Towards Ambient Intelligence in Smart Healthcare

John A. Stankovic
BP America Professor
Director, Link Lab
Department of Computer Science
University of Virginia
How I Got Here

• Teaching
  – High School Tutoring
  – Math and Engineering (lots of Math) – WHY?

• Research
  – ABM
My Work - Main Themes

• Wearables (Smart Watches) and In-situ

• Cognitive Assistance (ML and NLP)

• Conflict Detection (NLP)

• Acoustics (ML)

• Real Deployments
Current/Recent Projects

- Family Eating Dynamics for Obese Families - USC/Los Angeles – 23 families
- Smart Watch Reminder Systems – UVA Center for Telemedicine
- Alzheimer’s Patient – Caregiver Interaction – Ohio State and Univ. of Tennessee
- Smart Watch Handwashing Conformance, UVA PICU
- First Responders Cognitive Assistant – North Garden Fire and Rescue/ Richmond Fire, and Oxford Univ.
New Themes

• Use FM to integrate properties into ML models

• Address Uncertainties in ML Model Predictions
IOHT - Hype or Revolution

- Smart Watches
  - More and more sold; more and more sensors
- Smart Skin
- In-situ
- Smart Textiles
- …
Vision

An ambient healthcare intelligence

WebMD
Big Data Collections
ML/Analytics
Data Mining

General Population

Actuations

Nano-pills
Pacemaker

Holistic

Wearables
Sensors
Today’s Main Themes

• Wearables (Smart Watches)

• Cognitive Assistance (ML and NLP)
  – In situ => mood at distance
  – In situ => anxiety via MIL

Towards Ambient Healthcare Intelligence

Powerful Modality
Cognitive Assistance (on a smart watch)

- Towards General Healthcare Intelligence (comprehensive services)
  - Interact with Internet Healthcare Services
  - Support Conversations (esp. for elderly)
  - Reminders, suggestions, alarms - explainable
  - Physical and Mental Health
  - Pandemic-aware
  - Privacy-aware – not discussed today
iAdhere – verbal medication and exercise reminder system

For stroke patients

Using Apple Watch – with microphone and speaker

Applying in a Telemedicine setting
iAdhere

• Medication and exercise reminders
  – Supports general verbal questions
  – Allows rescheduling
  – Current: quality of exercise and pain

• EKG

• General healthcare dialogue support

Towards comprehensiveness
Demo – A Few Features

Earlier version called Medrem
Services Expanded for Pandemics

- Collections of services on a smart watch
  - Handwashing (or general hygiene/elderly)
  - Mood/Depression/Anxiety/Loneliness
  - Voice based conversations
    - Pandemic info
    - Reminders/Alerts/Advice
  - Physiological parameters and more
    - Symptoms
  - ...
Quality of Handwashing

- WHO guidelines
- Quality
- Solution: Hybrid CNN-RNN

- Supports conversations
  - Reminders based on time and when return home (beacons)
  - Info on quality of handwashing
How to handrub?
WITH ALCOHOL-BASED FORMULATION

1a
Apply a palmful of the product in a cupped hand and cover all surfaces.

1b

2
Rub hands palm to palm

3
right palm over left dorsum with interlaced fingers and vice versa

4
palm to palm with fingers interlaced

5
backs of fingers to opposing palms with fingers interlocked

6
rotational rubbing of left thumb clasped in right palm and vice versa

7
rotational rubbing, backwards and forwards with clasped fingers of right hand in left palm and vice versa

8
rinse hands with water

9
dry thoroughly with a single use towel

10
use towel to turn off faucet

How to handwash?
WITH SOAP AND WATER

0
Wet hands with water

1
apply enough soap to cover all hand surfaces.
Smartwatch App

- Wash your hands
- OK I will wash my hands now
Dialogue

It’s time to wash your hands

Ok, I will do it

After handwashing...

You did a good job!
But you should rub hands more, and wash for minimum 20 seconds

Thanks for your feedback

You are welcome
Solution – A Hybrid DNN

Evaluation

- Our own dataset

- 14 participants
  - Each 19 HW sessions

- 3 practice runs

- Video for Ground Truth
Acoustics: Exploiting Speech

- Distance Emotion Recognition
  - Happy, sad, angry, neutral

- Anxiety and Depression

Mental Health
Distance Emotion Recognition

Close to microphone  Fixed distance
24/7

A realistic indoor speech emotion recognition system

- Reverberation
- Ambient noise
- De-amplification of speech
- Overlapping of speech
Solution

1. Distance Agnostic Features/code words

2. Feature Modeling: Emo2vec

3. Classifier: LSTM
Select Robust Features

Consider 231 LLD features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mel-Frequency cepstral coefficients (MFCC) 1-25</td>
<td>25</td>
</tr>
<tr>
<td>Root-mean-square signal frame energy</td>
<td>1</td>
</tr>
<tr>
<td>The voicing probability computed from the ACF</td>
<td>1</td>
</tr>
<tr>
<td>The fundamental frequency computed from the Cepstrum</td>
<td>1</td>
</tr>
<tr>
<td>Pitch</td>
<td>1</td>
</tr>
<tr>
<td>Harmonics to noise ratio (HNR)</td>
<td>1</td>
</tr>
<tr>
<td>Zero-crossing rate of time signal</td>
<td>1</td>
</tr>
<tr>
<td>PLP cepstral coefficients compute from 26 Mel-frequency bands</td>
<td>6</td>
</tr>
<tr>
<td>The 8 line spectral pair frequencies computed from 8 LPC coefficients</td>
<td>8</td>
</tr>
<tr>
<td>Logarithmic power of Mel-frequency bands 0 - 7</td>
<td>32</td>
</tr>
</tbody>
</table>

Select 48 LLD features:
- 5 MFCC
- Voice probability
- Fundamental frequency
- Zero crossing rate
- 8 line spectral pair frequencies
- 32 logarithmic power of Mel-frequency bands

Delta and delta-delta of these 77 features
Audio (Code) Words

Audio segment

2 Dimensional Feature space

25 ms Small overlapping windows

[X,Y]
That means total $C$ words in our vocabulary.
Code Book Sizes

• Tested 500 to 2500 in increments of 500
  – K-means clustering

• Interesting Result: Different code book sizes for different emotions
Adaptation of Word2Vec: Emo2vec

- Convert audio words into vectors
- Words which occur in similar context (that means with similar neighbor words), for a specific emotions have similar vector representations.

\[(\text{word}, \{\text{Neighbour set}\}) \qquad (\text{word}, \{\text{Neighbour set}\})\]

\[(A, \{P, Q, R, S, T, U, V, W, M\}) \quad \vdots \quad (C, \{P, Q, R, S, T, U, V, W, X\})\]

\[(A, \{P, R, Q, S, T, U, V, W, N\}) \quad \vdots \quad (C, \{P, R, Q, S, T, U, V, W, N\})\]

\[(B, \{O, P, Q, R, S, T, U, V, W\}) \quad \vdots \quad (C, \{P, R, Q, S, T, U, V, M, N\})\]

\[(B, \{P, Q, R, S, T, N, U, V, W\}) \quad \vdots \quad (E, \{F, X, P, Y, Z, S, T, W\})\]

\[(B, \{P, R, S, T, U, V, W, M, Q\}) \quad \vdots \quad (F, \{A, P, E, G, H, H, J, J\})\]

\[(D, \{E, F, E, G, H, E, B, C\}) \quad \vdots \quad (J, \{E, F, J, M, M, K, N, P\})\]

\[(D, \{G, H, F, E, J, I, G, W\}) \qquad (D, \{F, O, X, D, K, M, N, J\})\]

Input corpus of happy $D_H$  
Input corpus of Not happy $D_N$
LSTM Classifier

100 Neurons

Dense output Layer

Separate for each emotion

20% Dropout Regularization (training)
Evaluation

• 2 literature datasets

• Our own family-discussion experiments
  – 12 families; 28 people
  – Spontaneous discussions
  – Similar performance to literature datasets

• 4 Baselines
  – Approximately 16% better than best baseline

Angry, Happy, Sad

Emo2vec approach on datasets
Observation

• In the past – Happy versus Angry difficult (acted datasets)

  – Why? - Use of Energy based features

  – In real setting: Laughter helps discriminate
Elimination of distorted features helpful?
Effect of Distance

• As we move from mic to 6m away from mic, drop in accuracy is about 5%

• State of art: the drop is about 12%
Mental Disorder - Anxiety

• 11% of Americans suffer from Anxiety

• Prior work – most use fully supervised learning
  – But what parts of the speech are representing anxiety
  – Very difficult to label
Contributions

- MIL $\Rightarrow$ weakly supervised learning
- Novel feature modeling $\Rightarrow$ NN2vec
- New classifier $\Rightarrow$ BLSTM-MIL
- 90% F-1 and accuracy
- 17% better than baselines
Weakly Labeled Data

Positive sample (from person with mental disorder)

Indicates anxiety disorder

True Positive

Indicates anxiety disorder

False Positive

Indicates anxiety disorder

True Negative
Positive Audio Clip

Segmentation → Feature Extraction → Feature Modeling

Code Words

Relationship among code words

NN2VEC (vector embedding)

and

Learned via DNN
NN2VEC Feature Modeling

Region indicative to Positive class: mental disorder

Frequent in Positive class

Frequent in Negative class

Positive Example
Classifier

• Prior Work (not RNNs)
  • SVM – MIL
  • DNN – MIL

• Fail to account for temporal dynamics in speech segment
BLSTM - MIL
**Evaluation**

<table>
<thead>
<tr>
<th>Social Anxiety</th>
<th>Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 105 Participants</td>
<td>• Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ)</td>
</tr>
<tr>
<td>• Mean Age: 19.24, SD: 1.84</td>
<td>• 142 participants</td>
</tr>
<tr>
<td>• Mean audio clip length: 3 minutes</td>
<td>• Mean audio clip length: 12 minutes</td>
</tr>
<tr>
<td>• Labeled by Licensed Clinical Psychologists</td>
<td></td>
</tr>
<tr>
<td>• Social Interaction Anxiety Scale (SIAS) and Social Phobia Scale (SPS)</td>
<td></td>
</tr>
</tbody>
</table>

Social Anxiety

Evaluation on Length of Instance

<table>
<thead>
<tr>
<th>Audio States</th>
<th>F-1 Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>77.2</td>
<td>78.9</td>
</tr>
<tr>
<td>1000</td>
<td>82.8</td>
<td>84.1</td>
</tr>
<tr>
<td>2000</td>
<td>87.9</td>
<td>89.1</td>
</tr>
<tr>
<td>2500</td>
<td>88.17</td>
<td>89.1</td>
</tr>
<tr>
<td>3000</td>
<td>89.13</td>
<td>90.1</td>
</tr>
<tr>
<td>3500</td>
<td>90.1</td>
<td>91</td>
</tr>
<tr>
<td>4000</td>
<td>90.1</td>
<td>91</td>
</tr>
<tr>
<td>4500</td>
<td>89.1</td>
<td>90</td>
</tr>
<tr>
<td>5000</td>
<td>88.17</td>
<td>89.1</td>
</tr>
</tbody>
</table>

Evaluation on Number of Audio States
Social Anxiety

15.4% improvement F-1 score compared to the best baseline in literature: \textit{I-vector with BLSTM}
Summary

• Smart and Connected Health Based on Wearables and in-situ systems

• Towards Ambient Intelligence: Cognitive Assistance for Healthcare

• Key modality: Acoustics/Speech
Thanks

• Abu Mondol, now at Amazon

• Sirat Samyoun, at UVA

• Asif Salekin, now at Syracuse

• Meiyi Ma, now at Vanderbilt

• Sarah Preum, now at Dartmouth
Real Deployment Observations

• 70% non-research software development

• Significant effort at daily monitoring of deployed systems

• Publication lag

• Where/what to publish
Future Work

• Multiple emotions at the same time
• Rare emotions
  – Frustration versus anger
• Longer distances from microphone
• Use of linguistics content
• Anxiety with elderly
• Chronic stress
Future Work

- Greater and greater intelligence
- Better support for self-help, e.g., dealing with conflicts
- Privacy