Sistema de grabación para detectar hablantes de Buenos Aires y Córdoba

Fernando Bugni, Agustín Gravano & Miguel Martínez Soler
Universidad de Buenos Aires - Facultad de Ciencias Exactas y Naturales
fernando.bugni@gmail.com, gravano@dc.uba.ar, miguelmsoler@gmail.com

Abstract

El uso de la lengua siempre ha caracterizado a las personas que la utilizan. La forma como nos comunicamos no sólo posee la información del mensaje a transmitir, sino que también posee características del hablante. Nos enfocaremos en distinguir estas características entre las regiones de Córdoba y Buenos Aires. En el presente trabajo desarrollamos un sistema de grabación a través de Internet que nos permitió recolectar grabaciones de hablantes provenientes de ambos grupos. Diseñamos un experimento que nos ayudó a comparar el habla de cada grupo haciendo hincapié en sus diferencias. Extrayendo estos atributos, utilizamos algoritmos de Machine Learning para la clasificación de hablantes en los dos grupos.

Introduction

Las diferencias estudiadas entre los dos grupos son:

- **Regla 1: Los hablantes de Córdoba estiran la sílaba anterior a la acentuada mientras los de Buenos Aires no lo hacen.** Ejemplo: 'Espectacular' posee su sílaba acentuada en '-lar'. La sílaba anterior, o sea '-cu-' se alarga solamente para hablantes de Córdoba.
- **Regla 2: Los hablantes de Córdoba aspiran y elisionan la /s/ al finalizar una palabra. Esto no sucede en Buenos Aires.** Ejemplo: 'Pájaros' posee el fonema /s/ al final. Utilizando la dialéctica de Córdoba, la /s/ final sería más suave que una de Buenos Aires.
- **Regla 3: Para hablantes de Córdoba, la /s/ antes de la /c/ o /t/ suenan más suaves que para hablantes de Buenos Aires.** Ejemplo: 'Mosca' en la variante de Córdoba posee el fonema /s/ más suave que en Buenos Aires.
- **Regla 4: La 'c' antes de la 't' se pronuncia con menor frecuencia para hablantes de Córdoba que para hablantes de Buenos Aires.** Ejemplo: 'Doctor' no debe sonar el fonema /c/.
- **Regla 5: Para hablantes cordobeces la 'y' y 'll' se pasa a 'i'. No sucede esto para Buenos Aires.** Ejemplo: 'lluvia' se debe pronunciar utilizando el fonema /j/.
- **Regla 6: En hablantes cordobeces la /r/ no vibra mientras que en Buenos Aires pasa lo contrario.** Ejemplo: 'Espárrago' en su fonema /r/ debe ser suave en comparación con Buenos Aires.

Normalmente estas reglas se producen en el **habla espontánea** y raramente en habla leída.

Diseño del experimento

Definimos dos tipos de frases para realizar el experimento: **Frases comunes** y **Frases Amper**.

1. **Frases comunes**:
   Agregamos 30 frases populares con el objetivo de captar pronunciación espontánea. Con ellas vamos a cubrir las reglas 2 a 6. Ejemplo: **'En la pelea se conoce al soldado, sólo en la victoria se conoce al caballero'**. **'victoria'** cubre la regla 4 que nos propone medir la duración de la */c/* antes de la */t/*. **'caballero'** para la regla 5: el fonema */ll/* se pasa a */i/*
2. **Frases Amper**:
   Para captar sílabas acentuadas y su sílaba anterior a la acentuada se agregó frases con una estructura fija. El objetivo es cubrir la regla 1. **Esquema:** ***Sujeto+" salió "+Adjetivo***. Donde **Sujeto**: *"El canapé"*, *"El repollo"*, *"El espárrago"*; **Adjetivo**: *"espectacular"*, *"delicioso"*, *"riquísimo"*. Todas las combinaciones agregan 9 frases más.

Combinamos las frases: orden de frases en un experimento

Frases comúnes | Frases Amper

Intercalado: 1 ó 3 Frases comunes cada una Amper

Datos Obtenidos

	Bs.As.	Cba.	Total
Conservar	220	90	310
Problemas en el habla	33	15	48
Mucho ruido de fondo	2	12	14
Sonido saturado	2	0	2

Table 1: Grabaciones y su clasificación

Utilizando nuestra página web de grabación clasificamos las grabaciones recolectadas en la **tabla 1**. La página de grabación también soporta varios intentos para una misma grabación. Descartamos estos casas quedándonos con el mejor caso. Esto nos da un total de **260 grabaciones, 181 para Buenos Aires y 79 para Córdoba.** Seguiremos con estos datos para el posterior análisis.

Extracción de información

Etiquetamos cada audio para marcar dónde comienza y termina cada fonema. Para ello utilizamos una librería llamada **ProsodyLab-Aligner**. Un ejemplo extrayendo el atributo 'kt' es el siguiente:

"en la pelea se konose al soldaDo
solo en la biktorja se konose al kaBaZero"

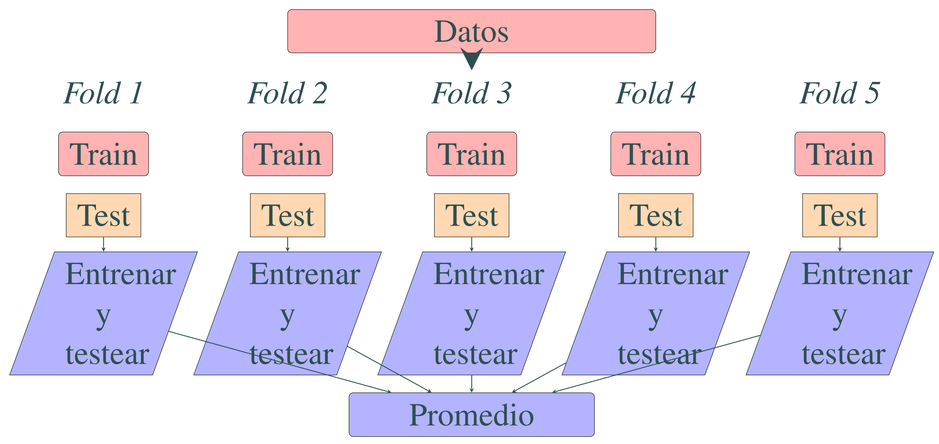
l a *sil* b i k t o r j a *sil* s e *sil*

La métrica utilizada es: $\frac{X-\mu}{\sigma}$ donde X es el valor a normalizar (por ej.: la duración de un fonema o sílaba dado). μ es el promedio de duración de la unidad utilizada en la grabación. σ es el desvío estándar de esta unidad utilizada. Esta métrica la realizamos para cada una de las reglas.

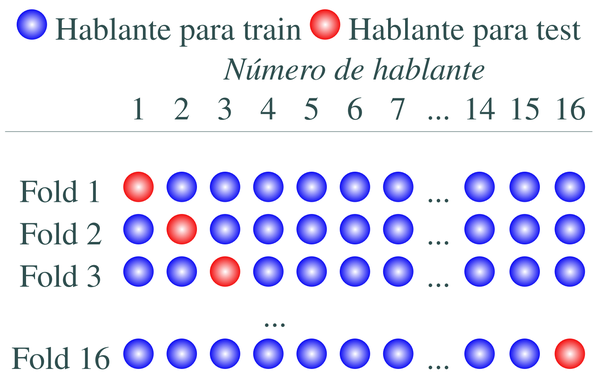
Análisis de datos y clasificación

Realizamos tres tipos de cross-validations para realizar la clasificación.

- **CV1 - Grupos de hablantes:** dividimos los hablantes en dos grupos en train y test. Esto lo hacemos para 5 folds.



- **CV2 - Dejando un hablante fuera promediando los atributos:** En este esquema dejamos un hablante afuera para test y entrenamos con el resto. También equilibramos la cantidad de hablantes: 8 Buenos Aires, 8 Córdoba. Para cada hablante promediamos sus atributos.



- **CV3 - Dejando un hablante fuera promediando los atributos desconocidos**: No es necesario promediar todos los atributos de cada hablante. Podemos promediar sólo los atributos desconocidos. El esquema de folds es igual al anterior.

Resultados

En la **tabla 2** podemos ver el resultado del promedio de cada fold según cada clasificador.

	Zero Rule	Ripper	C4.5	F. SMO	NaiveBayes
CV1	64	61	60	72	70
CV2	53.33	60	60	93.33	80
CV3	50	72.44	73.48	77.19	74.62

Table 2: Porcentaje de clasificación

Podemos ver cómo **Function SMO** y **NaiveBayes** pudieron mejorar la performance de clasificación utilizando los atributos que definimos. Utilizando Wilcoxon y T de Student pudimos confirmar que estos porcentajes son estadísticamente significativos.

References

[1] Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program] version 5.3.51, retrieved 2 june 2013 from http://www.praat.org/. 2013.

[2] Elena Vidal de Battini. Español en la argentina. 1964.

[3] Kyle Gorman, Jonathan Howell, and Michael Wagner. Prosodylab-aligner: A tool for forced alignment of laboratory speech. canadian acoustics. 39.3. 192–193. In *Proceedings of Acoustics Week in Canada, Quebec City*, 2011.

[4] Jorge A. Gurlekian, Reina Yanagida, Mónica Noemí Trípodi, and Guillermo Toledo. Amper-argentina: Variabilidad rÍtmica en dos corpus. 2009.

[5] María Beatriz Fontanella de Weinberg. El español en la argentina y sus variedades regionales. 2000.