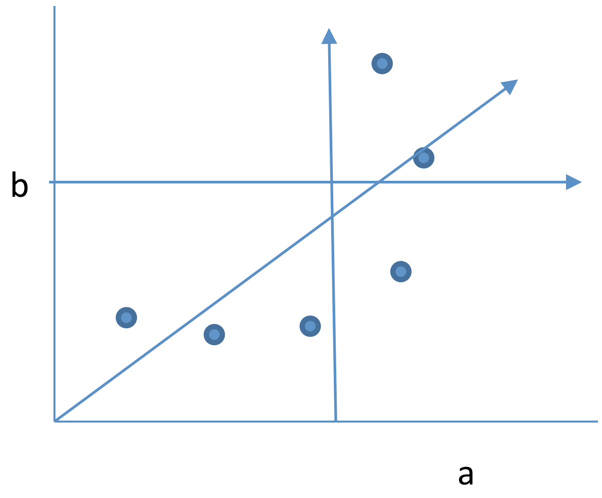
Reduced Feature number: 126

Accuracy: from 89.6% to 84%

<u>Verbal sexual advances:</u>

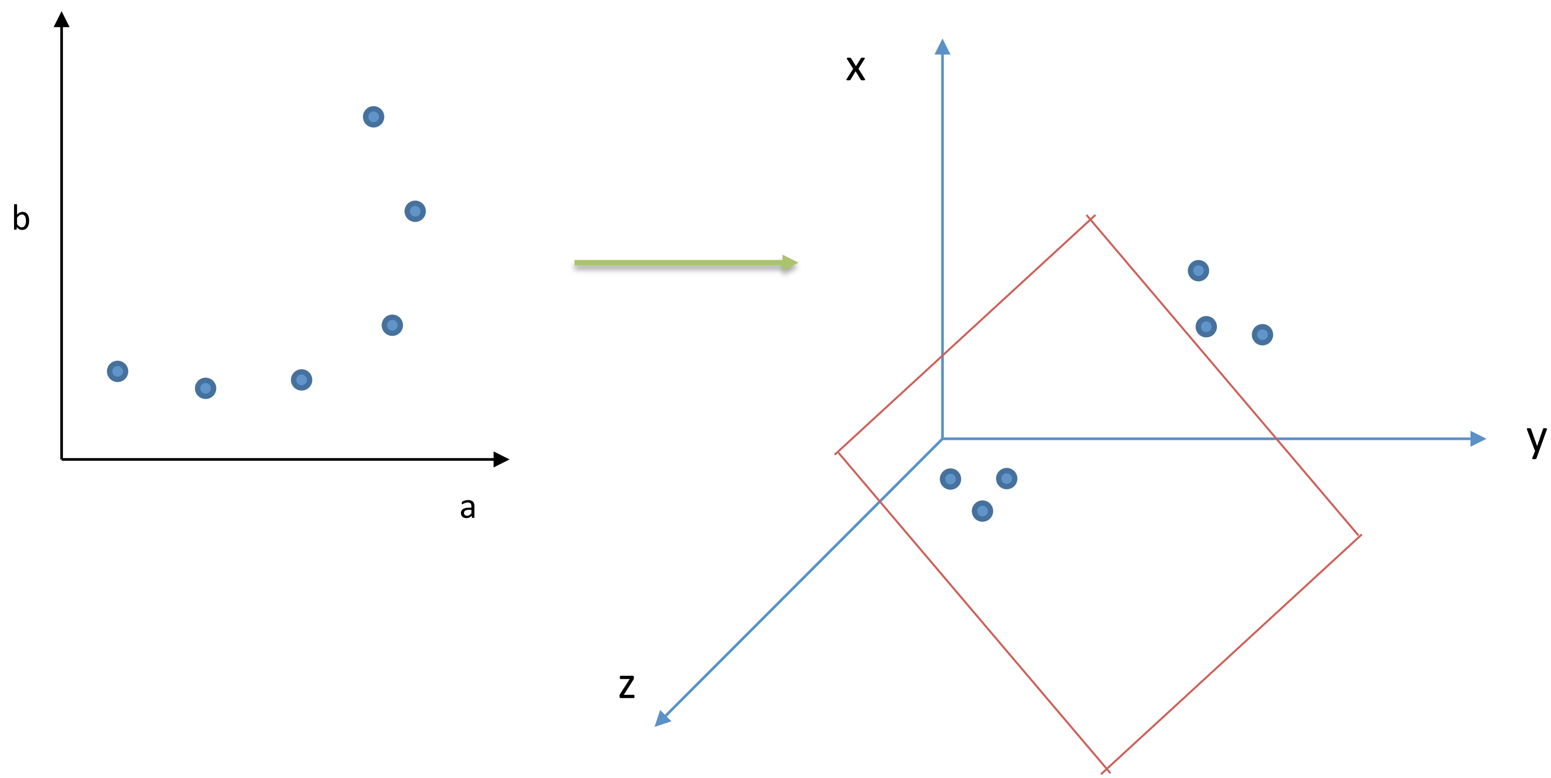
Reduced Feature number: 213

Accuracy: from 86.6% to 80%



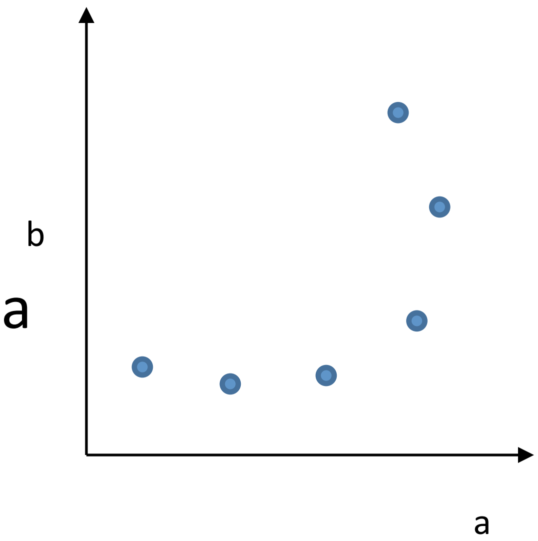
Project to higher dimension with RBF kernel

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^{n}(x_j - l_j^{(1)})^2}{2\sigma^2}\right)$$



Kernel PCA

- Increase the dimension up to number of training data
- Perform PCA on that higher dimension data
- Results:
- <u>Asking for help:</u> Accuracy from 89.6% to 73%
- <u>Verbal sexual advances:</u> Accuracy from 86.6% to 67.8%
- Limitation
  - We have limited number of training data so, can not increase dimension!!!
  - Hence, we need more data to use KPCA.



LPCA Steps
6 dimensions
Perform PCA on
6 dimensional data

Conclusion

- Combine text mining and signal processing
- First to detect cursing, asking for help, verbal sexual advances
- Textual features enhances classification performance
- Future works:
  - Evaluate on a larger dataset: validity of feature reduction
  - Include more contextual features

Thanks!

Question?