HIT SONG DETECTION USING LYRIC FEATURES ALONE

Abhishek Singhi **Daniel G. Brown**

University of Waterloo
Cheriton School of Computer Science
{asinghi,dan.brown}@uwaterloo.ca

ABSTRACT

We develop a hit detection model using 31 rhyme, syllable and meter features. Hits are songs which made it to the Billboard Year-End Hot 100 singles between the years 2008 and 2013; in some cases we use fewer than the top 100 songs. Flops are non-hit songs by 51 singers who have hits. We train a Bayesian network on 492 hits and 6323 flops. Using a 10 fold cross validation gives us recall and precision values of 0.451 and 0.214 respectively for the hits, which is much stronger than would be expected by random chance.

1. METHOD

We use the complete set of 24 rhyme and syllable features of the Rhyme Analyzer [2], which includes features like syllables per line, rhymes per line, and links per line. We add seven new meter features identifying the fraction of lines written in iambic, trochaic, spondaic, anapestic, dactylic, amphibrachic and pyrrhic meter, using the CMU Pronunciation Dictionary [5] to transcribe plain lyrics to a sequence of phonemes with indicated stress. Our hits are songs which made it to the Billboard Year-End Hot 100 singles between the years 2008-2013, eliminating duplicate songs repeated across two years and ones with noisy lyrics: wrong spellings, wrong lyrics or repetitions of meaningless syllables like "*lalalala*". Flops are the non-hit songs by 51 popular singers of hits. On manual inspection of 100 flop lyrics we observe that lyrics of flops that are shorter than thirty lines are very noisy; in particular, it is hard to reliably predict rhyme features. Thus we only study songs with at least thirty lines of lyrics.

We generate a Bayesian network for these 31-element feature vectors on the resulting 492 hits and 6323 flops using the Bayesian network module from Weka [1]. We use ten-fold cross validation and report the confusion matrices for the network that maximizes data likelihood.

2. RESULTS

The confusion matrix obtained from the Bayesian Network is shown in Table 1.

We were interested to see the performance of our approach when we restricted the definition of hit songs by selecting only the top k songs of the Billboard Year-End Hot 100. In this case our flop set is the other songs of some of the artists whose songs appear in the hit set; this makes the problem harder, as these artists' "flops" would be an enviable success for many artists.

Our approach works poorly when we seek to identify top 5 or 10 songs, but surprisingly well when used on the top 25 and 35 songs. The confusion matrix for $k = 15$, 25 and 35 is shown in Tables 2, 3 and 4 respectively.

True value	Prediction	
	Hits	Flops
Hits	222	270
Flops	813	5510

Table 1. The confusion matrix for our largest experiment. 222 hits were correctly classified as hits while 813 flops were misclassified as hits.

True value	Prediction	
	Hits	Flops
Hits	6	78
Flops	67	942

Table 2. The confusion matrix for $k = 15$, when we only have fifteen hit songs per year. Note that the flop set decreases in size as well: only authors of hits are used in our definition of what is a flop.

True value	Prediction	
	Hits	Flops
Hits	49	86
Flops	212	1722

Table 3. The confusion matrix for $k = 25$, when we only have twenty five hit songs per year. Note that the flop set decreases in size as well: only authors of hits are used in our definition of what is a flop.

Considering flops to be non-hit songs between the years 2008-2013 by the 51 singers of hit songs, and returning to the full top 100 hits from each year yields the confusion matrix shown in Table 5, which has similar precision, but lower recall compared to the values in table 1.

We observe that the performance of our algorithm increases considerably as the length of the lyrics increases.

We believe that this is because the probability of lyrics being noisy decreases as its length increases; we verify this by manually inspecting flops. Repeating the above experiment with flops which are at least fifty lines long we get the confusion matrix shown in Table 6. As most of the hit lyrics are lengthier and relatively noise free we do not eliminate them based on their line count. The results are considerably better than the ones in Table 5 and hence our model works extremely well for relatively noiseless lengthy lyrics.

True value	Prediction	
	Hits	Flops
Hits	80	106
Flops	281	1797

Table 4. The confusion matrix for k = 35, when we only have thirty five hit songs per year. Note that the flop set decreases in size as well: only authors of hits are used in our definition of what is a flop.

True value	Prediction	
	Hits	Flops
Hits	130	362
Flops	495	3164

Table 5. The confusion matrix obtained when we consider the flops between the years 2008-2013 only.

True value	Prediction	
	Hits	Flops
Hits	218	274
Flops	305	1780

Table 6. The confusion matrix obtained when we consider the flops which are at least 50 lines long between the years 2008-2013 only. Precision = 0.4168 and recall = 0.4430.

3. CONCLUSION

We are unaware of any hit detection model based on lyrics features. Thus, our work is novel, and it works surprisingly well, outperforming hit detection models that identify audio features using signal processing [3], [4]. It is interesting that as lyric length increases, the quality of outcome also improves. An extension might be to combine these features with features derived from either recordings or scores.

4. REFERENCES

[1] M. Hall et al.: "The WEKA data mining software: An update" *ACM SIGKDD Explorations Newsletter*, volume 11, number 1, pages 10–18, 2009.

[2] H. Hirjee and D.G. Brown: "Rhyme Analyzer: An Analysis Tool for Rap Lyrics," in *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010.

[3] F. Pachet and P. Roy: "Hit song science is not yet a science," in *Proceedings of the 9th International Conference on Music Information Retrieval*, pages 355-360, 2008.

[4] Y. Ni, R.S. Rodriguez, M. Mcvicar, and T. D. Bie, "Hit Song Science Once Again a Science?" in *Proceedings of the 4th International Workshop on Machine Learning and Music: Learning from Musical Structure*, 2011.

[5] H.S. Elovitz, R.W. Johnson, A. McHugh and J.E. Shore, "Automatic translation of English text to phonetics by means of letter-to-sound rules," US Naval Research Laboratory Report 7948, January 1976.