

Humberto Pérez Espinosa
1 de Octubre de 2009

Análisis de los trabajos presentados en el INTERSPEECH 2009 Emotion Challenge

Emotion Challenge

- En reconocimiento de emociones, a diferencia de en reconocimiento automático de habla y de hablante, no existe corpora y condiciones de prueba estandarizados para comparar resultados bajo las mismas condiciones
- El objetivo es reducir la brecha entre el trabajo de investigación en reconocimiento de emociones a partir de voz y la baja compatibilidad entre los resultados obtenidos
- El corpus FAU Aibo sirve como base, con subconjuntos de entrenamiento y prueba bien definidos

Modalidades

- Open Performance Sub-Challenge
 - Los participantes proponen su características y algoritmo de clasificación
- Classifier Sub-Challenge
 - Los participantes usan el conjunto de características proporcionado por los organizadores y proponen su estrategia de selección y clasificación
- Feature Sub-Challenge
 - Los participantes envían su conjunto de características y la clasificación es hecha por los organizadores
- Además se puede participar clasificando 2 y/o 5 clases

Experimento

- La idea es combinar un corpus de habla infantil con habla emocional *natural* mediante un experimento de Mago de Oz|
- Se les pidió a los niños que le dieran instrucciones al robot como si estuvieran hablando con un amigo



Corpus

- Emociones espontaneas en Alemán. 9.2 horas de habla.
- Cinco personas entrenadas escucharon cada grabación para etiquetarlas con una de las siguientes clases: Enojado, Enfático (Negativas) Neutral, Positivo y Resto (Neutras)
- El corpus está etiquetado a nivel de frase. Se cuenta con audio y transcripciones
- Para otorgar una clase a una palabra se hizo una votación entre las opiniones de los etiquetadores. Si tres o mas coinciden se atribuye la etiqueta a la palabra

Ejemplos

- Neutro / Neutro:
 - Noch nach links Aibo, nach rechts Aibo geh nach rechts Aibo. 
 - A la izquierda no Aibo, gira a la derecha Aibo ve a la derecha Aibo.
- Enfático / Negativo:
 - Komm weiter. Weiter nach rechts Aibo nach rechts. 
 - Vamos. Continua Aibo derecho a la derecha.
- Enojado / Negativo:
 - Dreh dich. Aibo stopp. 
 - Date la vuelta. Aibo alto.
- Positivo / Neutro
 - Ganz brav 
 - Muy bien

Participantes

	Autores	Institución	Modalidad
1	Polzehl & Sundaram	Technische Universität Berlin. Germany	Open / 2 c
2	Barra & Fernández	Universidad Politécnica de Madrid. Spain	Open/2 y 5 c
3	Kockmann & Burget	Brno University of Technology Czech Rep.	Open/2 y 5 c
4	Dumouchel & Dehak	École de Technologie Supérieure. Canada	Open/2 y 5 c
5	Vogt & André	University of Augsburg. Germany	Open/2 y 5 c
6	Luengo & Navas	University of Basque Country. Spain	Open/2 y 5 c
7	Planet & Iriondo	Univesitat Ramon Llull. Spain	Clas-Feat/5c
8	Bozkurt & Erzin	Koc University. Turkey	Open/2 y 5 c
9	Lee & Mower	SAIL University of Southern California. USA	Open / 5 c

Flujo de Trabajo



Extracción de Características

Características

1	1500	Lingüísticas, Prosódicas, Espectrales / Cepstrales (MFCCs), Calidad de Voz
2	42	Prosódicas: $\log F_0$, $\log E$ y derivativas. Tracto Vocal: MFCCs, Δ MFCC, $\Delta\Delta$ MFCC
3	-	MFCCs, Δ MFCC, $\Delta\Delta$ MFCC, SDC, Prosodia en silabas
4	-	MFCCs, Formantes, Energy, F_0
5	1451	Prosódicas, Espectrales / Cepstrales (MFCCs), Calidad de Voz, Segmentos sonóros
6	56+S	Prosódicas, Regresión, Calidad de Voz, Final de oración, Espectrales (LFPC)
7	384	Las propuestas por los organizadores
8	-	Prosódicas, Cepstrales (MFCCs), Dinámicas Δ MFCC, $\Delta\Delta$ MFCC, HMM based
9	384	Las propuestas por los organizadores

Selección de Características

Selección

- 1 Se usó Information Gain seleccionando 320 de 1500 características
- 2 No se hizo selección de características
- 3 Se agruparon por tipo de característica y se probaron por separado
- 4 No se menciona
- 5 Cfs Subset Evaluation 465 de 1451
- 6 Forward Selection Ranking 10 de 52 prosódicas, más 32 espectrales
- 7 Minimal-redundancy-maximal-relevance. Se seleccionaron 100 características.
- 8 ninguna
- 9 ninguna

Balanceo de Clases

Balanceo

- 1 Se igualó el número de instancias de ambas clases dividiendo en dos algunas de las instancias etiquetadas como NEG
- 2 No se menciona
- 3 No se menciona
- 4 No se menciona
- 5 SMOTE
- 6 Different Missclassification Cost Value para compensar influencia
- 7 SMOTE
- 8 No se menciona
- 9 Ajuste de umbral de decisión basado en la distribución de clases

Clasificadores Empleados

Clasificadores

- 1 Multi Layer Perceptron, SVM
- 2 Graphical Models/Multi-Space Probability Distributions, Dinamic Bayesian Network
- 3 Gaussian Mixture Models- UBM, Maximum Likelihood, Maximum Mutual Information, Join Factor Analisis
- 4 Gaussian Mixture Models - UBM + SVM
- 5 Naïve Bayes
- 6 Gaussian Mixture Models- UBM (spectral), SVM(Prosody)
- 7 J48, IBk-A, Naïve Bayes, SVM
- 8 Gaussian Mixture Models- UBM
- 9 Bayesian Logistic Regresion, SVM

Estrategia de Clasificación

Estrategia

- 1 Se calcularon medidas de confianza normalizadas y ponderadas para ambos clasificadores(lingüístico y acústico). Se selecciona la salida con mayor confianza
- 2 En el entrenamiento se generan DBN para cada emoción, en la evaluación el sistema acumula las calificaciones para cada modelo y selecciona el más alto
- 3 Se probaron todas las fusiones de 2 clasificadores vs grupos de características y filtros
- 4 Un GMM por clase. Se aplicó Logistic Regresión para fusionar resultados
- 5 Ninguna
- 6 Un clasificador para características espectrales (GMM) y otro para prosódicas (SVM). Se fusionan con SVM. Todo se hace independientemente para cada clase.
- 7 Se hace un árbol de clasificadores de 2 niveles para cada clase. + / -, TP – FP / FN – TN
- 8 Un GMM por clase. Se fusionan usando una técnica de suma ponderada de decisiones
- 9 Árbol de clasificadores binarios. La clase de un clasificador es pasada al siguiente

Herramientas Empleadas

Herramientas

1	Janus / Ibis Toolkit
2	HTK, HTS, GMTK Toolkit
3	Multiclass Linear Regression Tool
4	No se mencionan
5	EmoVoice, Weka
6	No se mencionan
7	Matlab, Weka, Praat
8	No se mencionan
9	No se mencionan

Resultados

	Propuesta	2 Clases		5 clases	
		RP	RnP	RP	RnP
1	Fusionar los resultados de la clasificación hecha con atributos acústicos y la hecha con atributos lingüísticos	72.62	67.55	-	-
2	Dinamic Bayesian Networks para lidiar con la temporalidad de las características	-	67.06	-	38.24
3	Tomar en cuenta la evolución en el tiempo de los MFCCs y su fusión de clasificadores	65.80	68.30	43.90	41.30
4	Buen preprocesamiento y Optimización de GMM con SVM	68.68	70.29	52.08	39.40
5	Discretización de características, métodos rápidos	66.60	66.40	41.1	39.40
6	Separación de características para clasificar y fusión de clasificadores. Doble fusión, características y clases.	-	67.19	-	41.38
7	Clasificación mediante árboles de clasificadores	-	-	60.20	32.63
8	Tomar en cuenta la evolución en el tiempo de las características. Un clasificador por clase	63.03	67.90	44.17	41.59
9	Combinación de clasificadores binarios. Árbol de clasificación primero los más fáciles	-	-	-	41.57
RF	SVM, balanceo por SMOTE y estandarización acústica	65.50	67.70	39.20	38.20

Características Lingüísticas [1] Polzehl & Sundaram

- Se usan las transcripciones como características para la clasificación de contenido emocional
- Se diseñó un reconocedor automático de habla para obtener las transcripciones del conjunto de prueba. Se entrenó con adultos leyendo noticias
- Para clasificar una instancia se usaron “emotion salience words” calculadas en referencias o en hipótesis

Características Lingüísticas [1] Polzehl & Sundaram

- La salida es un puntaje para cada clase
- Para hacer una fusión de características se definen estadísticas en puntajes de “emotional word salience” dependientes de la clase
- Las palabras con mayor trascendencia emocional fueron “Aibolein”, “pfui”, “stopp”

Conclusiones

- No se pudo mejorar significativamente los resultados dados como referencia
- Es muy importante el preprocesamiento de la señal de audio
- Son necesarias características que tomen en cuenta el comportamiento a través del tiempo
- Los modelos dinámicos de clasificación GMM se está volviendo el estándar