Identifying Sentiment and Emotion in Low Resource Languages

Julia Hirschberg and Zixiaofan (Brenda) Yang

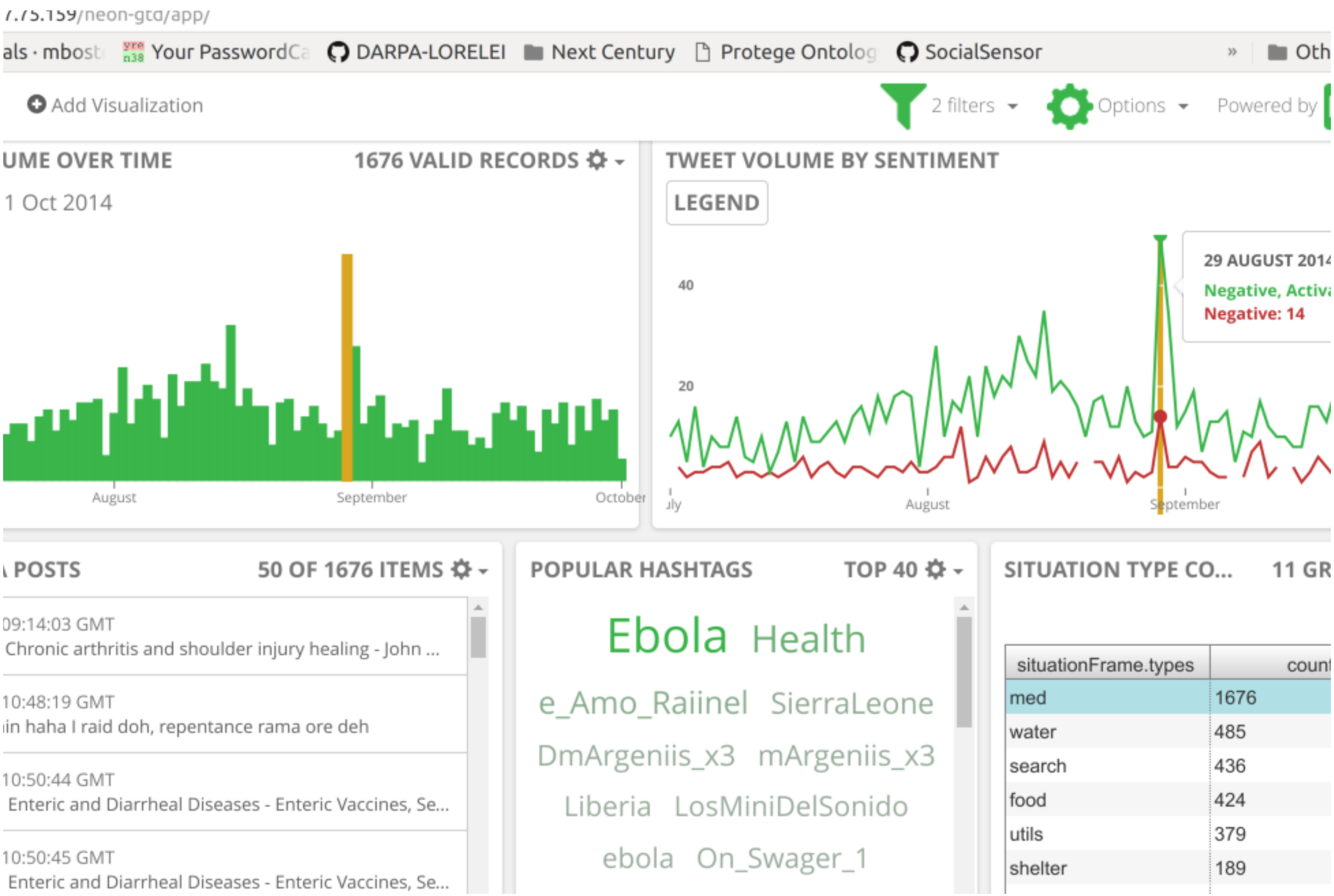
Columbia University

Humanitarian Assistance and Disaster Relief (HADR) in DARPA Lorelei

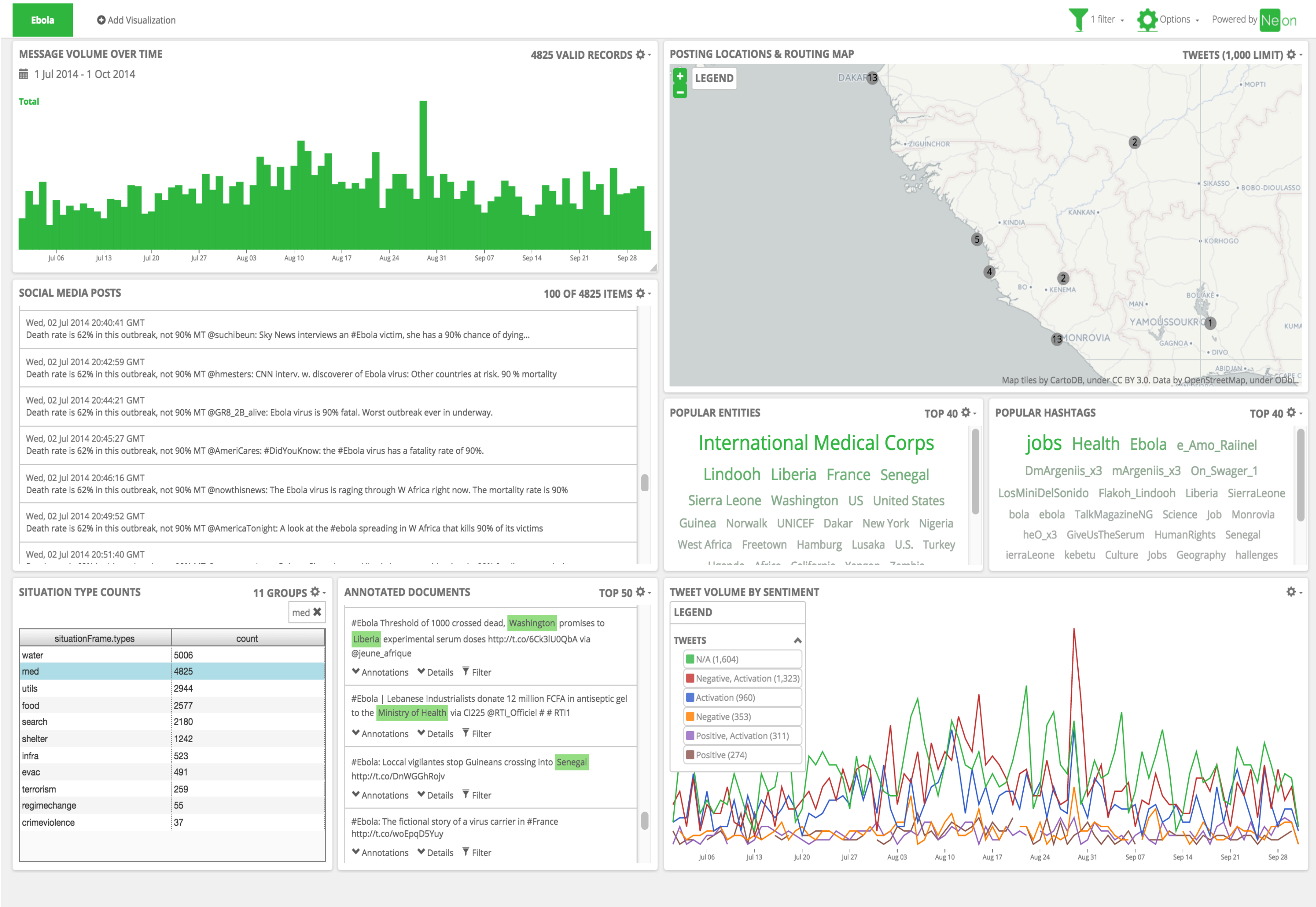
- In disaster situations, international responders may not speak the language of the area in distress and may have little ***reliable*** access to local informants
  - 7100+ active languages in the world -- hard to predict which languages will be needed next
    - 44 in Boko Haram area (Hausa, Kanuri) ~522 languages in all of Nigeria
    - 19 in Ebola outbreak areas in Liberia, Sierra Leone, and Guinea
    - 20+ Mayan languages spoken by Central American refugee children
  - Current methods require 3 years and $10M’s per language (mostly to prepare training corpora)
    - Would require $70B and 230K person-years to handle all languages

Challenge

- How can we develop language technologies quickly to help first responders understand text and speech information vital to their mission (social media, hotline msgs, news broadcasts)?
  - Triage information by urgency and sentiment/emotion (anger, stress, fear, happiness)
  - Display information in a form that relief workers can easily understand



7.75.159/neon-gtd/app/
als · mbost
Your PasswordCa
DARPA-LORELEI
Next Century
Protege Ontolog
SocialSensor
Oth
Add Visualization
2 filters
Options
Powered by
UME OVER TIME
1676 VALID RECORDS
1 Oct 2014
August
September
October
TWEET VOLUME BY SENTIMENT
LEGEND
40
20
29 AUGUST 2014
Negative, Activ
Negative: 14
ly
August
September
POSTS
50 OF 1676 ITEMS
09:14:03 GMT
Chronic arthritis and shoulder injury healing - John ...
10:48:19 GMT
in haha I raid doh, repentance rama ore deh
10:50:44 GMT
Enteric and Diarrheal Diseases - Enteric Vaccines, Se...
10:50:45 GMT
Enteric and Diarrheal Diseases - Enteric Vaccines, Se...
POPULAR HASHTAGS
TOP 40
Ebola
Health
e_Amo_Raiinel
SierraLeone
DmArgeniis_x3
mArgeniis_x3
Liberia
LosMiniDelSonido
ebola
On_Swager_1
SITUATION TYPE CO...
11 GR
situationFrame.types
count
med 1676
water 485
search 436
food 424
utils 379
shelter 189


Ebola
Add Visualization
1 filter
Options
Powered by Neon
MESSAGE VOLUME OVER TIME
4825 VALID RECORDS
1 Jul 2014 - 1 Oct 2014
Total
Jul 06 Jul 13 Jul 20 Jul 27 Aug 03 Aug 10 Aug 17 Aug 24 Aug 31 Sep 07 Sep 14 Sep 21 Sep 28
POSTING LOCATIONS & ROUTING MAP
TWEETS (1,000 LIMIT)
LEGEND
DAKAR
ZIGUINCHOR
KINDIA
KANKAN
SIKASSO
BOBO-DIOULASSO
KORHOGO
BO
KENEMA
MAN
BOUAKÉ
YAMOUSSOUKRO
MONROVIA
GAGNOA
DIVO
ABIDJAN
Map tiles by CartoDB, under CC BY 3.0. Data by OpenStreetMap, under ODbL.
SOCIAL MEDIA POSTS
100 OF 4825 ITEMS
Wed, 02 Jul 2014 20:40:41 GMT
Death rate is 62% in this outbreak, not 90% MT @suchibeun: Sky News interviews an #Ebola victim, she has a 90% chance of dying...
Wed, 02 Jul 2014 20:42:59 GMT
Death rate is 62% in this outbreak, not 90% MT @hmesters: CNN interv. w. discoverer of Ebola virus: Other countries at risk. 90 % mortality
Wed, 02 Jul 2014 20:44:21 GMT
Death rate is 62% in this outbreak, not 90% MT @GR8_2B_alive: Ebola virus is 90% fatal. Worst outbreak ever in underway.
Wed, 02 Jul 2014 20:45:27 GMT
Death rate is 62% in this outbreak, not 90% MT @AmeriCares: #DidYouKnow: the #Ebola virus has a fatality rate of 90%.
Wed, 02 Jul 2014 20:46:16 GMT
Death rate is 62% in this outbreak, not 90% MT @nowthisnews: The Ebola virus is raging through W Africa right now. The mortality rate is 90%
Wed, 02 Jul 2014 20:49:52 GMT
Death rate is 62% in this outbreak, not 90% MT @AmericaTonight: A look at the #ebola spreading in W Africa that kills 90% of its victims
Wed, 02 Jul 2014 20:51:40 GMT
POPULAR ENTITIES
TOP 40
International Medical Corps
Lindooh Liberia France Senegal
Sierra Leone Washington US United States
Guinea Norwalk UNICEF Dakar New York Nigeria
West Africa Freetown Hamburg Lusaka U.S. Turkey
POPULAR HASHTAGS
TOP 40
jobs Health Ebola e_Amo_Raiinel
DmArgeniis_x3 mArgeniis_x3 On_Swager_1
LosMiniDelSonido Flakoh_Lindooh Liberia SierraLeone
bola ebola TalkMagazineNG Science Job Monrovia
heO_x3 GiveUsTheSerum HumanRights Senegal
ierraLeone kebetu Culture Jobs Geography hallenges
SITUATION TYPE COUNTS
11 GROUPS
med
situationFrame.types | count
water | 5006
med | 4825
utils | 2944
food | 2577
search | 2180
shelter | 1242
infra | 523
evac | 491
terrorism | 259
regimechange | 55
crimeviolence | 37
ANNOTATED DOCUMENTS
TOP 50
#Ebola Threshold of 1000 crossed dead, Washington promises to Liberia experimental serum doses http://t.co/6Ck3lU0QbA via @jeune_afrique
Annotations Details Filter
#Ebola | Lebanese Industrialists donate 12 million FCFA in antiseptic gel to the Ministry of Health via CI225 @RTI_Officiel # # RTI1
Annotations Details Filter
#Ebola: Loccal vigilantes stop Guineans crossing into Senegal http://t.co/DnWGGhRojv
Annotations Details Filter
#Ebola: The fictional story of a virus carrier in #France http://t.co/woEpqD5Yuy
Annotations Details Filter
TWEET VOLUME BY SENTIMENT
LEGEND
TWEETS
N/A (1,604)
Negative, Activation (1,323)
Activation (960)
Negative (353)
Positive, Activation (311)
Positive (274)
Jul 06 Jul 13 Jul 20 Jul 27 Aug 03 Aug 10 Aug 17 Aug 24 Aug 31 Sep 07 Sep 14 Sep 21 Sep 28

Our Goal

- Identify sentiment and emotion in written and spoken data to share with relief workers
  - Provide additional, extra-propositional meaning
    - Fear and stress of victims
    - Happiness at success of relief efforts
    - Anger at relief workers
  - Method: Develop ways to recognize and interpret sentiment and emotion in LRLs by training on High Resource Languages and other LRLs

Three Main Possibilities

- Can we recognize emotions relevant to Lorelei from labeled speech (e.g. anger, stress, fear)
- Can ***text-trained*** sentiment systems be used to label ***unlabeled*** speech transcripts to train sentiment recognition in speech?
- Can systems trained on emotion/sentiment in speech of one language be used to recognize emotion/sentiment in another?

Anger vs. Neutral

- Corpus: Mandarin Affective Speech
- Language : Mandarin
  - Neutral sentences (e.g. “It will rain tonight.”) and words (e.g. “train,” “apple”)
  - 5 basic emotions (neutral, anger, elation, panic, sadness) simulated by 68 students
- Our study: Anger: 5100 vs. Neutral: 5100

Feature Extraction Using openSMILE

- Baseline features (384)
    - ‘Standard’ simple low-level acoustic features
      (e.g., MFCC’s; max, min and mean frame energy)
    - ‘Unique’ features (e.g. slope and offset of a linear approximation of MFCC1-12)
  - Larger feature set (6552)
    - More Functionals and Low-Level Descriptors

Machine Learning Results

- Random forest: (*Scikit-learn*)
  - Train decision tree classifiers on various sub-samples of the training set using 384 feature set
  - Uses averaging to improve the predictive accuracy and control over-fitting
- **Weighted F-measure: 0.88 (0.50 baseline); P=.88; R=.88**

Useful Features

- Arithmetic mean and max value of MFCC[1] (mel frequency cepstral coefficients)
- The offset of linear approximation of root-mean-square frame energy
- Arithmetic mean and max value of MFCC[2]
- Range, max value, quadratic error and standard deviation of the 1st order delta coefficient of MFCC[1]
- Offset of linear approximation of MFCC[1]
- Arithmetic mean of root-mean-square frame energy

English: Stress vs. Neutral

- Corpus: SUSAS (Speech under simulated and actual stress)
  - Neutral words (e.g. “break” or “eight”) simulated by 9 speakers
  - Stress produced doing single tracking tasks
  - Stress: 630; Neutral: 631
- Classification result on random forest model: Weighted F-measure: **0.7031** (.50 baseline); P=.70; R=.70

Multi-labeled Semaine Corpus

- Queen’s U Belfast, http://semaine-db.eu
- Natural interactions in English between users and an ’operator’ simulating a Sensitive Artificial Listener (SAL) agent
- SAL agent examples：
  - ‘Do tell me all the delicious details.’
  - ‘Ohh.... that would be lovely.’
  - ‘What are your weaknesses?’
  - ‘It's all rubbish.’

- Annotations by 6-8 raters for each conversation
  - Full rating for valence, activation, power, expectation/anticipation, intensity
  - Optional rating for basic emotions: anger, happiness, sadness, fear, contempt…
- Solid SAL part : 87 conversations, each lasting approximately 5 minutes

Examples

- Valence score : -0.8819
- Valence score : 0.1125
- Valence score : 0.5803
- Valence score : 0.8308

Comparing Human Sentiment Labels to Automatic Labels

- Question:
  - Suppose we have unlabeled speech, can we annotate transcripts automatically with a sentiment annotation system and use those labels for unlabeled speech instead of manual labels?
- Method:
  - Segment transcripts into sentences and align with speech
  - Turn Semaine manual, continuous pos/neg labels into binary for use as gold standard
  - Label training transcript sentences using text-trained sentiment analyzer to label positive/negative/neutral
  - Build classifier from sentiment-labeled speech and compare to classifier built using manual Semaine speech labels

English Text-based Sentiment Analysis

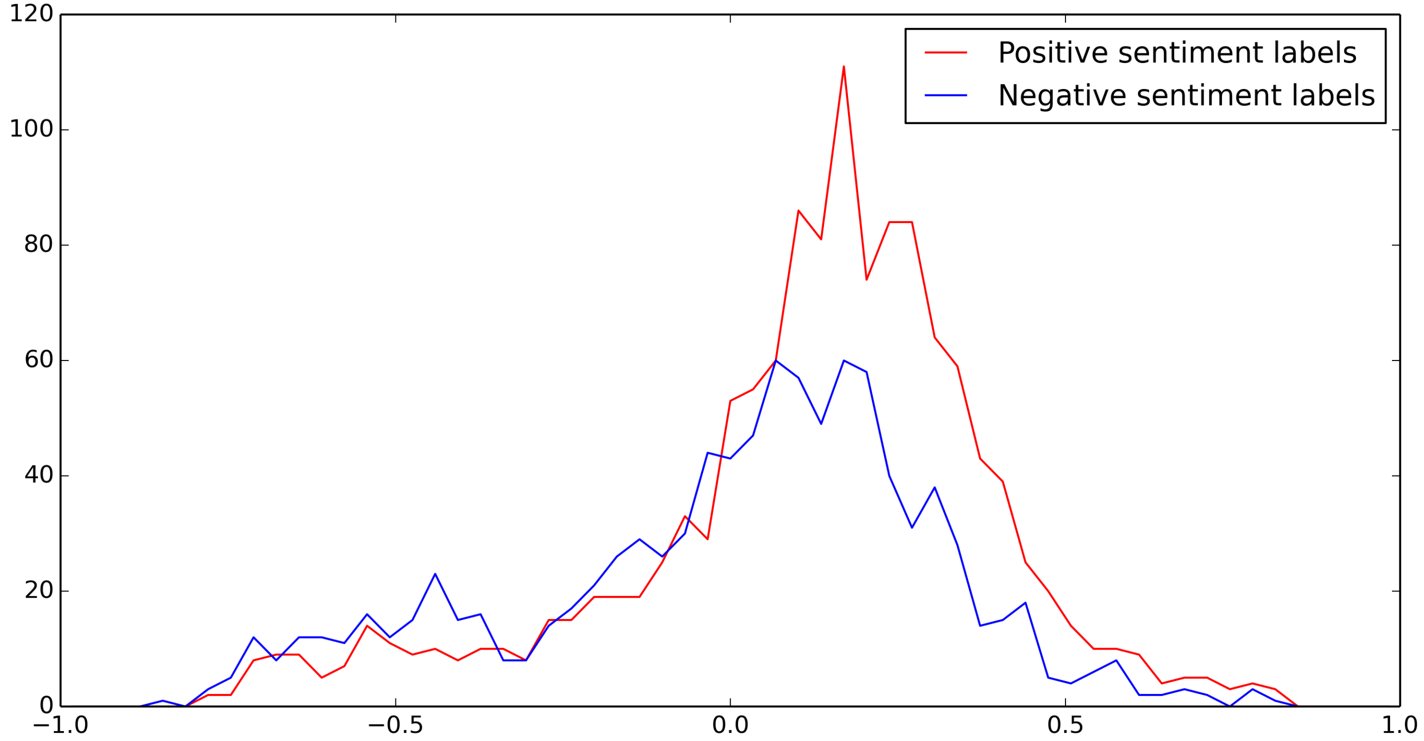
- Sentiment detection system (Rosenthal 2014)
- Features(lexical, syntactic):
  - Dictionary of Affect and Language (DAL)
  - WordNet 3.0
  - Wiktionary
  - POS tags
  - Top 500 n-gram features
- Output label: positive/negative/neutral

Comparison of Sentiment Labels vs. Valence Scores

- Examples:
  - Anyway he would probably do all the wrong shopping.
    - Sentiment analysis output label: **Negative**
    - Valence score: **- 0.4420**
  - There must be lot's of happy things in your life.
    - Sentiment analysis output label: **Positive**
    - Valence score: **0.7451**
  - *And how am I going to wrap all the presents?
    - Sentiment analysis output label: **Neutral**
    - Valence score: **- 0.4090**
  - *Life is very bad, I don't suppose yours is any better.
    - Sentiment analysis output label: **Positive**
    - Valence score: **- 0.7500**

Comparison of Sentiment Labels vs. Valence Scores

- Sentiment: Positive : 1301, Negative: 978, Neutral: 1177
- Distribution of sentiment labels over valence scores:



Results of Sentiment Analysis of Transcripts

- Manually annotated valence scores are unbalanced:
  - 2363 sentences with positive score(score >= 0)
  - 1093 sentences with negative score(score < 0)
- Set ‘neutral’ threshold to 0.118
  - 1728 sentences with positive/negative score
- Precision of sentiment labels using new threshold:
  - Positive label precision: 57.88%
  - Negative label precision: 60.22%

Experiments: Sentiment Labels vs. Valence Scores

- openSMILE baseline (384) feature set
- 4 speech experiments:
  - Train on sentiment labels; test on sentiment labels
  - *Train on sentiment labels; test on (human) valence scores
  - Train on (human) valence scores; test on sentiment labels
  - *Train on (human) valence scores; test on (human) valence scores
- 10-fold cross validation; weighted f-measure

Experiments: Sentiment Labels vs. Valence Scores

- Unbalanced classes in training data:
  - Moving threshold score for a balanced division
  - **Up sampling**
  - Down sampling
- Machine learning algorithms: (Scikit-learn)
  - Linear models: Linear regression; Ridge; Lasso
  - Nearest neighbors model: KNN
  - Tree model: Decision tree
  - Ensemble models: **Random forest**; Ada Boost
- Unbalanced classes in test data:
  - Evaluation: Weighted F-measure

Experiments:
Sentiment Labels vs. Valence Scores

- Baseline: Majority class (positive)

Train on	Sentiment Labels		SemaineValence Scores	
Test on	Sentiment Labels	Valence Scores	Sentiment Labels	Valence Scores
Baseline	0.4140	0.5526	0.4140	0.5526
Random Forest	0.5425	***0.6111***	0.4979	***0.6897***

- Should improve when we add lexical features to acoustic ones

Cross-Lingual Training

- Given a corpus of anger in English, can we predict anger in Mandarin?
- Given a corpus of anger in Mandarin, can we predict anger in English?
- Train on English Semaine, test on Mandarin Affect Corpus: F1=0.56 (cf. Mand/Mand 0.88) Train on Mandarin Affect, test on English Semaine: F1=0.62 (cf. Eng/Eng: F1=.77)

Conclusions and Future Work

- We can detect emotions like anger and stress from labeled Mandarin and English speech reasonably well
- We can detect emotions (e.g. anger) by training on one language and testing on another with performance above the baselines
- We can detect manually labeled English emotionsl speech from transcripts automatically labeled with sentiment, also with promising results
- Future: Appen Lorelei and Babel languages (Turkish, Mandarin,Uyghur)

– Develop text-based sentiment detectors cross-lingually for LRL
– Detect sentiment in Appen transcripts
– Label aligned speech
– Train sentiment models on ”labeled” speech
– Deep Learning

Thank you!

Questions?