

DEEP CONTENT-BASED MUSIC RECOMMENDATION

A. van der Oord, S. Dieleman & B. Schrauwen, 2013, Ghent University



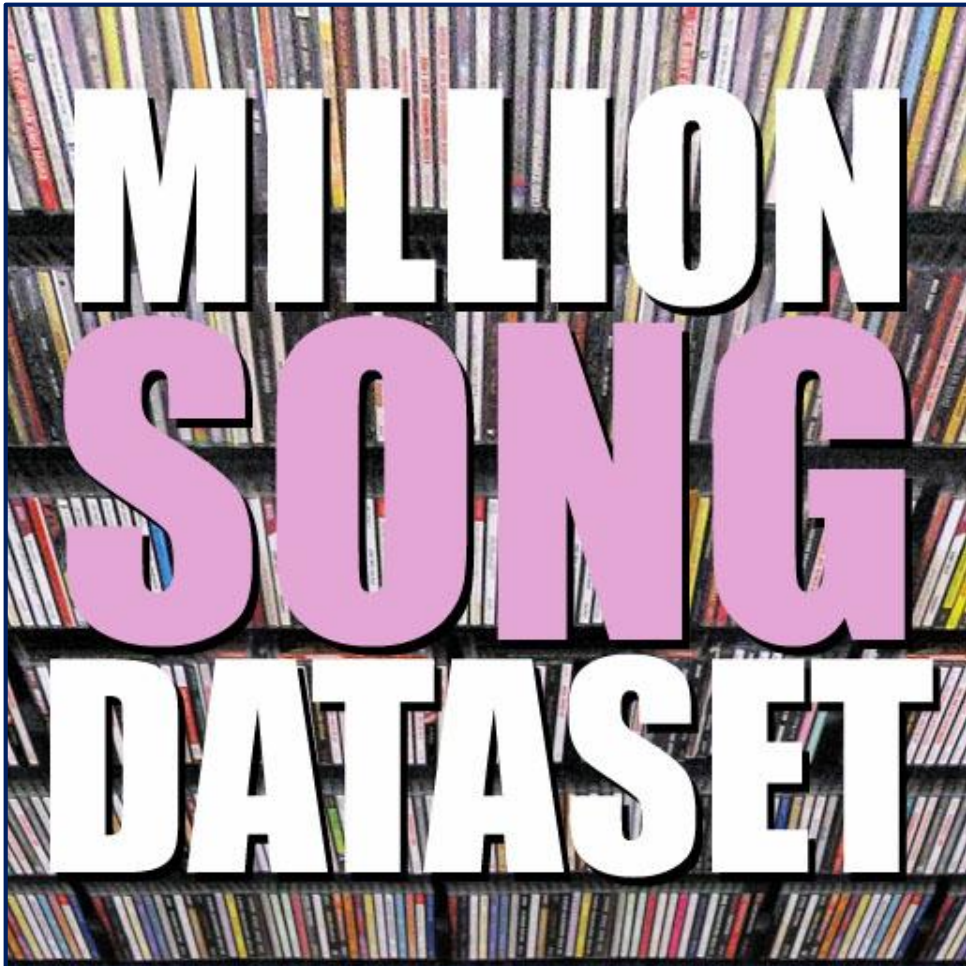
PEDRO BAHAMONDES W.
IIC3633 – SISTEMAS RECOMENDADORES
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
11 DE OCTUBRE DE 2018

INTRODUCCIÓN

- Recomendación automática de música
 - ✓ Descubrir música nueva
 - ✓ Acertar en el gusto del usuario
- Dificultades del problema
 - Variedad de estilos y géneros
 - Gustos influenciados por factores geográficos y culturales
 - Gran número de ítems (canciones)
 - Álbumes y artistas no homogéneos
- Collaborative Filtering vs. Content-based
 - ❖ CF genera mejores predicciones que CB en general
 - ❖ Problema de *cold start* común en música (*power law*)
 - ❖ CB todavía sujeto a enorme brecha semántica
- Este trabajo busca reducir la brecha semántica existente usando redes neuronales convolucionales profundas



DATASET



- *The Million Song Dataset*. Contiene 1M de canciones con **metadata** y *features* precalculadas de los audios.
- *Echo Nest Taste Profile Subset*. Subconjunto de 382K canciones para 1M de usuarios.
 - Utilizado en *The Million Song Dataset Challenge*.
- Extracción de muestras de 29s para ~99% de las canciones desde 7digital.com.
- Datos contienen *implicit feedback*, como el número de veces que una canción ha sido escuchada por un usuario.

WEIGHTED MATRIX FACTORIZATION

- Representación por **factores latentes** para usuarios de la base de datos especializado para *implicit feedback*
- Número de reproducciones r_{ui} , preferencia p_{ui} , confianza c_{ui}

$$p_{ui} = I(r_{ui} > 0)$$

$$c_{ui} = 1 + \alpha \log(1 + \epsilon^{-1} r_{ui})$$

- Función objetivo WMF:

$$\min_{x_*, y_*} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

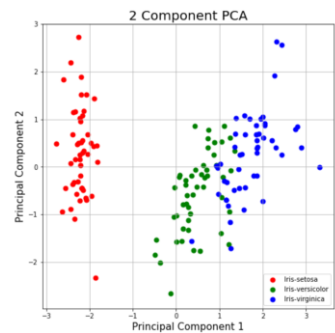
Regularizador L2 Factor latente ítem i

Factor latente usuario u

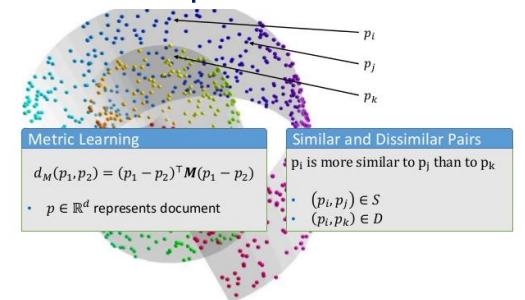
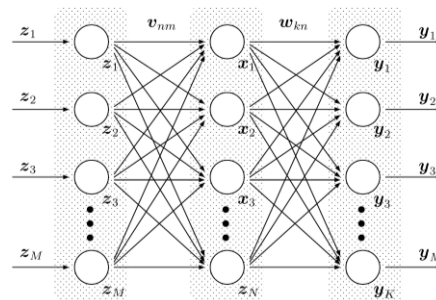
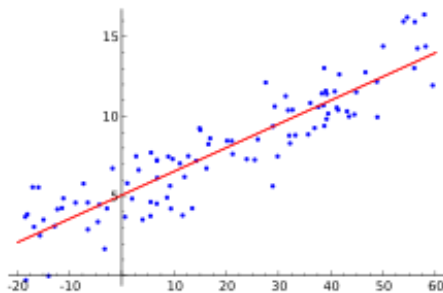
PREDICCIÓN DE FACTORES LATENTES

- Representación por *Bag-Of-Words* (BoW)
 - Los *Mel-frequency cepstral coefficients* (MFCC) son extraídos desde las señales de audio:
 - 39 coeficientes por señal
 - Se **vectorizan** los MFCC:
 - diccionarios de 4000 elementos con *clustering* de algoritmo de K-medias.
 - Se representan los vectores como **bag-of-words**:
 - para cada canción, se mide cuántas veces concurre con cada media.

- Principal Component Analysis (PCA) para **reducir la dimensionalidad** de la representación



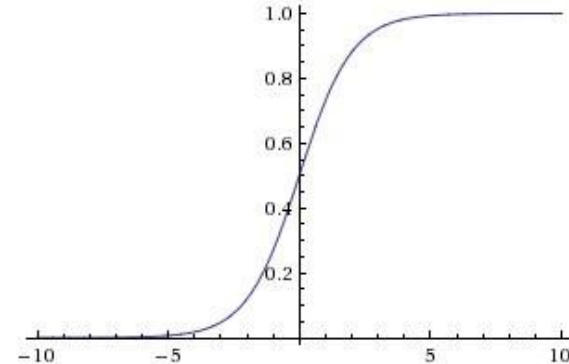
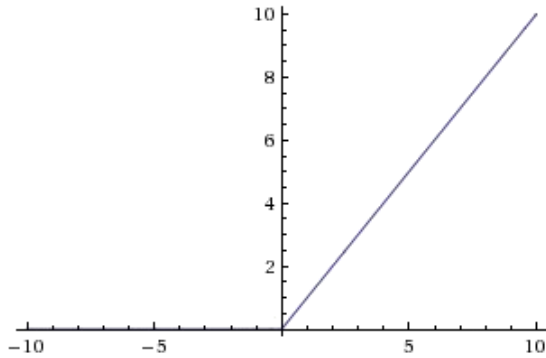
- Sobre BoW, se aplica una **regresión lineal**, una red neuronal tipo **perceptrón multicapa** con 1000 unidades ocultas y una *metric learning to rank* como base de comparación para los experimentos.



PREDICCIÓN DE FACTORES LATENTES

➤ Uso de Red Neuronal Convolutiva (CNN)

- Convergencia acelerada usando *Rectified Linear Units* en lugar de *Sigmoid Nonlinearities*



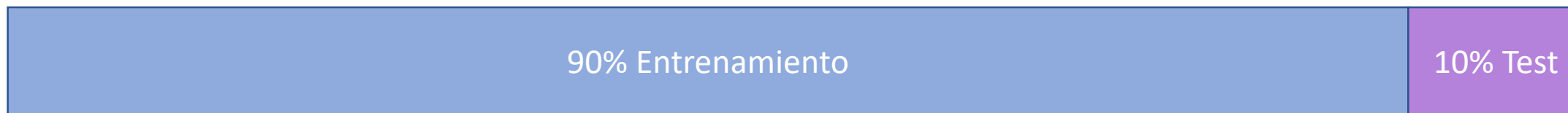
- Paralelización en GPU para rapidez en entrenamiento
- Representación sobre *features* intermedios calculados sobre extractos de 3 segundos
- Funciones objetivo minimizan *Mean Square Error* (MSE) y *Weighted Prediction Error* (WPE)

$$\min_{\Theta} \sum_i \|y_i - y'_i\|^2$$

$$\min_{\Theta} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2$$

EVALUACIÓN

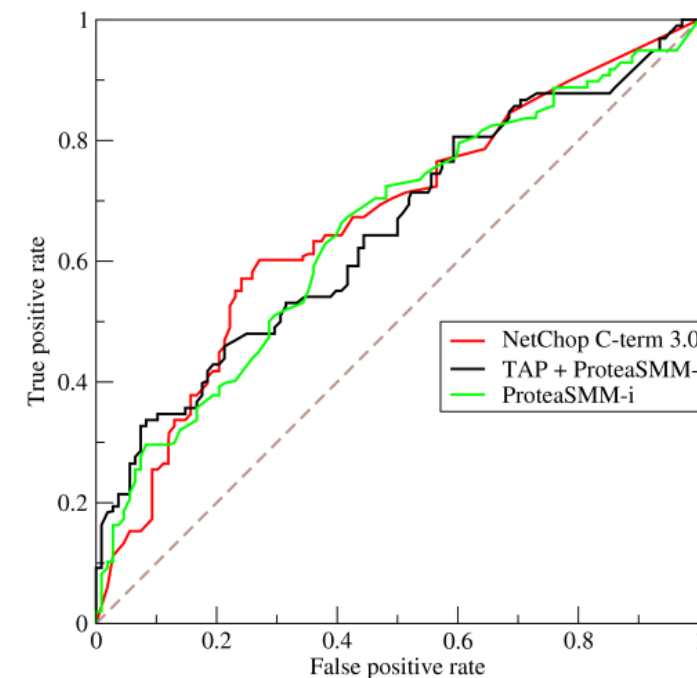
➤ *10-fold cross validation*



- Recomendaciones con **mayor puntaje primero** para rankear
- Evaluación cuantitativa con *Mean Average Precision* (mAP) y con **área bajo la curva ROC** (AUC)

$$mAP = \frac{\sum_{q=1}^Q AveP(q)}{Q}$$

- **Evaluación cualitativa** observando y categorizando recomendaciones.



EVALUACIÓN

- Primera evaluación sobre subconjunto de 9.330 canciones más populares sobre 20.000 usuarios

Modelo	Mean Average Precision (mAP)	Área bajo la curva ROC (AUC)
Metric Learning to Rank (MLR)	0,01801	0,60608
Regresión Lineal	0,02389	0,63518
Perceptrón Multicapa (MLP)	0,02536	0,64611
CNN con MSE	0,05016	0,70987
CNN con WPE	0,04323	0,70101

- Segunda evaluación sobre las 382.410 canciones para 1 millón de usuarios

Modelo	Mean Average Precision (mAP)	Área bajo la curva ROC (AUC)
Aleatorio	0,00015	0,49935
Regresión Lineal	0,00101	0,64522
CNN con MSE	0,00672	0,77192
Cota superior (WMF)	0,23278	0,96070

EVALUACIÓN

➤ Evaluación cualitativa mirando las **canciones más similares** según WMF y CNN

Query	Most similar tracks (WMF)	Most similar tracks (predicted)
Jonas Brothers - Hold On	Jonas Brothers - Games Miley Cyrus - G.N.O. (Girl's Night Out) Miley Cyrus - Girls Just Wanna Have Fun Jonas Brothers - Year 3000 Jonas Brothers - BB Good	Jonas Brothers - Video Girl Jonas Brothers - Games New Found Glory - My Friends Over You My Chemical Romance - Thank You For The Venom My Chemical Romance - Teenagers
Beyoncé - Speechless	Beyoncé - Gift From Virgo Beyoncé - Daddy Rihanna / J-Status - Crazy Little Thing Called Love Beyoncé - Dangerously In Love Rihanna - Haunted	Daniel Bedingfield - If You're Not The One Rihanna - Haunted Alejandro Sanz - Siempre Es De Noche Madonna - Miles Away Lil Wayne / Shanell - American Star
Coldplay - I Ran Away	Coldplay - Careful Where You Stand Coldplay - The Goldrush Coldplay - X & Y Coldplay - Square One Jonas Brothers - BB Good	Arcade Fire - Keep The Car Running M83 - You Appearing Angus & Julia Stone - Hollywood Bon Iver - Creature Fear Coldplay - The Goldrush
Daft Punk - Rock'n Roll	Daft Punk - Short Circuit Daft Punk - Nightvision Daft Punk - Too Long (Gonzales Version) Daft Punk - Aerodynamite Daft Punk - One More Time / Aerodynamic	Boys Noize - Shine Shine Boys Noize - Lava Lava Flying Lotus - Pet Monster Shotgun LCD Soundsystem - One Touch Justice - One Minute To Midnight

Table 4: A few songs and their closest matches in terms of usage patterns, using latent factors obtained with WMF and using latent factors predicted by a convolutional neural network.

EVALUACIÓN

- Visualización usando técnica de *T-distributed Stochastic Neighbor Embedding* (t-SNE)

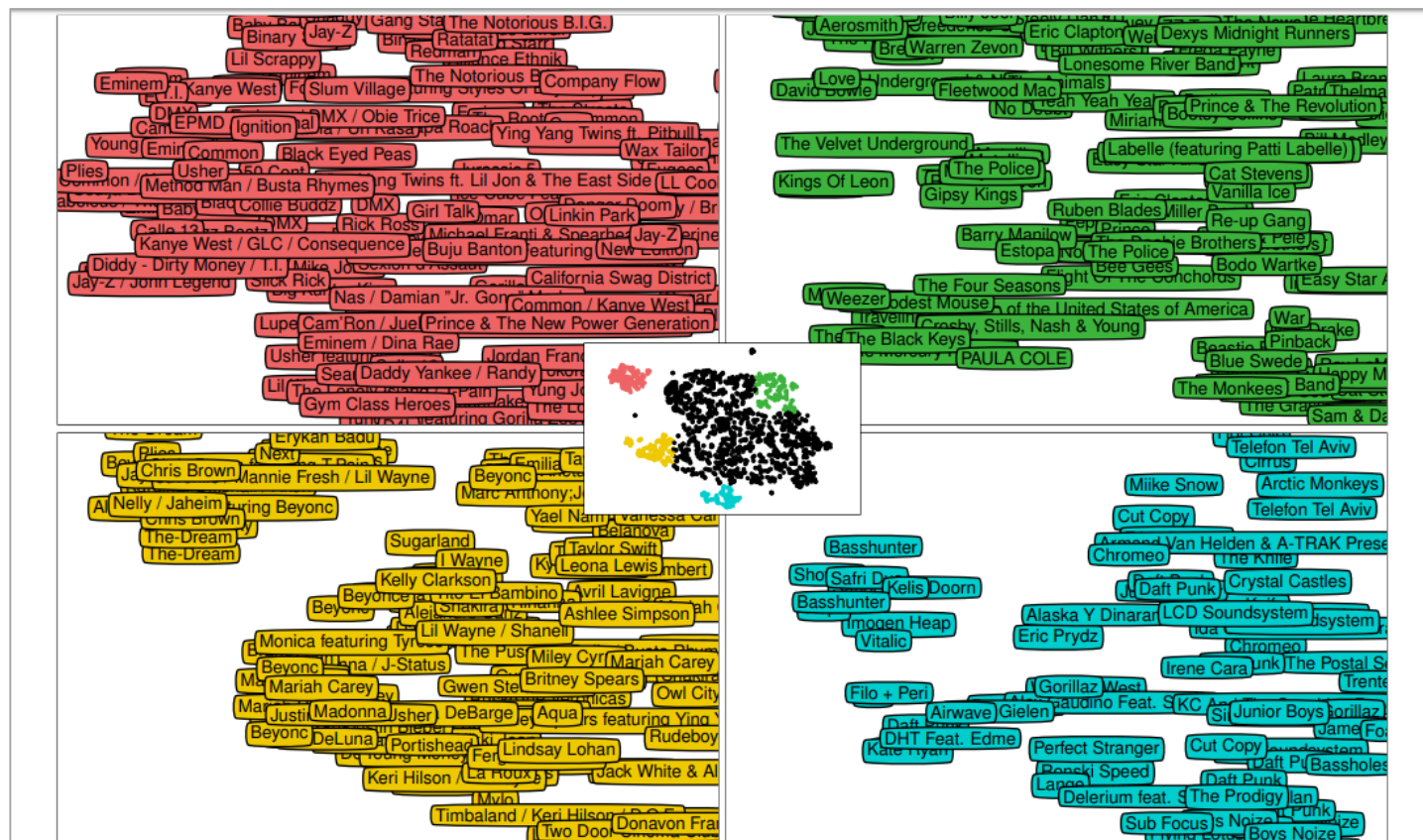


Figure 1: t-SNE visualization of the distribution of predicted usage patterns, using latent factors predicted from audio. A few close-ups show artists whose songs are projected in specific areas. We can discern hip-hop (red), rock (green), pop (yellow) and electronic music (blue). This figure is best viewed in color.

TRABAJOS RELACIONADOS

- Recomendación de artículos científicos con matriz de factorización probabilística extendido con modelo que considera tópicos (Wang et al., 2011).
- Variación de *Metric Learning to Rank* en el que se cambian los *features* extraídos de los audios para obtener mejores resultados (McFee et al., 2012).
- Recomendación de un ítem utilizando un *query* con información semántica, sin ganancias significativas respecto a CF (Weston et al., 2012).

* Este trabajo utiliza funciones de pérdida distintas en su evaluación, por temas de escalabilidad.

SÍNTESIS Y CONCLUSIONES

- ✓ En el trabajo de los autores, se investiga el uso de redes neuronales convolucionales profundas para predecir los factores latentes de audios de música.
- ✓ Trabajo sobre datos a escala industrial.
- ✓ Recomendaciones sensibles a los gustos de los usuarios a pesar de limitaciones de representación.
- ✓ Predecir factores latentes parece ser viable para recomendar música nueva e impopular.
- ✓ CNN aplicadas superan significativamente a las representaciones con BoW.

DEEP CONTENT-BASED MUSIC RECOMMENDATION

A. van der Oord, S. Dieleman & B. Schrauwen, 2013, Ghent University



PEDRO BAHAMONDES W.
IIC3633 – SISTEMAS RECOMENDADORES
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
11 DE OCTUBRE DE 2018