Modelling Chart Trajectories Using Song Features

Jonathan Vi Perrie

August 16^{th}, 2019

UNIVERSITY OF WATERLOO

Outline

1. **Overview:** What are chart trajectories? Why do we want to model them?
2. **Related work:** What areas do we borrow from?
3. **Data & pre-processing:** What songs and specific attributes are we interested in analyzing?
    - Song features
    - Target variables
4. **Modelling:** How can we use modelling to show the significance of our target variables?
    - Hit or flop?
    - Peak position versus weeks on chart
    - Multiclass classification
5. **Critique:** How do our findings compare to a related study?
6. **Conclusions:** What did we learn?
    - Summary
    - Limitations

Overview

Motivations

In hit song science, songs are usually classified as hits or flops based on one dimension of song success [AM17, IKW+18, BP18, DL05].

Contribution

We propose multiple definitions for song success that explicitly represent a song's position and time on the charts.

Significance

With richer definitions for song success, one can better comprehend the impact of song features.

Related work

Modelling

- **Prediction:** Genre classification [TC02], emotion detection [HDE09], and hit song science [DL05]
- **Explanation:** What attributes increase the likelihood of a song associating with some class? [AM17]

Feature analysis

- **Lyric features:** Topic modelling [JRJR13]
- **Audio features:** Signal processing [SCB$^+$12]

Data & pre-processing: Song features

Datasets

- The Billboard Hot 100
- The Million Song Dataset
    - The musiXmatch Dataset
- Spotify’s The Echo Nest audio features

Data & pre-processing: Song features

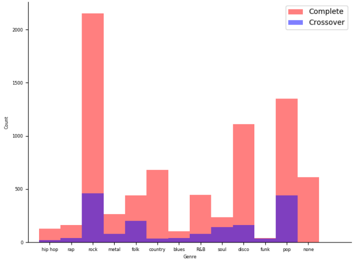
Pre-processing

- **Genre:** Custom set of genres based on artist Spotify tags.
- **Topic mixtures:** Latent Dirichlet allocation (LDA) topic mixture lyric features
- **Similarity measures:** Chart similarity, genre similarity, and artist similarity
  - The Echo Nest audio features
  - LDA lyric features
- **Baseline song attributes:** Control variables
  - Time blocks
  - Past charting songs by an artist

Data & pre-processing: Genre

Method

1. Find tags associated with each genre.
2. Sub-select most frequent tags above some threshold.
3. Find top set of most distinct tags using tf-idf for each genre.
4. Match artists to genres (crossovers are ties).

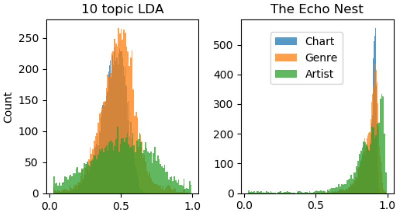


Data & pre-processing: Similarity measures

Weighted cosine similarity measure normalized over all songs.

$$w = \mathrm{e}^{\frac{-wks}{52}} \tag{1}$$

1. **Chart similarity:** Charting songs from previous year.
2. **Genre similarity:** Charting songs from previous year and same genre.
3. **Artist similarity:** Charting songs from same artist.



Data & pre-processing: Target variable

Chart position data

Songs have temporal position data for their chart activity.

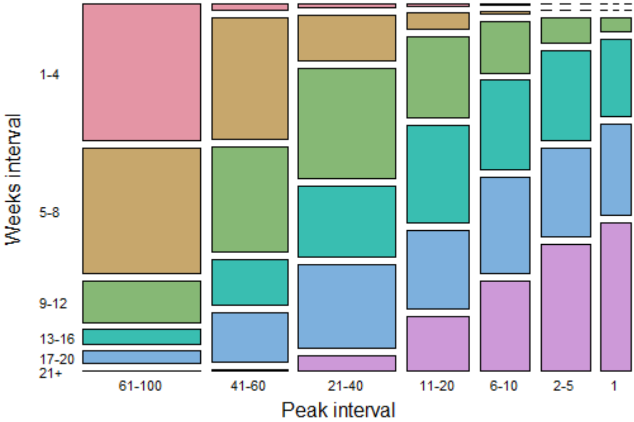
Interval target variable

- Can derive peak position and weeks on chart measures.
- Taking the product of the discretized version of these measures produces the interval classes.

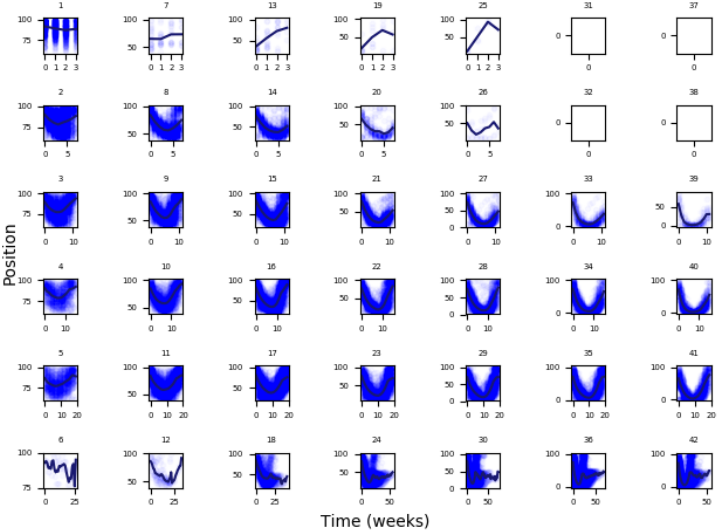
Alignment target variable

- Alignment time series using dynamic time warping.
- Clustering alignment distances using complete linkage hierarchical clustering produces the alignment classes.

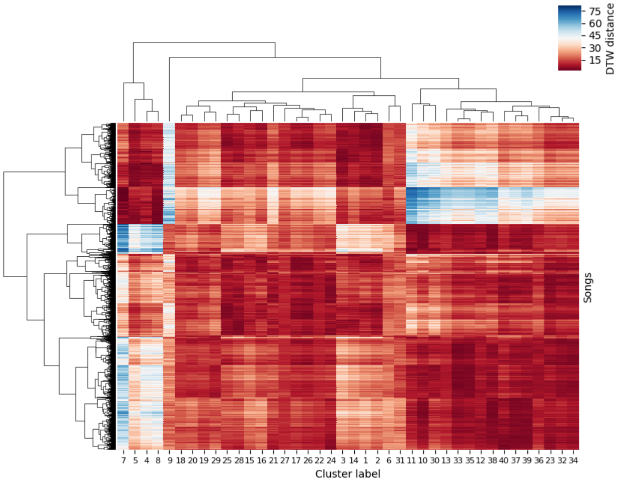
Data & pre-processing: Interval target variable



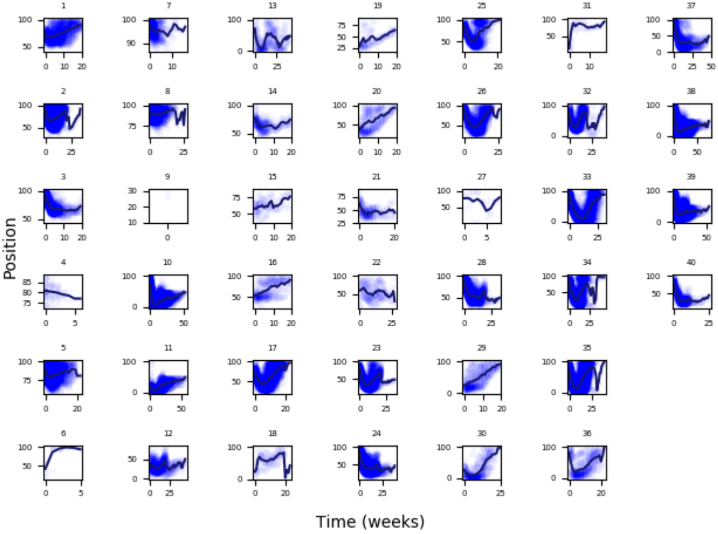
Data & pre-processing: Interval target variable



Data & pre-processing: Alignment target variable



Data & pre-processing: Alignment target variable



Data & pre-processing: Summary

- Gathered and processed data from sources to build song features.
- Constructed feature sets from baseline, The Echo Nest, and 10, 20, 40, and 80 topic mixture features.
- Developed interval and alignment target variables using song chart data.
- Interval target variables have more arch-like trajectories.
- Alignment target variables are able to represent more obscure trajectories.

Modelling: Research questions

1. Can we model hits and flops using our target variable definitions?

2. Are the same features important for distinguishing songs by peak position and weeks on chart?

3. Can we perform multiclass classification on our data using our target variable definitions?

Modelling: Hit or flop?

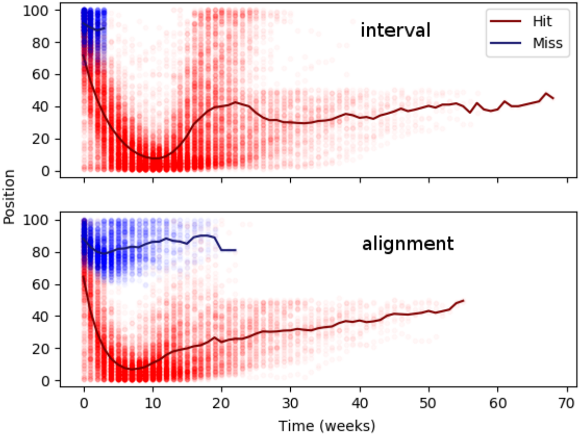
Modelling binary target variable tasks using logistic regression with various feature sets.

Model evaluation

- **Target variables:** Interval versus alignment target variable
- **Feature sets:** Comparing baseline, The Echo Nest, 10, 20, 40 and 80 topic mixture feature sets
- **Explanatory measure:** Akaike Information Criterion (AIC)
- **Predictive measure:** F1 score under 10-fold cross-validation

Modelling: Hit or flop?

Binary classes



Modelling: Hit or flop?

Explanatory results

- **AIC:** Alignment (The Echo Nest) / interval (20 topic mixture)
- **Similarity:** Time blocks had significant effects with large magnitude
- **Contrast:** Acousticness and topic mixture effects

Predictive results

Interval target variable

	metadata	The Echo Nest	10 topic	20 topic	40 topic	80 topic
F1	0.769	0.778	0.772	0.783	0.774	0.775

Alignment target variable

	metadata	The Echo Nest	10 topic	20 topic	40 topic	80 topic
F1	0.809	0.816	0.815	0.811	0.802	0.783

Modelling: Peak position versus weeks on chart

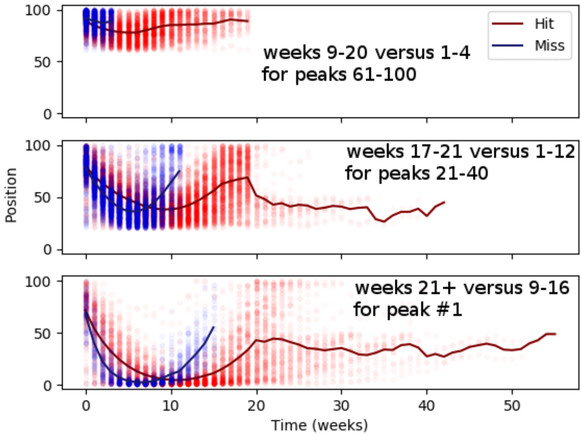
Modelling one-dimensional tasks using lasso logistic regression with some optimal interval target variable feature set.

Model evaluation

- **Target variables:** Peak position versus weeks on chart
- **Task target variable ranges:** Differences in values at separate fixed intervals
- **Baseline attributes:** Full versus reduced feature sets
- **Explanatory measure:** % deviations explained relative to null model
- **Predictive measure:** F1 score under 10-fold cross-validation

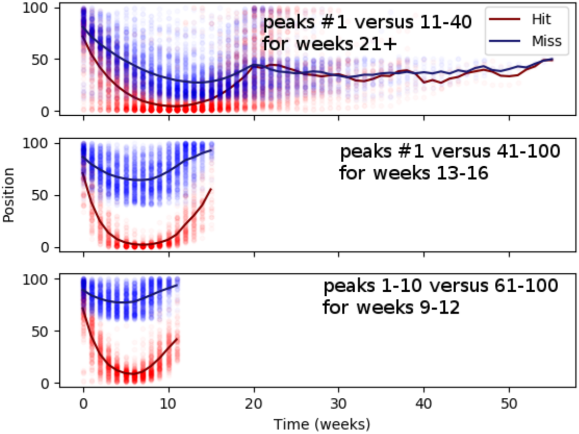
Modelling: Peak position versus weeks on chart

Time task binary classes



Modelling: Peak position versus weeks on chart

Peak task binary classes



Modelling: Peak position versus weeks on chart

Explanatory results

- **% Deviations:** Weeks on chart (C, C) / peak position (C, B)
- **Similarity:** Acousticness and valence within task sets
- **Contrast:** Acousticness and valence between target variables

Predictive results

weeks on chart tasks

	A	**A control**	**B**	**B control**	**C**	**C control**
F1	0.661	0.575	0.849	0.710	0.946	0.828

Peak position tasks

	A	**A control**	**B**	**B control**	**C**	**C control**
F1	0.697	0.638	0.894	0.750	0.936	0.790

Modelling: Multiclass classification

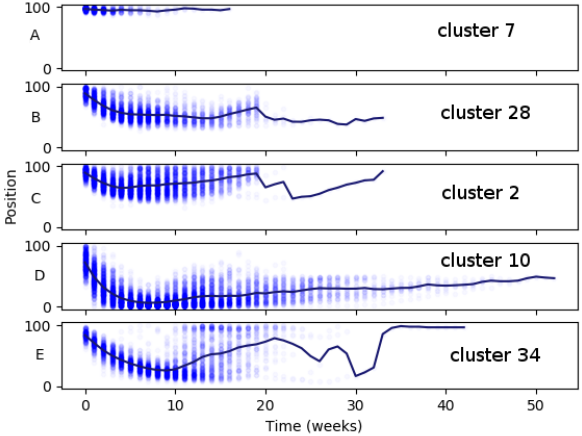
Modelling multiple alignment classes using a one-against-all lasso logistic regression with some feature set.

Model evaluation

- **Baseline attributes:** Full versus reduced feature sets
- **Explanatory measure:** % deviations explained relative to null model
- **Predictive measure:** Accuracy under 10-fold cross-validation

Modelling: Multiclass classification

Alignment classes



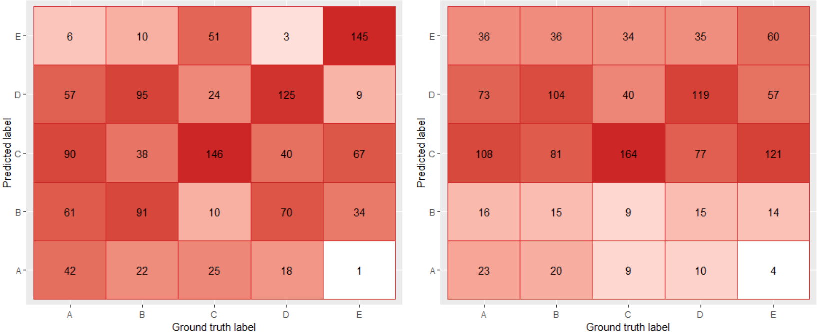
Modelling: Multiclass classification

Explanatory results

- **% Deviations:** Full (20.9%) / Reduced (3.5%)
- **Similarity:** Acousticness effects sign
- **Contrast:** Acousticness spread across classes

Modelling: Multiclass classification

Predictive results



Modelling: Summary

- We demonstrated that our classes are separable.
- Much of the separability can be attributed to baseline attributes.
- There are substantial differences in some of the feature effects for peak position and weeks on chart tasks.
- Multiclass classification faces the same challenges with baseline attributes.

Critique: Overview

Inspiration

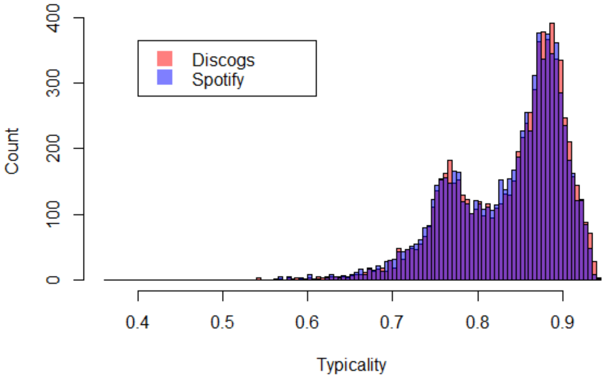
Our work is partially based on Askin & Mauskapf's research into modelling song performance based on a song similarity [AM17].

Workflow

- Used songs from Billboard Hot 100.
- Gathered The Echo Nest audio features and baseline attributes.
- Computed a genre-weighted typicality measure to test the importance of song similarity in explaining performance.
- Modelled peak position using an ordered logit with the features

Critique: Areas of concern

Typicality range



Critique: Areas of concern

Genre sensitivity analysis

- Used Spotify genre definitions instead of Discogs.
- Typicality effect shrunk and became less significant in one model.

	Dependent variable:			
	Peak position (inverted)			
	(Discogs)	(Discogs)	(Spotify)	(Spotify)
typicality	-2.020^{***}	8.079	-1.175^{*}	9.999
	(0.681)	(5.486)	(0.673)	(6.205)
typicality^2		-6.413^{*}		-7.071^{*}
		(3.482)		(3.917)
Observations	7,683	7,683	7,683	7,683
Note:		*p$<$0.1; **p$<$0.05; ***p$<$0.01		

Conclusions

Summary

- Proposed target variables that incorporate both temporal and position-based aspects of a song's performance.
- Identified shortcomings of a related explanatory work.

Limitations

- Data coverage
- Feature granularity

Thank you!

Thanks to...

- **Robin:** For making being a TA so easy.
- **Jesse:** For teaching Affective Computing.
- **Peter:** For checking in on me when things went silent.
- **Dan:** For being supportive when things were bumpy.

- Noah Askin and Michael Mauskapf, *What makes popular culture popular? product features and optimal differentiation in music*, American Sociological Review **82** (2017), no. 10, 910–944.
- Jonah Berger and Grant Packard, *Are atypical things more popular?*, Psychological Science **29** (2018), no. 7, 1178–1184.
- Ruth Dhanaraj and Beth Logan, *Automatic prediction of hit songs.*, ISMIR, 2005, pp. 488–491.
- Xiao Hu, J Stephen Downie, and Andreas F Ehmann, *Lyric text mining in music mood classification*, ISMIR, 2009, pp. 411–416.

- Myra Interiano, Kamyar Kazemi, Lijia Wang, Jienian Yang, Zhaoxia Yu, and Natalia L Komarova, *Musical trends and predictability of success in contemporary songs in and out of the top charts*, Royal Society Open Science **5** (2018), no. 5.
- Cora Johnson-Roberson and Matthew Johnson-Roberson, *Temporal and regional variation in rap lyrics*, NIPS Workshop on Topic Models: Computation, Application and Evaluation, 2013.
- Joan Serrà, Álvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll Arcos, *Measuring the evolution of contemporary western popular music*, Scientific Reports **2** (2012), 521–526.
- George Tzanetakis and Perry Cook, *Musical genre classification of audio signals*, IEEE Transactions on Speech and Audio Processing **10** (2002), no. 5, 293–302.

UNIVERSITY OF WATERLOO