Cold-Start Music Recommendation Using Multimodal Deep Architectures

ORIOL NIETO

ONIETO@PANDORA.COM

SYSTEMATIC APPROACHES TO DEEP LEARNING METHODS FOR AUDIO

ESI WORKSHOP

VIENNA, AUSTRIA

SEP 15, 2017

Outline

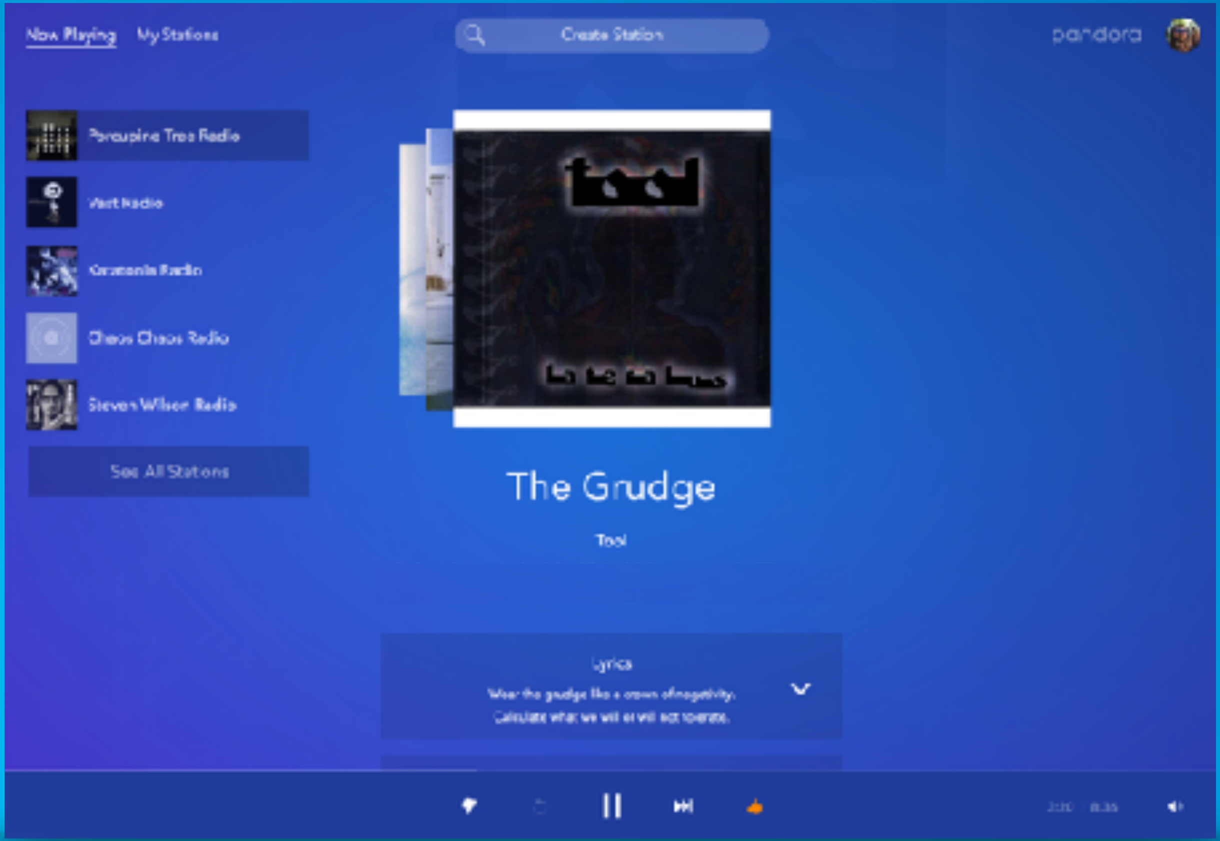
- Motivation: The Cold-Start Problem

- Background: Collaborative Filtering

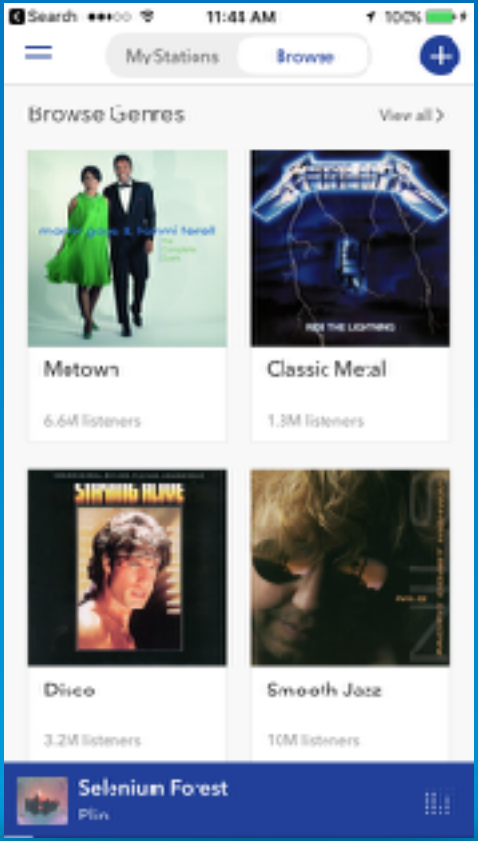
- Cold-Start Music Recommendation:
    - Estimate Collaborative Factors from Audio
    - The Music Genome Project™
    - Multimodal Deep Architectures

Outline

- **Motivation: The Cold-Start Problem**
- Background: Collaborative Filtering
- Cold-Start Music Recommendation:
  - Estimate Collaborative Factors from Audio
  - The Music Genome Project™
  - Multimodal Deep Architectures

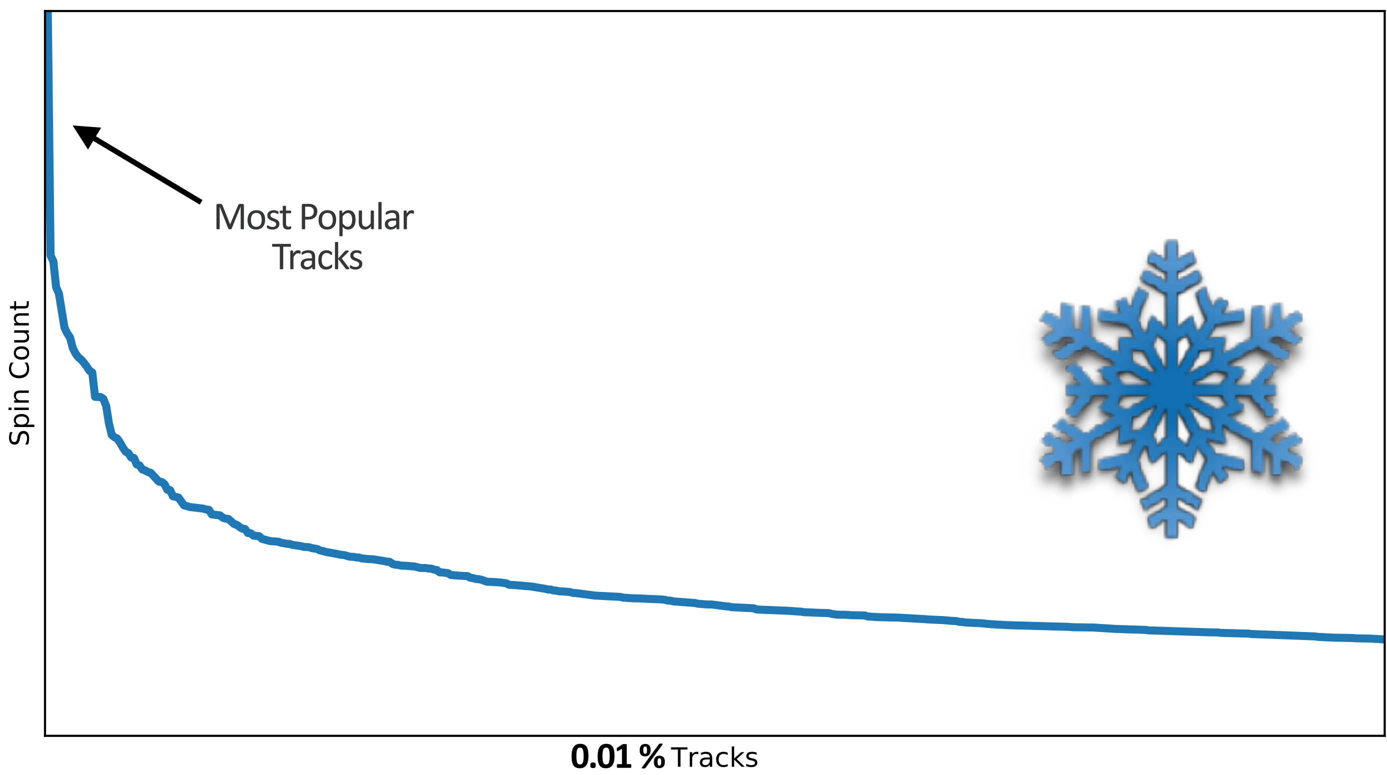


Wear the grudge like a crown of negativity.
Calculate what we will or will not tolerate.



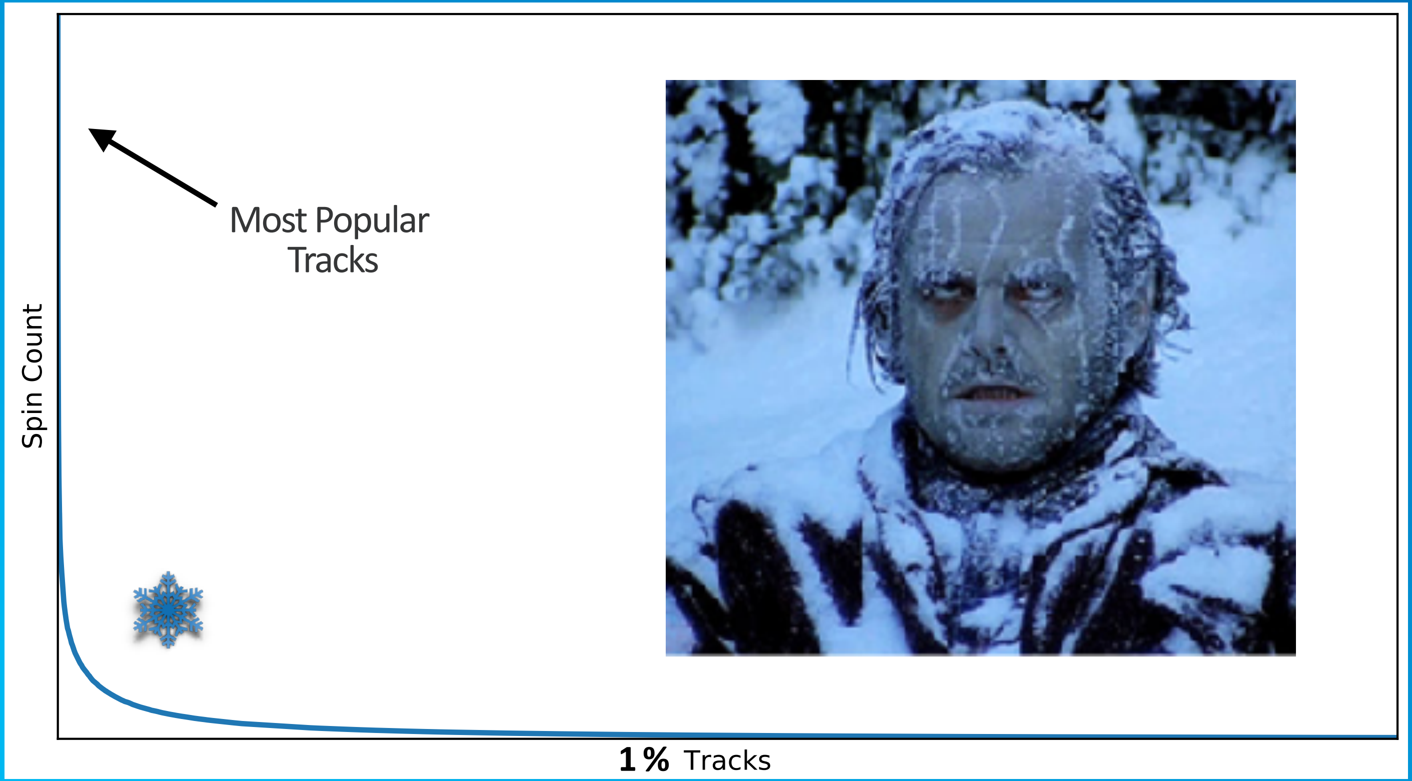
Cold-Start Problem

THE LONG TAIL



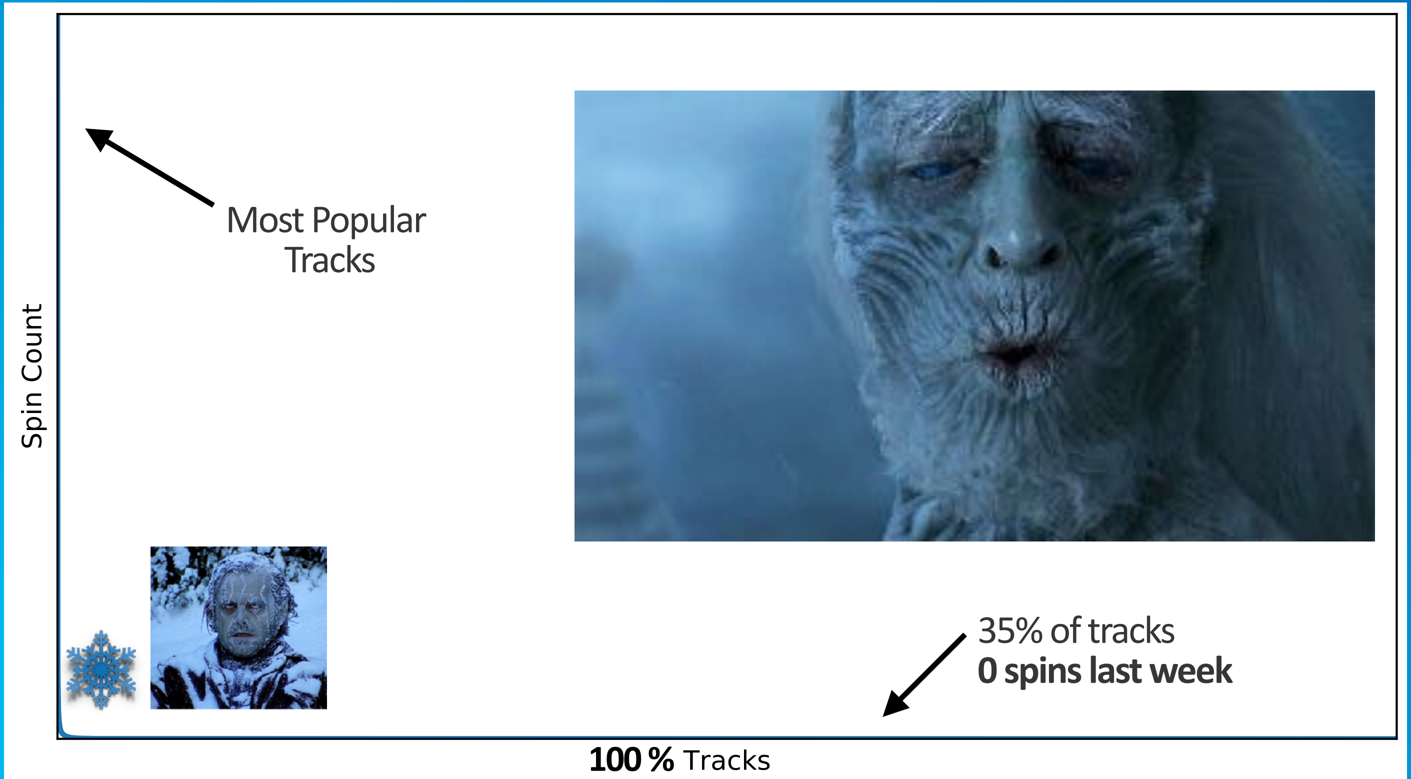
Cold-Start Problem

THE LONG TAIL



Cold-Start Problem

THE LONG TAIL



Most Popular Tracks

35% of tracks
0 spins last week

Outline

- Motivation: The Cold-Start Problem
- **Background: Collaborative Filtering**
- Cold-Start Music Recommendation:
    - Estimate Collaborative Factors from Audio
    - The Music Genome Project™
    - Multimodal Deep Architectures

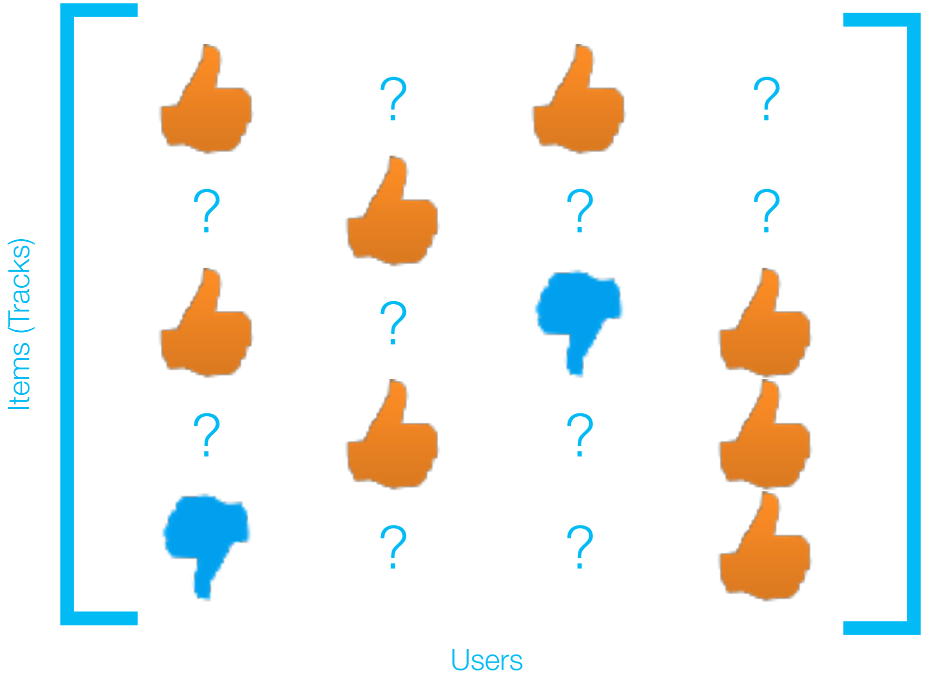
Collaborative Filtering

PROBLEM OVERVIEW

Explicit

Thumbs (up and down)

Station Creation



👍	?	👍	?
?	👍	?	?
👍	?	👎	👍
?	👍	?	👍
👎	?	?	👍

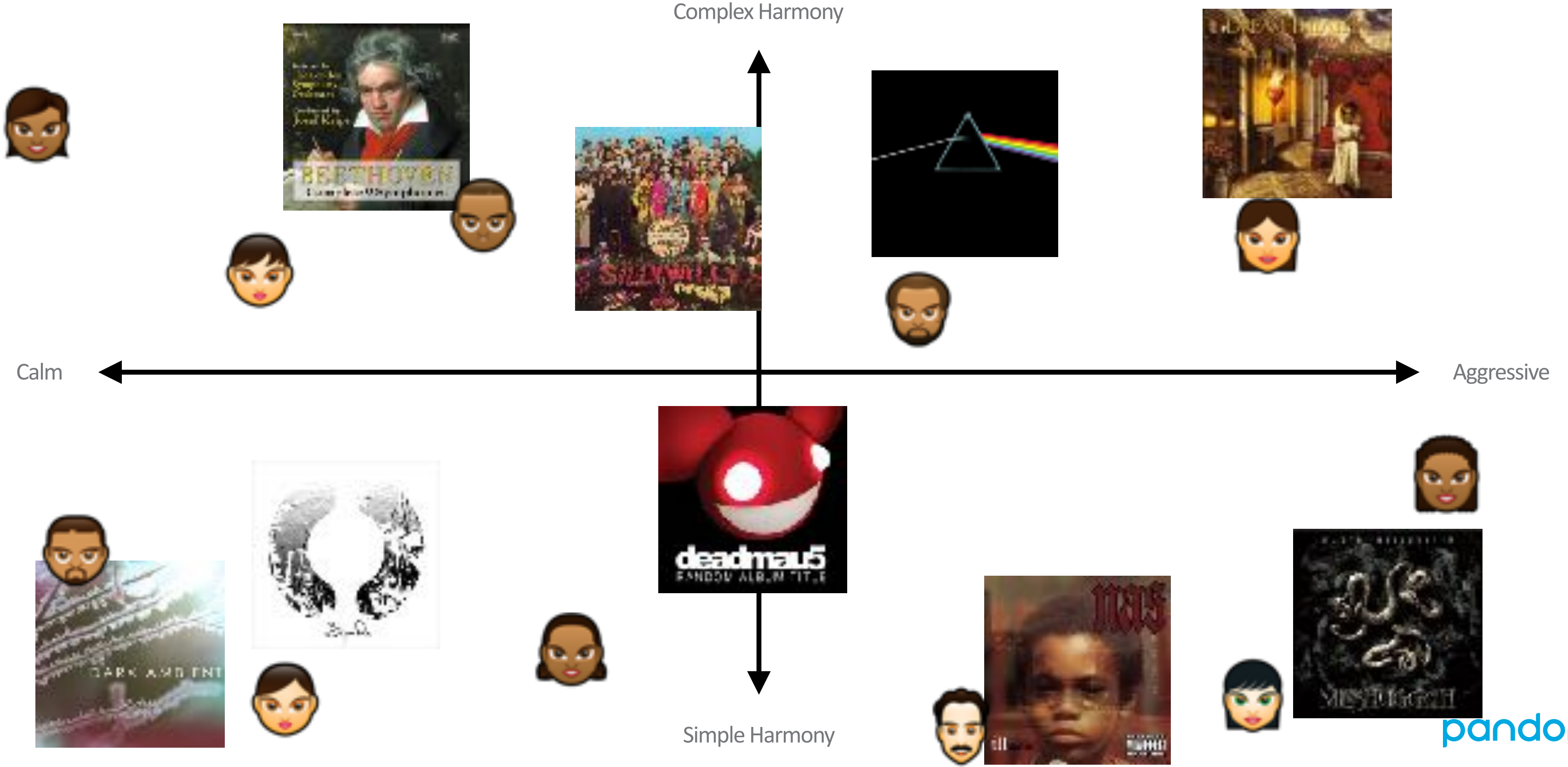
Implicit

Track Completion

Track Skips

Collaborative Filtering

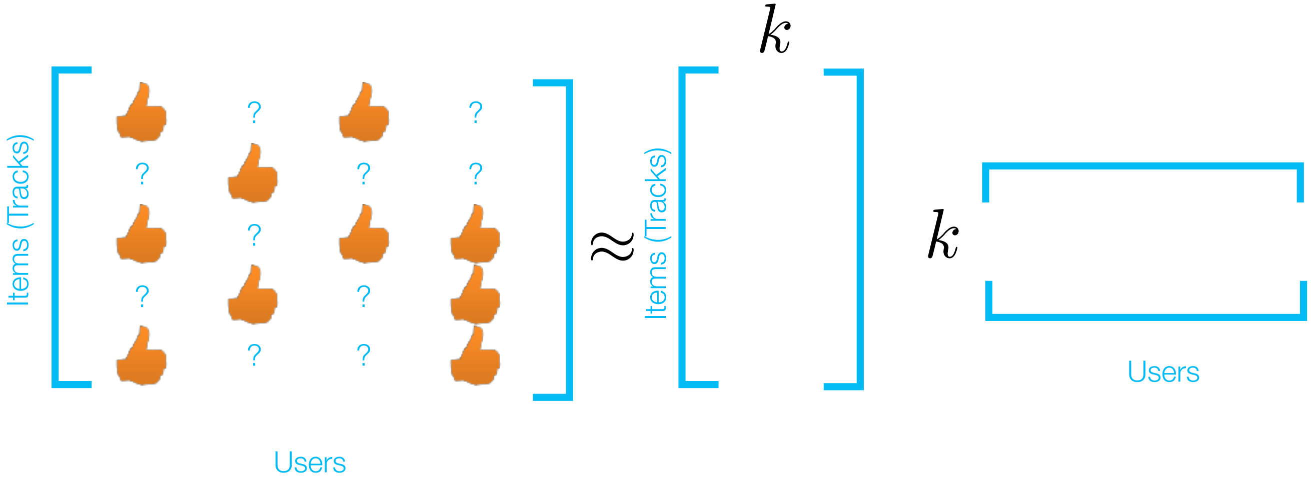
LATENT FACTORS



pandora®

Collaborative Filtering

MATRIX FACTORIZATION



Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 42–49.

Collaborative Filtering

PROBLEM FORMULATION

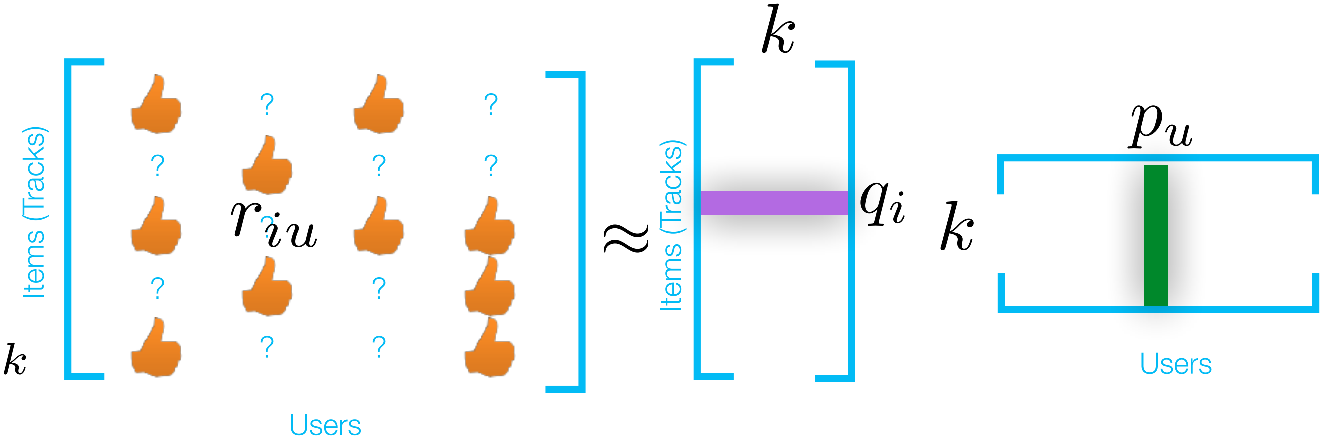
Given Item ***i*** and User ***u***:

Rating: r_{iu}

Item Latent Factor: $q_i \in \mathbb{R}^k$

User Latent Factor: $p_u \in \mathbb{R}^k$

Rating Approximation: $\hat{r}_{iu} = q_i^T p_u$



$$\text{argmin}_{q*,p*} \sum_{u,i \in \mathcal{S}} (r_{ui} - q_i^T p_u)^2 \quad + \lambda(||q_i||^2 + ||p_u||^2)$$

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer, 42(8), 42–49.

Collaborative Filtering

EXAMPLE

	Artist	Title
Query Track	**The Beatles**	**While My Guitar Gently Weeps**
Ranked 1	The Beatles	A Day In The Life
Ranked 2	The Beatles	A Day In The Life (Love Version)
Ranked 3	The Beatles	Across The Universe

Collaborative Filtering

EXAMPLE

	Artist	Title
Query Track	**The Beatles**	**While My Guitar Gently Weeps**
Ranked 35	George Harrison	While My Guitar Gently Weeps (Live)
Ranked 82	George Harrison	My Sweet Lord (Live)
Ranked 91	Paul McCartney & Eric Clapton	Something (Live)
Ranked 158	Led Zeppelin	Tangerine

Collaborative Filtering

THE GOOD AND THE BAD

Rich preference-driven similarity space

Powerful at matching the right song with the right listener

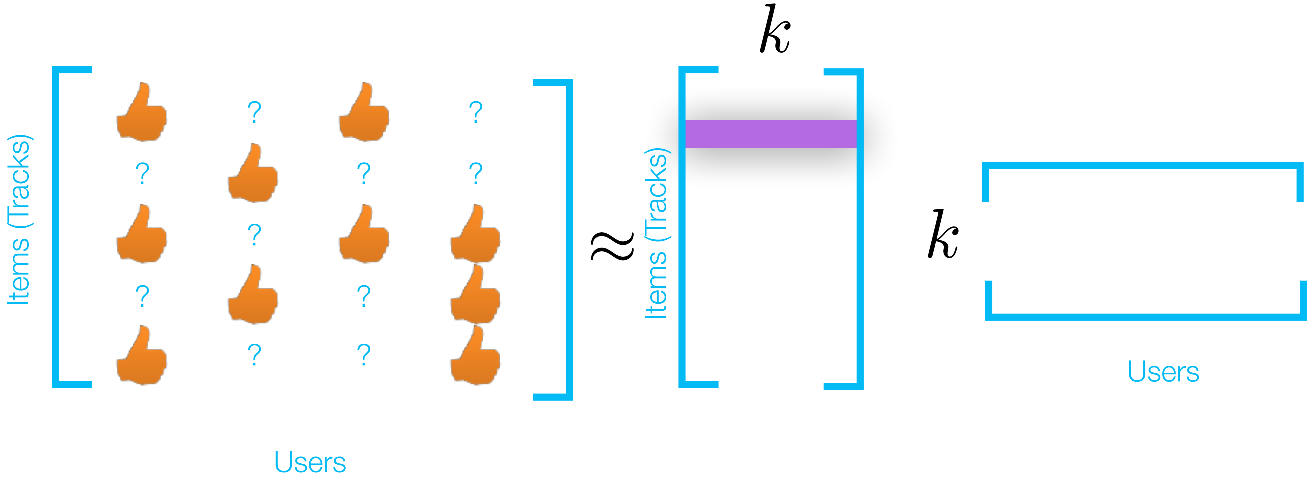
Latent space is generally not interpretable

Can only recommend items that have already been rated

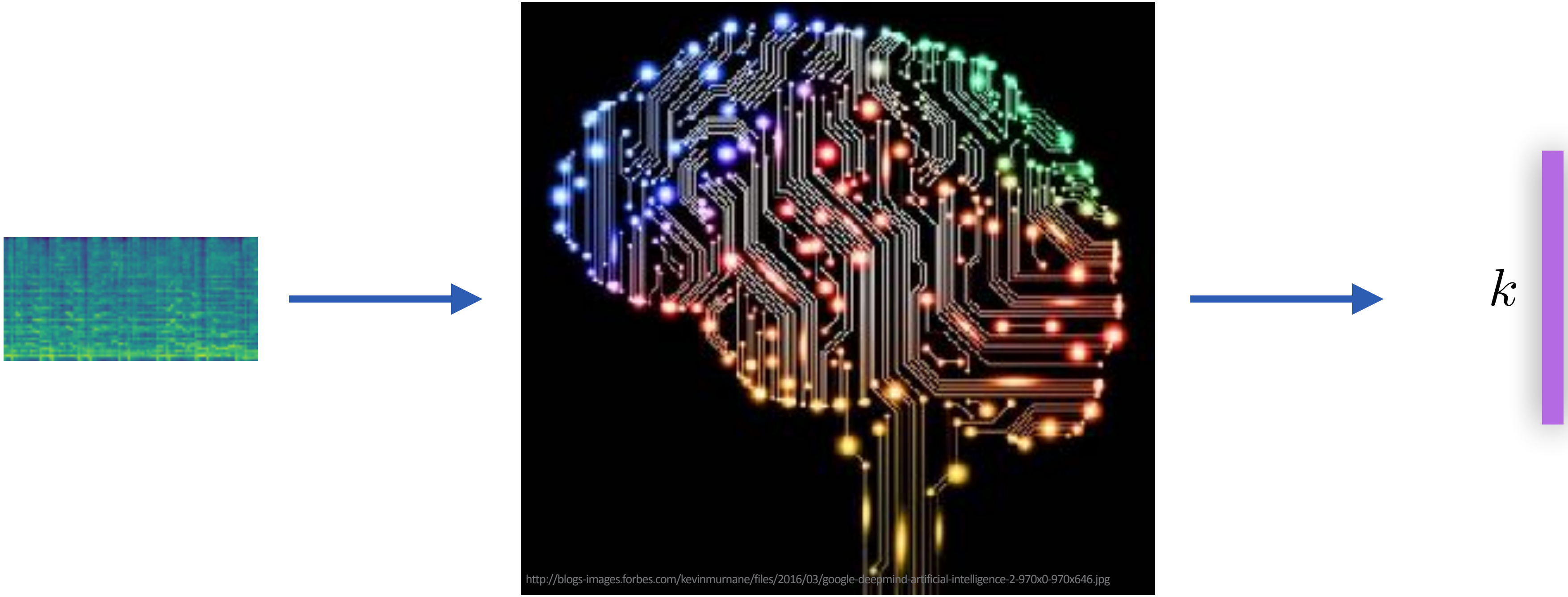
Outline

- Motivation: The Cold-Start Problem
- Background: Collaborative Filtering
- **Cold-Start Music Recommendation:**
  - **Estimate Collaborative Factors from Audio**
  - The Music Genome Project™
  - Multimodal Deep Architectures

Estimate Collaborative Factors



Approximate Item Factors using Audio

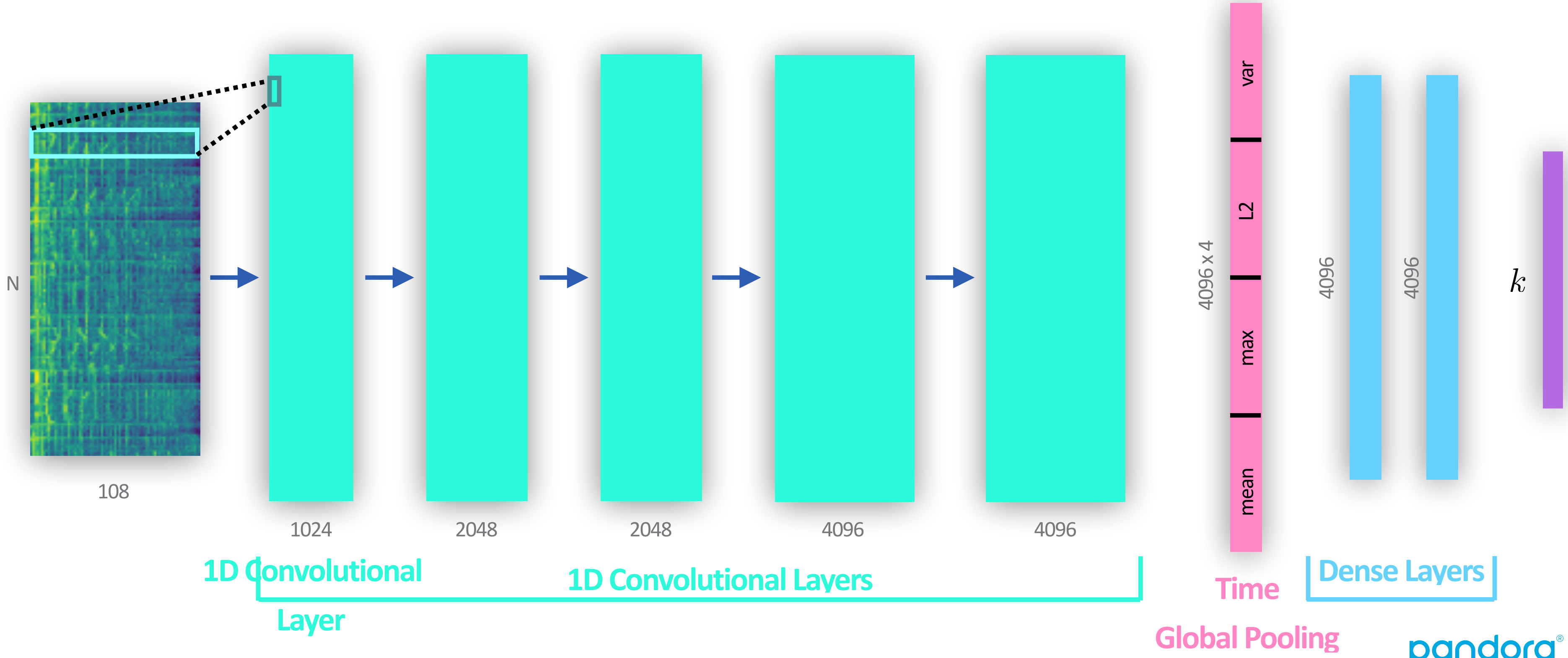


http://blogs-images.forbes.com/kevinmurnane/files/2016/03/google-deepmind-artificial-intelligence-2-970x0-970x646.jpg

Oord, A. Van Den, Dieleman, S., & Schrauwen, B. (2013). Deep Content-based Music Recommendation. Advances in Neural Information Processing Systems, 2643–2651.

Approximating Factors using Audio

WITH DEEP LEARNING

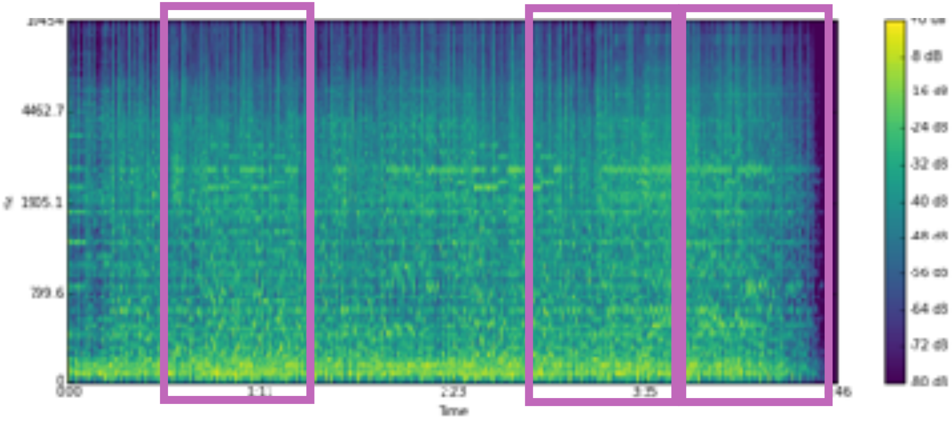


pandora®

Approximating Item Factors using Audio

TRAINING DATA

- (Small) Data set $\{\mathbb{X}, \mathbb{Y}\}$:
  - 83k tracks
  - 3 patches of 35 seconds per track (251k patches = M)
    - (Patches only for training!)
  - Splits:
    - Train: 80%
    - Validation: 10%
    - Test: 10%



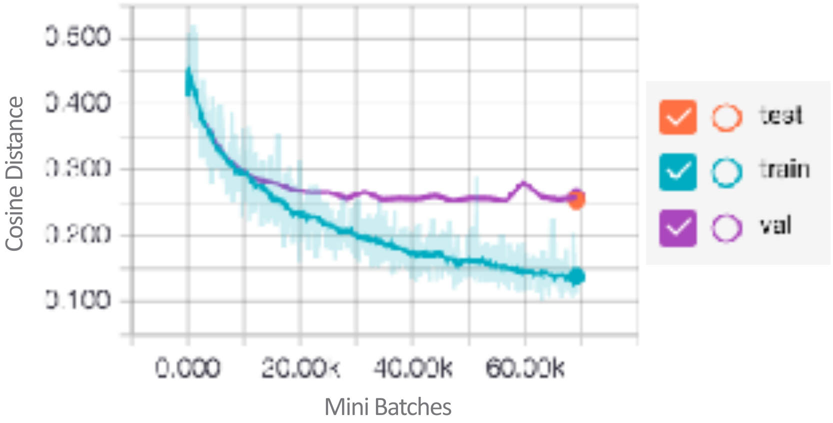
Approximating Item Factors

TRAINING

- Loss function:
  - Cosine Distance

$$\mathcal{L}(\theta) = 1 - \frac{1}{M} \sum_{X\in\mathbb{X},\mathbf{y}\in\mathbb{Y}} \frac{f(X;\theta)^T\mathbf{y}}{||f(X;\theta)||_2||\mathbf{y}||_2}$$

- Optimization:
  - Adam (default params)
  - 50% Dropout on Dense Layers
  - Early Stopping
  - Mini-batches of 64 examples



Approximating Item Factors using Audio

RESULTS

Input	Cos Distance	# Epochs	Time / Epoch
Audio (35s Patches)	0.25	22	~2h

Approximating Item Factors using Audio

RESULTS

Input	Cos Distance	# Epochs	Time / Epoch
Audio (35s Patches)	0.25	22	~2h
Audio (Full Tracks)	0.21	-	-

Outline

- Motivation: The Cold-Start Problem

- Background: Collaborative Filtering

- **Cold-Start Music Recommendation:**
  - Estimate Collaborative Factors from Audio
  - **The Music Genome Project™**
  - Multimodal Deep Architectures

The Music Genome Project™

Attribute Examples

Breathy Voice

Nasal Voice

Odd Meter

Has Banjo

Joyful Lyrics

…

>1.5 Million tracks manually analyzed

~400 attributes per track

pandora®

Recommending Music using the MGP™

EXAMPLE

	Artist	Title
Query Track	**The Beatles**	**While My Guitar Gently Weeps**
Ranked 1	IV Thieves	The Sound And The Fury
Ranked 2	Journey	Too Late
Ranked 3	Albert Lee	Look Out Cleveland

Recommending Music using the MGP™

EXAMPLE

	Artist	Title
Query Track	The Beatles	While My Guitar Gently Weeps
Ranked 1	**IV Thieves**	**The Sound And The Fury**
Ranked 2	Journey	Too Late
Ranked 3	Albert Lee	Look Out Cleveland

Recommending Music using the MGP™

EXAMPLE

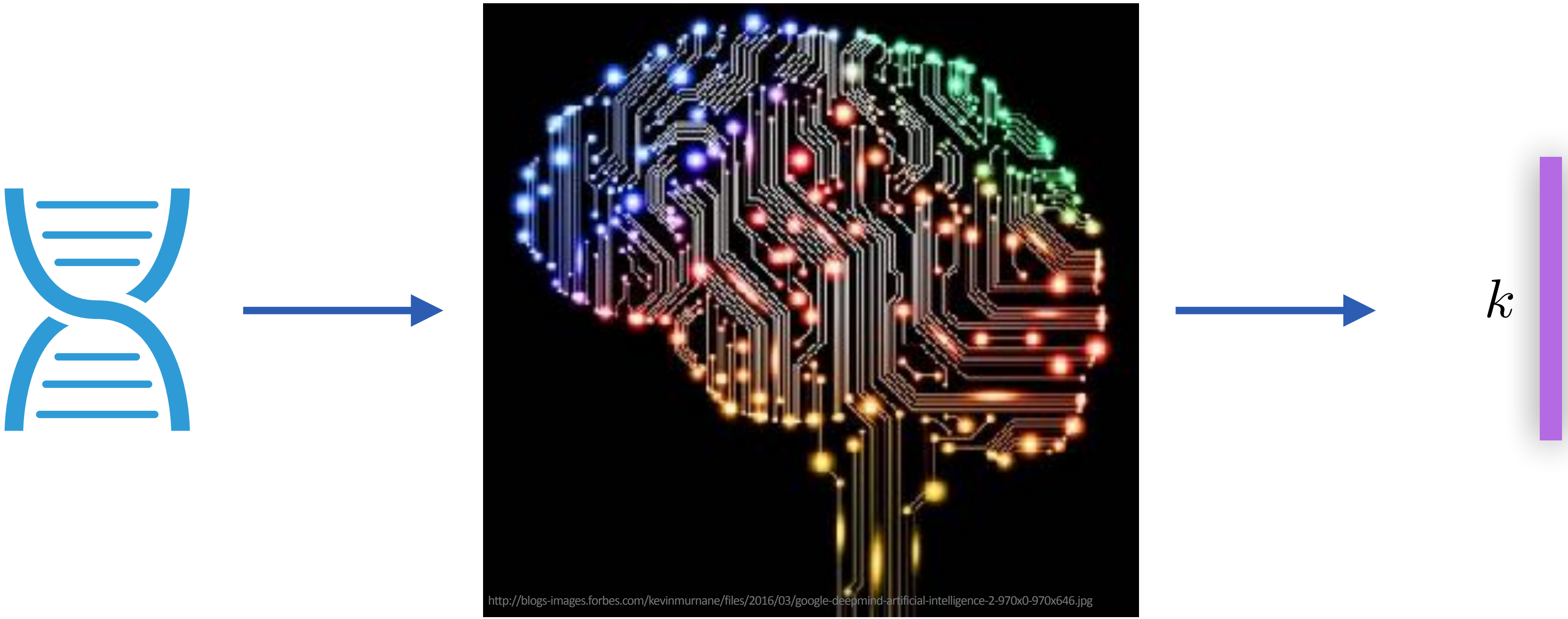
	Artist	Title
Query Track	The Beatles	While My Guitar Gently Weeps
Ranked 1	IV Thieves	The Sound And The Fury
Ranked 2	**Journey**	**Too Late**
Ranked 3	Albert Lee	Look Out Cleveland

Recommending Music using the MGP™

EXAMPLE

	Artist	Title
Query Track	The Beatles	While My Guitar Gently Weeps
Ranked 1	IV Thieves	The Sound And The Fury
Ranked 2	Journey	Too Late
Ranked 3	**Albert Lee**	**Look Out Cleveland**

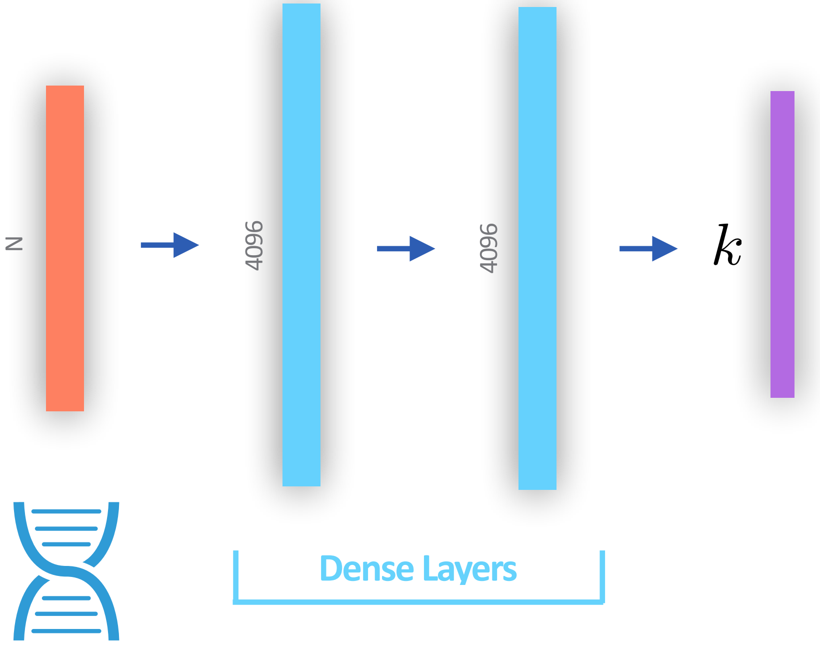
Approximate Factors using the MGP™



http://blogs-images.forbes.com/kevinmurnane/files/2016/03/google-deepmind-artificial-intelligence-2-970x0-970x646.jpg

Approximate Factors using the MGP

DEEP ARCHITECTURE



Approximating Item Factors using the MGP™

TRAINING DATA

- (Small) Data set $\{\mathbb{X}, \mathbb{Y}\}$:
  - 83k tracks (M)
  - Splits:
    - Train: 80%
    - Validation: 10%
    - Test: 10%

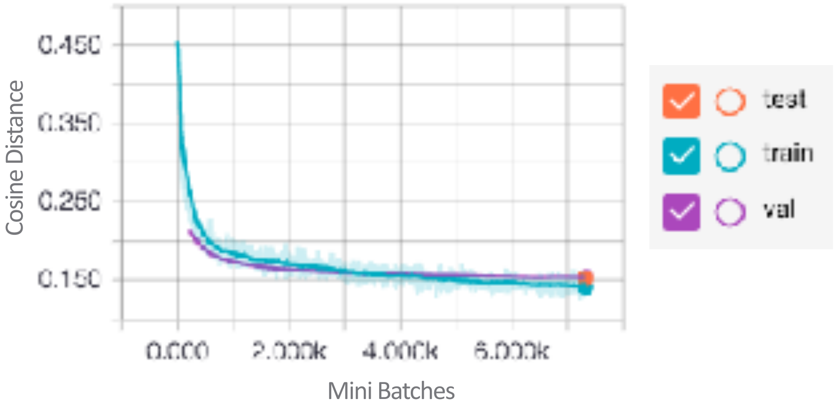
Approximating Item Factors using the MGP™

TRAINING

- Loss function:
  - Cosine Distance

$$\mathcal{L}(\theta) = 1 - \frac{1}{M} \sum_{\mathbf{x}\in\mathbb{X},\mathbf{y}\in\mathbb{Y}} \frac{f(\mathbf{x};\theta)^T\mathbf{y}}{||f(\mathbf{x};\theta)||_2||\mathbf{y}||_2}$$

- Optimization:
  - Adam (default params)
  - 50% Dropout on Dense Layers
  - Early Stopping
  - Mini-batches of 256 examples

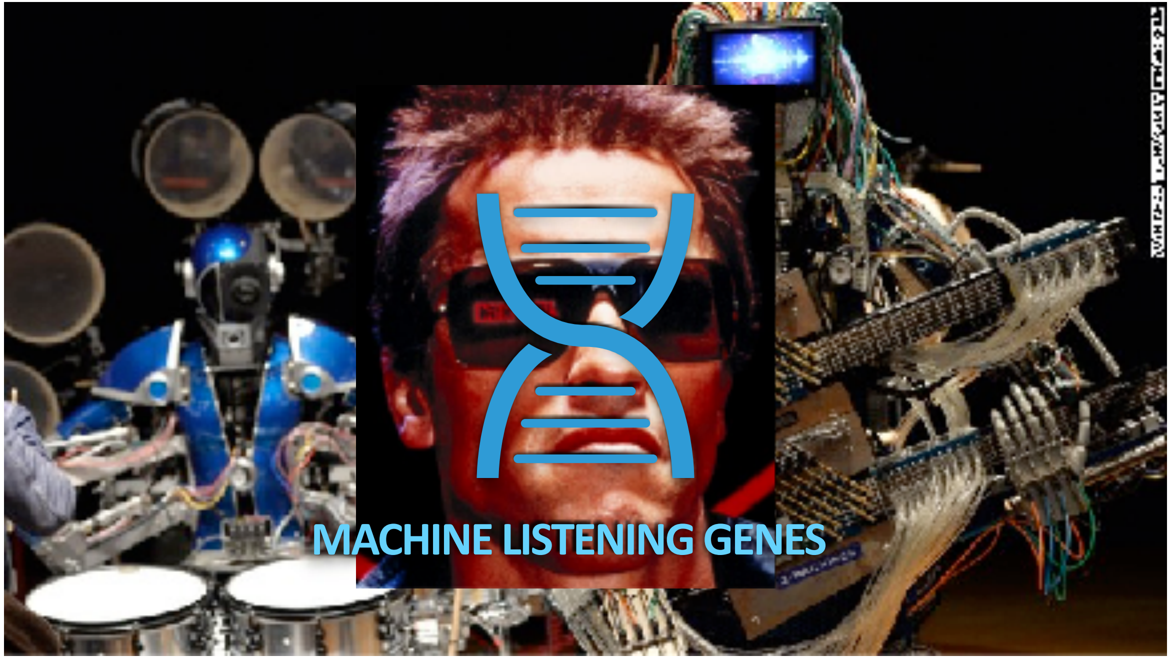


Approximating Item Factors

RESULTS

Input	Cos Distance	# Epochs	Time / Epoch
Audio (35s Patches)	0.25	22	~2h
Audio (Full Tracks)	0.21	-	-
MGP	0.15	37	7s

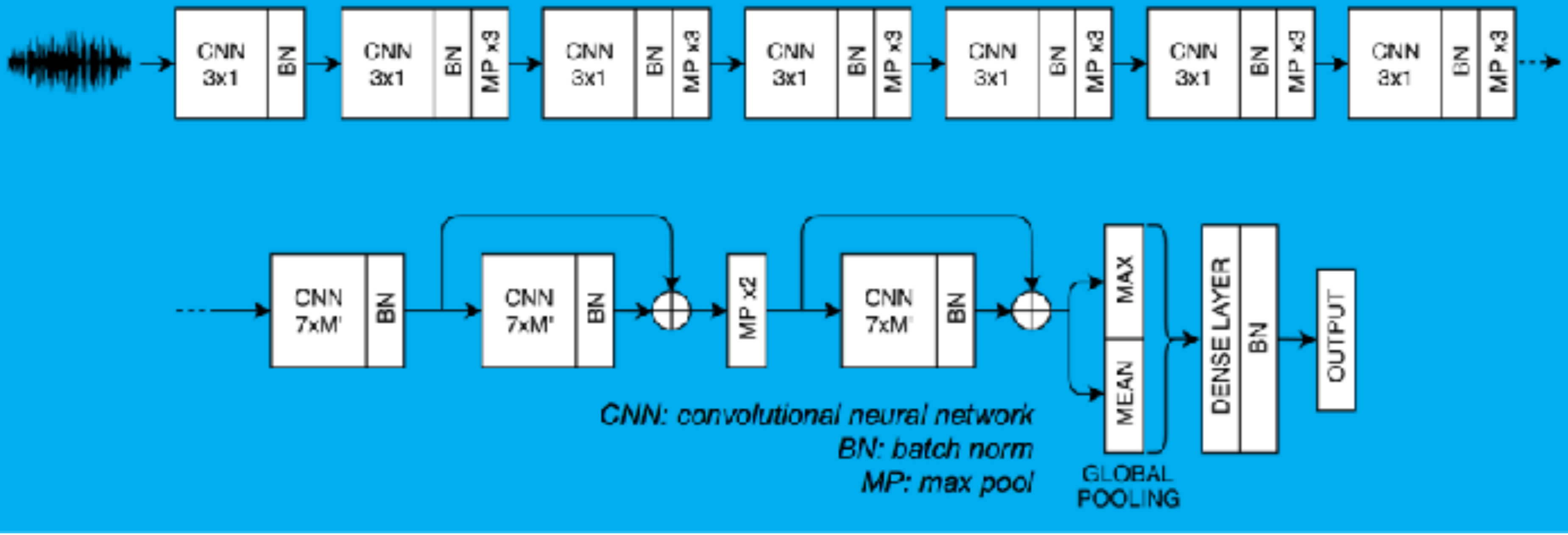
Beyond the MGP™



APPROXIMATE MGP WITH MACHINE LISTENING

(Coming soon: MGP™ Estimation with Waveforms!)

(Lee *et al.*, 2017)

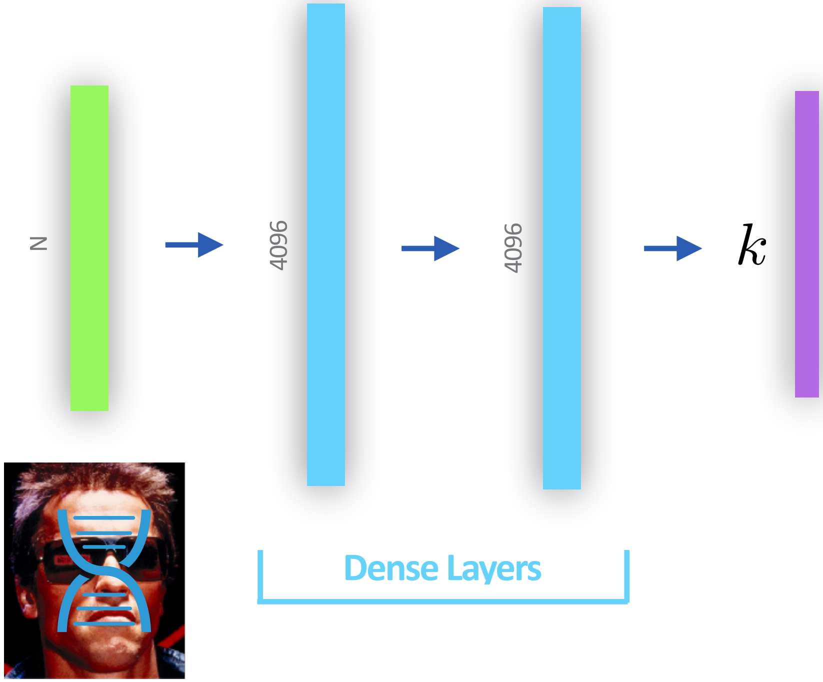


CNN: convolutional neural network
BN: batch norm
MP: max pool

APPROXIMATE MGP WITH MACHINE LISTENING

Approximate Factors using MLG

DEEP ARCHITECTURE



Approximating Item Factors

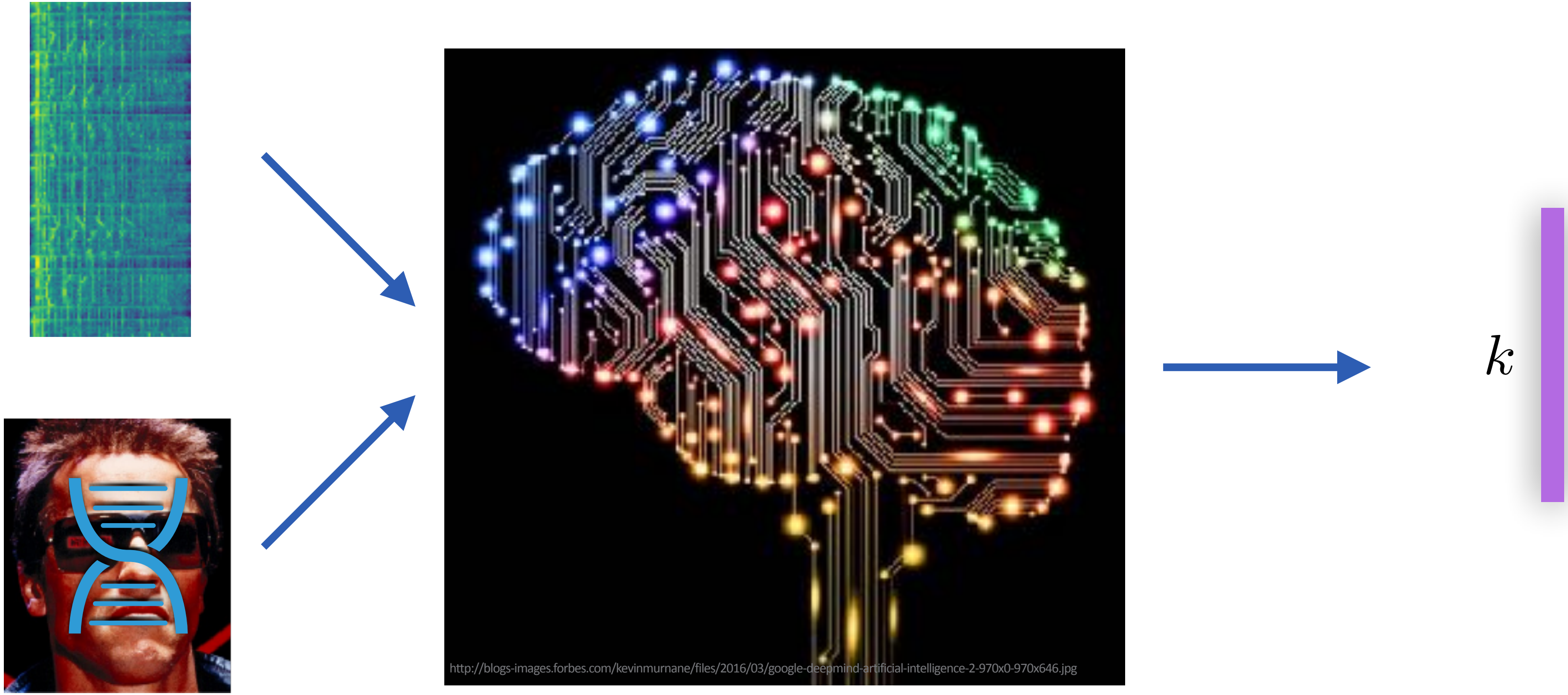
RESULTS

Input	Cos Distance	# Epochs	Time / Epoch
Audio (35s Patches)	0.25	22	~2h
Audio (Full Tracks)	0.21	-	-
MGP	0.15	37	7s
MLG	0.22	37	7s

Outline

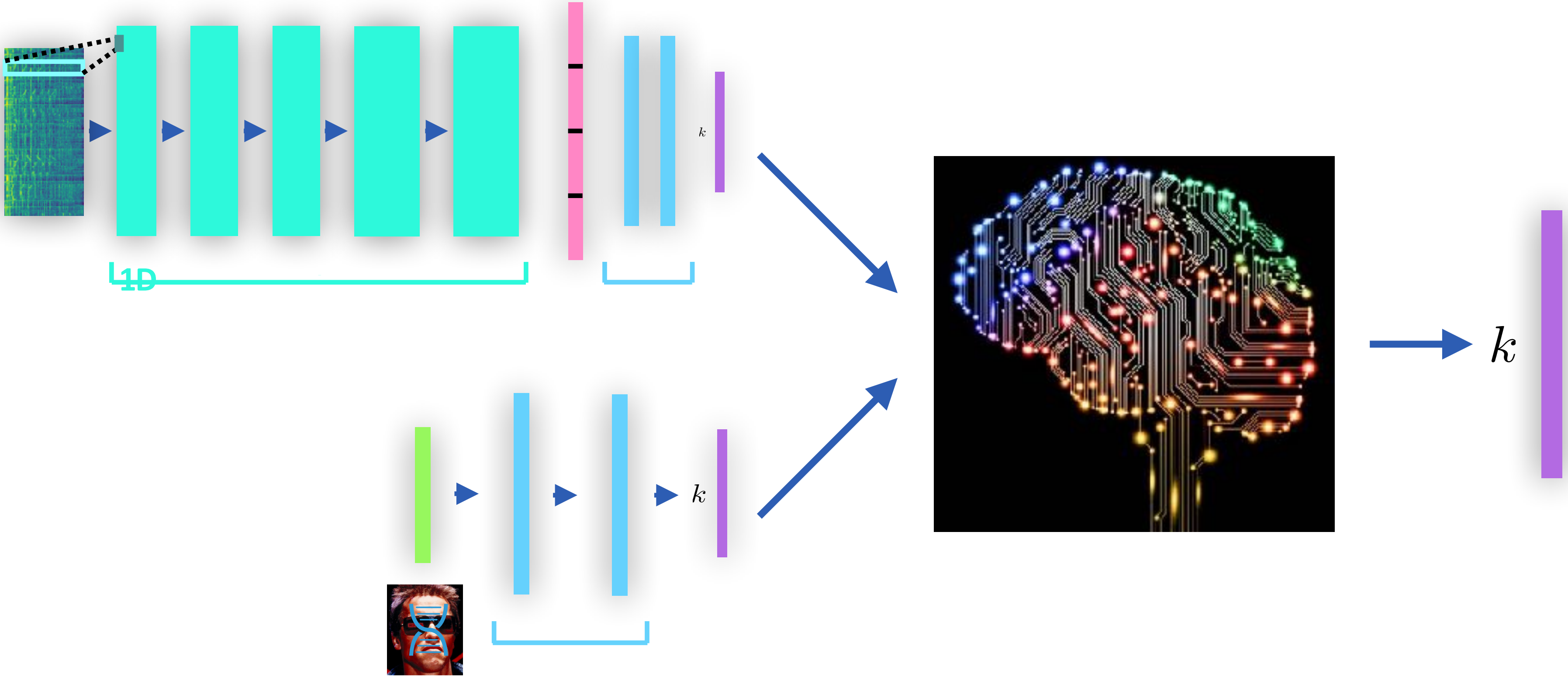
- Motivation: The Cold-Start Problem
- Background: Collaborative Filtering
- **Cold-Start Music Recommendation:**
  - Estimate Collaborative Factors from Audio
  - The Music Genome Project™
  - **Multimodal Deep Architectures**

Combine Methods to Approximate Factors



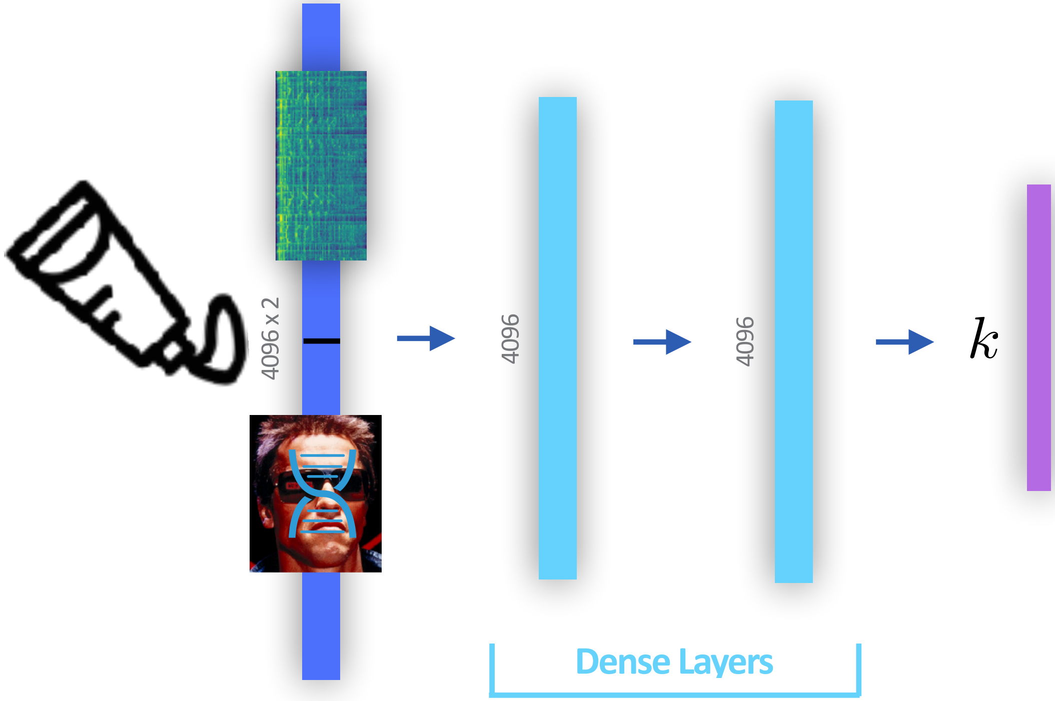
http://blogs-images.forbes.com/kevinmurnane/files/2016/03/google-deepmind-artificial-intelligence-2-970x0-970x646.jpg

Combine Methods to Approximate Factors



Combine Methods to Approximate Factors

LATE-FUSION DEEP ARCHITECTURE

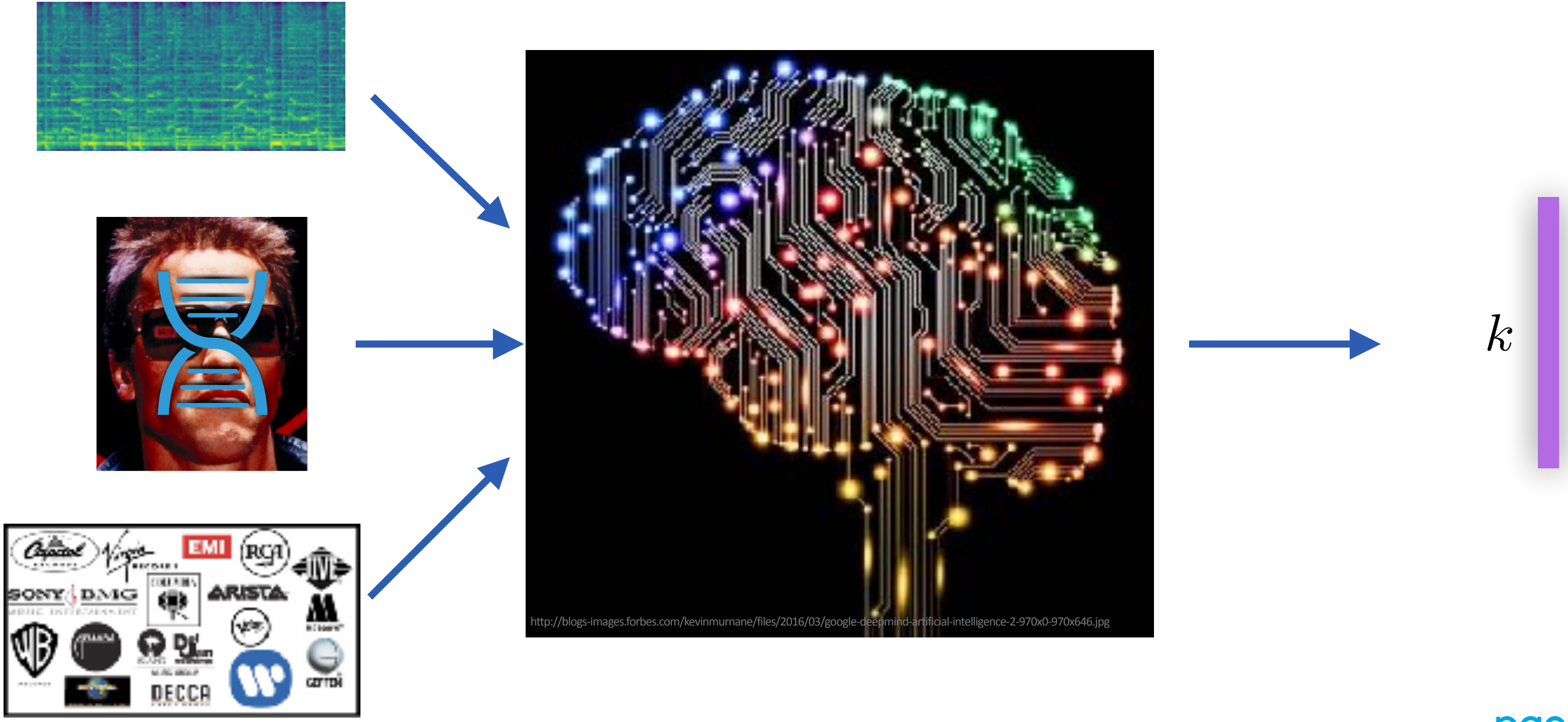


Approximating Item Factors

RESULTS

Input	Cos Distance	# Epochs	Time / Epoch
Audio (35s Patches)	0.25	22	~2h
Audio (Full Tracks)	0.21	-	-
MGP	0.15	37	7s
MLG	0.22	37	7s
Audio + MLG	0.19	37	7s

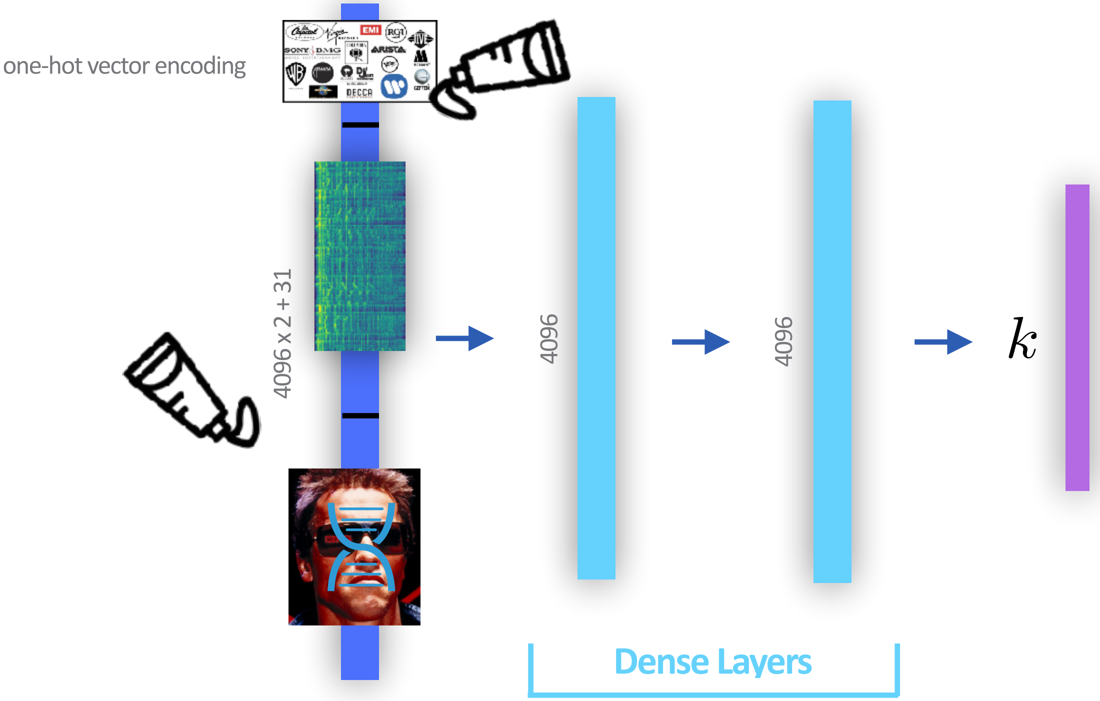
Further Multimodality to Approximate Factors



http://blogs-images.forbes.com/kevinmurnane/files/2016/03/google-deepmind-artificial-intelligence-2-970x0-970x646.jpg

Further Multimodality to Approximate Factors

LATE-FUSION DEEP ARCHITECTURE



one-hot vector encoding

Approximating Item Factors

RESULTS

Input	Cos Distance	# Epochs	Time / Epoch
Audio (35s Patches)	0.25	22	~2h
Audio (Full Tracks)	0.21	-	-
MGP	0.15	37	7s
MLG	0.22	37	7s
Audio + MLG	0.19	37	7s
Audio + MLG + genres	0.16	37	7s

More data

IS ALRIGHT

- LARGE Data set $\{\mathbb{X}, \mathbb{Y}\}$:
  - ~900k most popular tracks
  - 3 patches of 35 seconds per track (~2.7M patches =M)

Input	Trained on	Test Set	Cos Distance
Audio	SMALL	SMALL	0.21

More data

IS ALRIGHT

- LARGE Data set $\{\mathbb{X}, \mathbb{Y}\}$:
  - ~900k most popular tracks
  - 3 patches of 35 seconds per track (~2.7M patches =M)

Input	Trained on	Test Set	Cos Distance
Audio	SMALL	SMALL	0.21
Audio	LARGE	LARGE	0.37

pandora®

More data

IS ALRIGHT

- LARGE Data set $\{\mathbb{X}, \mathbb{Y}\}$:
  - ~900k most popular tracks
  - 3 patches of 35 seconds per track (~2.7M patches =M)

Input	Trained on	Test Set	Cos Distance
Audio	SMALL	SMALL	0.21
Audio	LARGE	LARGE	0.37
Audio	SMALL	LARGE	**0.64**

More data

IS ALRIGHT

- LARGE Data set $\{\mathbb{X}, \mathbb{Y}\}$:
  - ~900k most popular tracks
  - 3 patches of 35 seconds per track (~2.7M patches =M)

Input	Trained on	Test Set	Cos Distance
Audio	SMALL	SMALL	0.21
Audio	LARGE	LARGE	0.37
Audio	SMALL	LARGE	**0.64**
Audio	LARGE	SMALL	**0.21**

Recommendation Examples

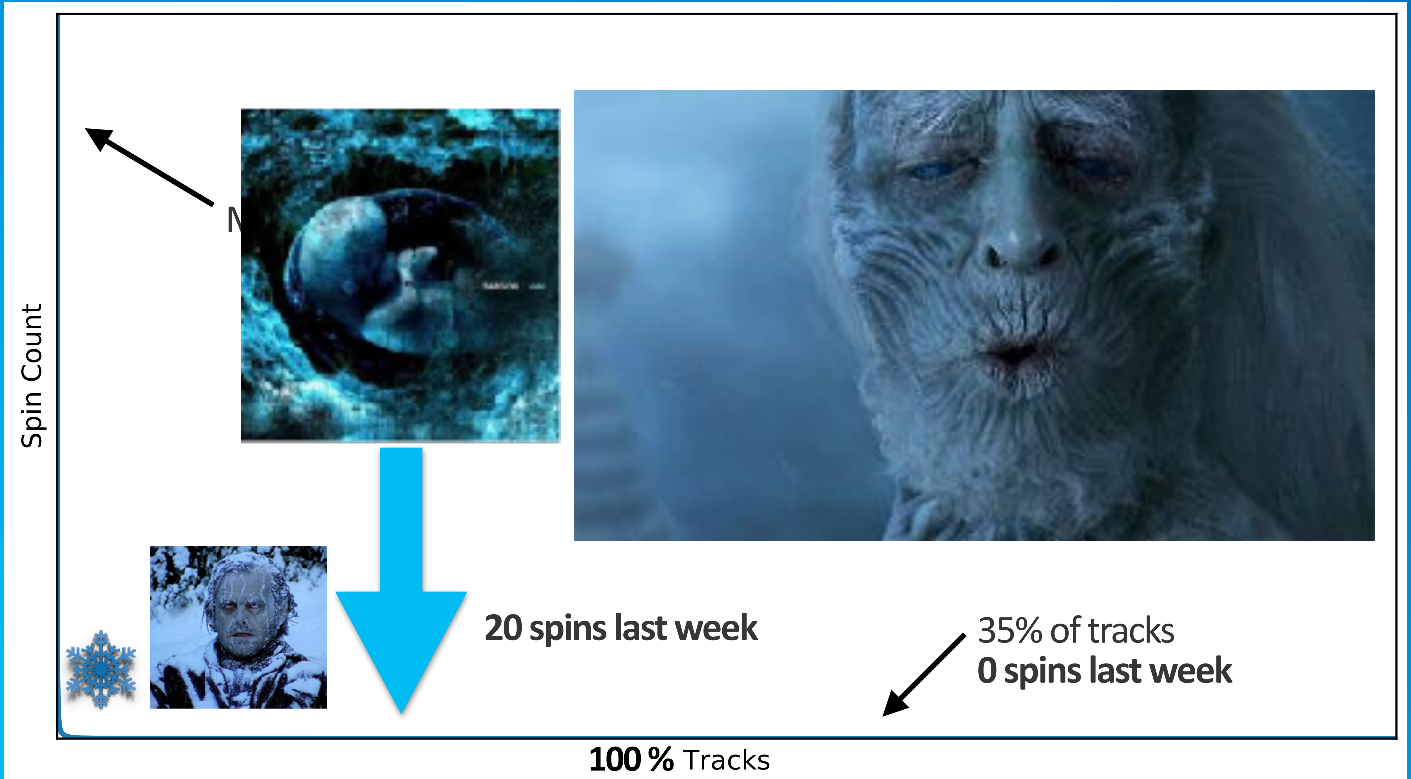
	Artist	Title
Query Track	**The Beatles**	**While My Guitar Gently Weeps**
Ranked 1	Bob Dylan	Knockin’ On Heavens Door
Ranked 2	Neil Young	Heart Of Gold
Ranked 3	The Rolling Stones	Angie

Recommendation Examples

	Artist	Title
Query Track	**Sargon**	**Continuarà**
Ranked 1	Mudvayne	Happy?
Ranked 2	Mudvayne	Forget To Remember
Ranked 3	Stone Sour	Hell & Consequences

pandora®

Long Tail Context



20 spins last week

35% of tracks
0 spins last week

100 % Tracks

Recommendation Examples

	Artist	Title
Query Track	**Sargon**	**Continuarà**
Ranked 1	Mudvayne	Happy?
Ranked 2	Mudvayne	Forget To Remember
Ranked 3	Stone Sour	Hell & Consequences

Recommendation Examples

	Artist	**Title**
Query Track	Sargon	Continuarà
Ranked 1	**Mudvayne**	**Happy?**
Ranked 2	Mudvayne	Forget To Remember
Ranked 3	Stone Sour	Hell & Consequences

Recommendation Examples

	Artist	Title
Query Track	Sargon	Continuarà
Ranked 1	Mudvayne	Happy?
Ranked 2	**Mudvayne**	**Forget To Remember**
Ranked 3	Stone Sour	Hell & Consequences

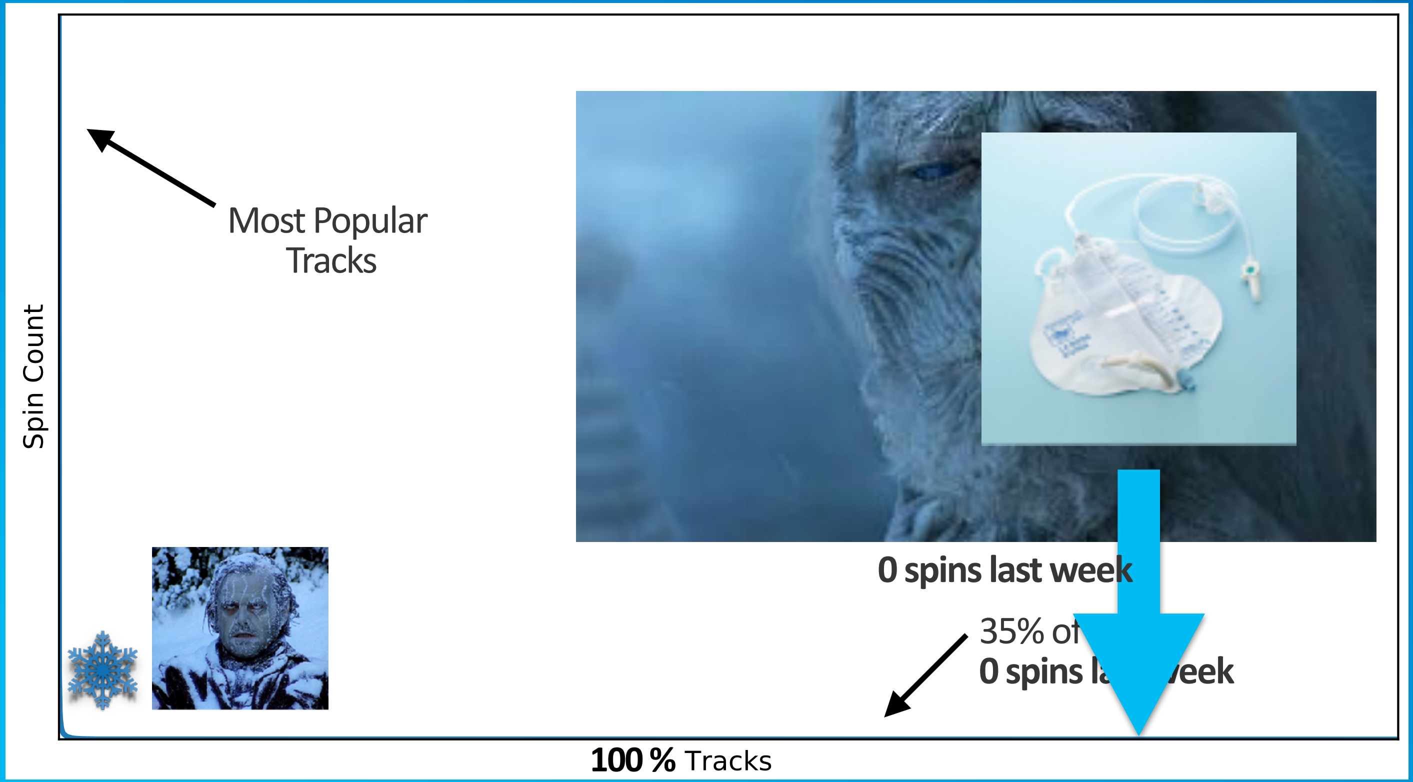
Recommendation Examples

	Artist	Title
Query Track	Sargon	Continuarà
Ranked 1	Mudvayne	Happy?
Ranked 2	Mudvayne	Forget To Remember
Ranked 3	**Stone Sour**	**Hell & Consequences**

Recommendation Examples

	Artist	Title
Query Track	**La Bossa d'Urina**	**El Tiempo**
Ranked 1	Il Divo	Hallelujah
Ranked 2	Sarah Brightman & The London Symphony Orchestra	Time To Say Goodbye
Ranked 3	Andrea Bocelli	Amapola

Long Tail Context



Recommendation Examples

	Artist	Title
Query Track	**La Bossa d'Urina**	**El Tiempo**
Ranked 1	Il Divo	Hallelujah
Ranked 2	Sarah Brightman & The London Symphony Orchestra	Time To Say Goodbye
Ranked 3	Andrea Bocelli	Amapola

Recommendation Examples

	Artist	**Title**
Query Track	La Bossa d'Urina	El Tiempo
Ranked 1	**Il Divo**	**Hallelujah**
Ranked 2	Sarah Brightman & The London Symphony Orchestra	Time To Say Goodbye
Ranked 3	Andrea Bocelli	Amapola

Recommendation Examples

	Artist	**Title**
Query Track	La Bossa d’Urina	El Tiempo
Ranked 1	Il Divo	Hallelujah
Ranked 2	**Sarah Brightman & The London Symphony Orchestra**	**Time To Say Goodbye**
Ranked 3	Andrea Bocelli	Amapola

Recommendation Examples

	Artist	**Title**
Query Track	La Bossa d’Urina	El Tiempo
Ranked 1	Il Divo	Hallelujah
Ranked 2	Sarah Brightman & The London Symphony Orchestra	Time To Say Goodbye
Ranked 3	**Andrea Bocelli**	**Amapola**

ENSEMBLE OF RECOMMENDERS MAY PRODUCE OPTIMAL RECOMMENDATIONS

MAN vs MACHINE?

MAN + MACHINE

MAN + MACHINE

“Mix of Art and Science”

Oramas, S., Nieto, O., Sordo, M., Serra, X., A Deep Multimodal Approach for Cold-start Music Recommendation. Deep Learning for Recommender Systems Workshop, RecSys, Como, Italy 2017

Oramas, S., Nieto, O., Barbieri, F., Serra, X., Multi-label Music Genre Classification From Audio, Text, and Images Using Deep Features. Proc. of the 18th International Society for Music Information Retrieval Conference (ISMIR). Suzhou, China, 2017

MAN + MACHINE

"Mix of Art and Science"

Oramas, S., Nieto, O., Sordo, M., Serra, X., A Deep Multimodal Approach for Cold-start Music Recommendation. Deep Learning for Recommender Systems Workshop, RecSys, Como, Italy 2017

Oramas, S., Nieto, O., Barbieri, F., Serra, X., Multi-label Music Genre Classification From Audio, Text, and Images Using Deep Features. Proc. of the 18th International Society for Music Information Retrieval Conference (ISMIR). Suzhou, China, 2017

MAN + MACHINE

“Mix of Art and Science”

THANKS!

ONIETO@PANDORA.COM