ARE POETRY AND LYRICS ALL THAT DIFFERENT?

Abhishek Singhi **Daniel G. Brown**
University of Waterloo
Cheriton School of Computer Science
{asinghi,dan.brown}@uwaterloo.ca

ABSTRACT

We hypothesize that different genres of writing use different adjectives for the same concept. We test our hypothesis on lyrics, articles and poetry. We use the English Wikipedia and over 13,000 news articles from four leading newspapers for the article data set. Our lyrics data set consists of lyrics of more than 10,000 songs by 56 popular English singers, and our poetry dataset is made up of more than 20,000 poems from 60 famous poets. We find the probability distribution of synonymous adjectives in all the three different categories and use it to predict if a document is an article, lyrics or poetry given its set of adjectives. We achieve an accuracy level of 67% for lyrics, 80% for articles and 57% for poetry. Using these probability distribution we show that adjectives more likely to be used in lyrics are more rhymable than those more likely to be used in poetry, but they do not differ significantly in their semantic orientations. Furthermore we show that our algorithm is successfully able to detect poetic lyricists like Bob Dylan from non-poetic ones like Bryan Adams, as their lyrics are more often misclassified as poetry.

1. INTRODUCTION

The choice of a particular word, from a set of words that can instead be used, depends on the context we use it in, and on the artistic decision of the authors. We believe that for a given concept, the words that are more likely to be used in lyrics will be different from the ones which are more likely to be used in articles or poems, because lyricists have different objectives typically. We test our hypothesis on adjective usage in these categories of documents. We use adjectives, as a majority have synonyms that can be used depending on context. To our surprise, just the adjective usage is sufficient to separate documents quite effectively.

Finding the synonyms of a word is still an open problem. We used three different sources to obtain synonyms for a word – the WordNet, Wikipedia and an online thesaurus. We prune synonyms, obtained from the three sources, which fall below an experimentally determined threshold for the semantic distance between the synonyms and the word. The list of relevant synonyms obtained after pruning was used to obtain the probability distribution over words.

A key requirement of our study is that there exists a difference, albeit a hazy one, between poetry and lyrics. Poetry attracts a more educated and sensitive audience while lyrics are written for the masses. Poetry, unlike lyrics, is often structurally more constrained, adhering to a particular meter and style. Lyrics are often written keeping the music in mind while poetry is written against a silent background. Lyrics, unlike poetry, often repeat lines and segments, causing us to believe that lyricists tend to pick more rhymable adjectives; of course, some poetic forms also repeat lines, such as the villanelle. For twenty different concepts we compare adjectives which are more likely to be used in lyrics rather than poetry and vice versa.

Even in my heart I see
You're not bein' **true** to me
Deep within my soul I feel
Nothing's like it used to be
Sometimes I wish I could turn back time
Impossible as it may seem
But I wish I could so **bad** baby
Quit playin' games with my heart

Figure 1. The bold-faced words are the adjectives our algorithm takes into account while classifying a document, which in this case in a snippet of lyrics by the Backstreet Boys.

We use a bag of words model for the adjectives, where we do not care about their relative positions in the text, but only their frequencies. Finding synonyms of a given word is a vital step in our approach and since it is still considered a difficult task improvement in synonyms finding approaches will lead to an improvement in our classification accuracy. Our algorithm has a linear run time as it scans through the document once to come up with the prediction, giving us an accuracy of 68% overall. Lyricists with a relatively high percentage of lyrics misclassified as poetry tend to be recognized for their poetic style, such as Bob Dylan and Annie Lennox.

2. RELATED WORK

We do not know of any work on the classification of documents based on the adjective usage into lyrics, poetry or articles nor are we aware of any computational work which discerns poetic from non-poetic lyricists. Previous works have used adjectives for various purposes like sentiment analysis [1]. Furthermore in Music Information Retrieval, work on poetry has focused on poetry translator, automatic poetry generation.

Chesley et al. [1] classifies blog posts according to sentiment using verb classes and adjective polarity, achieving accuracy levels of 72.4% on objective posts, 84.2% for positive posts, and 80.3% for negative posts. Entwisle et al. [2] analyzes the free verbal productions of ninth-grade males and females and conclude that girls use more adjectives than boys but fail to reveal differential use of qualifiers by social class.

Smith et al. [13] use of tf-idf weighting to find typical phrases and rhyme pairs in song lyrics and conclude that the typical number one hits, on average, are more clichéd. Nichols et al. [14] studies the relationship between lyrics and melody on a large symbolic database of popular music and conclude that songwriters tend to align salient notes with salient lyrics.

There is some existing work on automatic generation of synonyms. Zhou et al. [3] extracts synonyms using three sources - a monolingual dictionary, a bilingual corpus and a monolingual corpus, and use a weighted ensemble to combine the synonyms produced from the three sources. They get improved results when compared to the manually built thesauri, WordNet and Roget.

Christian et al. [4] describe an approach for using Wikipedia to automatically build a dictionary of named entities and their synonyms. They were able to extract a large amount of entities with a high precision, and the synonyms found were mostly relevant, but in some cases the number of synonyms was very high. Niemi et al. [5] add new synonyms to the existing synsets of the Finnish WordNet using Wikipedia's links between the articles of the same topic in Finnish and English.

As to computational poetry, Jiang et al. [6] use statistical machine translation to generate Chinese couplets while Genzel et al. [7] use statistical machine translation to translate poetry keeping the rhyme and meter constraints.

3. DATA SET

The training set consists of articles, lyrics and poetry and is used to calculate the probability distribution of adjectives in the three different types of documents. We use these probability distributions in our document classification algorithms, to identify poetic from non-poetic lyricists and to determine adjectives more likely to be used in lyrics rather than poetry and vice versa.

3.1 Articles

We take the English Wikipedia and over 13,000 news articles from four major newspapers as our article data set. Wikipedia, an enormous and freely available data set is edited by experts. Both of these are extremely rich sources of data on many topics. To remove the influence of the presence of articles about poems and lyrics in Wikipedia we set the pruning threshold frequency of adjectives to a high value, and we ensured that the articles were not about poetry or music.

3.2 Lyrics

We took more than 10,000 lyrics from 56 very popular English singers. Both the authors listen to English music and hence it was easy to come up with a list which included singers from many popular genres with diverse backgrounds. We focus on English-language popular music in our study, because it is the closest to "universally" popular music, due to the strength of the music industry in English-speaking countries. We do not know if our work would generalize to non-English Language songs. Our data set includes lyrics from the US, Canada, UK and Ireland.

3.3 Poetry

We took more than 20,000 poems from more than 60 famous poets, like Robert Frost, William Blake and John Keats, over the last three hundred years. We selected the top poets from Poem Hunter [19]. We selected a wide time range for the poets, as many of the most famous English poets are from that time period. None of the poetry selected were translations from another language. Most of the poets in our dataset are poets from North America and Europe. We believe that our training data, is representative of the mean, as a majority of poetry and poetic style are inspired by the work of these few extremely famous poets.

3.4 Test Data

For the purpose of document classification we took 100 from each category, ensuring that they were not present in the training set. While collecting the test data we ensured the diversity, the lyrics and poets came from different genres and artists and the articles covered different topics and were selected from different newspapers.

To determine poetic lyricists from non-poetic ones we took eight of each of the two types of lyricists, none of whom were present in our lyrics data sets. We ensured that the poetic lyricists we selected were indeed poetic by looking up popular news articles or ensuring that they were poet along with being lyricists. Our list for poetic lyricists included Bob Dylan and Annie Lennox etc. while the non-poetic ones included Bryan Adams and Michael Jackson.

4. METHOD

These are the main steps in our method:

1) Finding the synonyms of all the words in the training data set.
2) Finding the probability distribution of word for all the three types of documents.
3) The document classification algorithm.

4.1 Extracting Synonyms

We extract the synonyms for a term from three sources: WordNet, Wikipedia and an online thesaurus.

WordNet is a large lexical database of English where words are grouped into sets of cognitive synonyms (synsets) together based on their meanings. WordNet interlinks not just word forms but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. The synonyms returned by WordNet need some pruning.

We use **Wikipedia** redirects to discover terms that are mostly synonymous. It returns a large number of words, which might not be synonyms, so we need to prune the results. This method has been widely used for obtaining the synonyms of named entities *e.g.* [4], but we get decent results for adjectives too.

We also used an **online Thesaurus** that lists words grouped together according to similarity of meaning. Though it gives very accurate synonyms, pruning is necessary to get better results.

We prune synonyms obtained from the three sources, which fall below an experimentally determined threshold for the semantic distance between the synonyms and the word. To calculate the semantic similarity distance between words we use the method described by Pirro et al. [8]. Extracting synonyms for a given word is an open problem and with improvement in this area our algorithm will achieve better classification accuracy levels.

4.2 Probability Distribution

We believe that the choice of an adjective to express a given concept depends on the genre of writing: adjectives used in lyrics will be different from ones used in poems or in articles. We calculate the probability of a specific adjective for each of the three document types.

First, WordNet is used to identify the adjectives in our training sets. For each adjective we compute the frequency of that were in the training set and the frequency of it and its synonyms; the ratio of these is the frequency with which that adjective represents its synonym group in that class of writing.

We exclude adjectives that occur infrequently (fewer than 5 times in our lyrics/poetry set or 50 in articles). The enormous size of the Wikipedia justifies the high threshold value.

4.3 Document classification algorithm

We use a simple linear time algorithm which takes as input the probability distributions for adjectives, calculated above, and the document(s) to be classified, calculates the score of the document being an article, lyrics or poetry, and labels it with the class with the highest score. The algorithm takes a single pass along the whole document and identifies adjectives using WordNet.

For each word in the document we check its presence in our word list. If found, we add the probability to the score, with a special penalty of -1 for adjectives never found in the training set and a special bonus of +1 for words with probability 1. The penalty and boosting values used in the algorithm were determined experimentally. Surprisingly, this simple approach gives us much better accuracy rates than Naïve Bayes, which we thought would be a good option since it is widely used in classification tasks like spam filtering. We have decent accuracy rates with this simple, naïve algorithm; one future task could be to come up with a better classifier.

5. RESULTS

First, we look at the classification accuracies between lyrics, articles and poems obtained by our classifier. We show that the adjectives used in lyrics are much more rhymable than the ones used in poems but they do not differ significantly in their semantic orientations. Furthermore, our algorithm is able to identify poetic lyricists from non-poetic ones using the word distributions, calculated in earlier section. We also compare adjectives for a given concepts which are more likely to be used in lyrics rather than poetry and vice versa.

5.1 Document Classification

Our test set consists of the text of 100 each of our three categories. Using our algorithm with the adjective distributions we get an accuracy of 67% for lyrics, 80% for articles and 57% for poems.

The confusion matrix, Table 1 we find the best accuracy for articles. This might be because of the enormous size of the article training set which consisted of all English Wikipedia articles. A slightly more number of articles get misclassified as lyrics than poetry.

Surprisingly, a large number of misclassified poems get classified as articles rather than poetry, but most misclassified lyrics get classified as poems.

5.2 Adjective Usage in Lyrics versus Poems

Poetry is written against a silent background while lyrics are often written keeping the melody, rhythm, instrumentation, the quality of the singer's voice and other qualities of the recording in mind. Furthermore, unlike most poetry, lyrics include repeated lines. This led us to believe the adjectives which were more likely to be used in lyrics rather than poetry would be more rhymable.

We counted the number of words an adjective in our lyrics and poetry list rhymes with from the website rhymezone.com. The values are tabulated in Table 2.

From the values in Table 2, we can clearly see that the adjectives which are more likely to be used in lyrics to be much more rhymable than the adjectives which are more likely to be used in poetry.

	Predicted		
Actual	Lyrics	Articles	Poems
Lyrics	67	11	22
Articles	11	80	6
Poems	10	33	57

Table 1. The confusion matrix for document classification. Many lyrics are categorized as poems, and many poems as articles.

	Lyrics	Poetry
Mean	33.2	22.9
Median	11	5
25^{th} percentile	2	0
75^{th} percentile	38	24

Table 2. Statistical values for the number of words an adjective rhymes with.

	Lyrics	Poetry
Mean	-.05	-.053
Median	0.0	0.0
25^{th} percentile	-0.27	-0.27
75^{th} percentile	0.13	0.13

Table 3. Statistical values for the semantic orientation of adjectives used in lyrics and poetry.

We were also interested in finding if the adjectives used in lyrics and poetry differed significantly in their semantic orientations. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity. We calculated the semantic orientations, which take a value between -1 and +1, using SentiWordNet, of all the adjectives in the lyrics and poetry list, the values are in Table 3. They show no difference between adjectives in poetry and those in lyrics.

5.3 Poetic vs non-Poetic Lyricists

There are lyricists like Bob Dylan [15], Ani DiFranco [16], and Stephen Sondheim [17,18], whose lyrics are considered to be poetic, or indeed, who are published poets in some cases. The lyrics of such poetic lyricists possibly could be structurally more constrained than a majority of the lyrics or might adhere to a particular meter and style. While selecting the poetic lyricists we ensured that popular articles supported our claim or by going to their Wikipedia page and ensuring that they were poets along with being lyricists and hence the influence of their poetry on lyrics.

Our algorithm consistently misclassifies a large fraction of the lyrics of such poetic lyricists as poetry while the percentage of misclassified lyrics as poetry for the non-poetic lyricists is significantly much less. These values for poetic and non-poetic lyricists are tabulated in table 4 and table 5 respectively.

Poetic Lyricists	% of lyrics misclassified as poetry
Bob Dylan	42%
Ed Sheeran	50%
Ani Di Franco	29%
Annie Lennox	32%
Bill Callahan	34%
Bruce Springsteen	29%
Stephen Sondheim	40%
Morrissey	29%
Average misclassification rate	**36%**

Table 4. Percentage of misclassified lyrics as poetry for poetic lyricists.

Non-Poetic Lyricists	% of lyrics misclassified as poetry
Bryan Adams	14%
Michael Jackson	22%
Drake	7%
Backstreet Boys	23%
Radiohead	26%
Stevie Wonder	17%
Led Zeppelin	8%
Kesha	18%
Average misclassification rate	**17%**

Table 5. Percentage of misclassified lyrics as poetry for non-poetic lyricists.

From the values in table 4 and 5 we see that there is a clear separation between the misclassification rate between poetic and non-poetic lyricists. The maximum misclassification rate for the non-poetic lyricists i.e. 26% is less than the minimum mis-classification rate for poetic lyricists i.e. 29%. Furthermore the difference in average misclassification rate between the two groups of lyricists is 19%. Hence our simple algorithm can accurately identify poetic lyricists from non-poetic ones, based only on adjective usage.

5.4 Concept representation in Lyrics vs Poetry

We compare adjective uses for common concepts. To represent physical beauty we are more likely to use words like “sexy” and “hot” in lyrics but “gorgeous” and “handsome” in poetry. For 20 of these, results are tabulated in Table 6. The difference could possibly be because unlike lyrics, which are written for the masses, poetry is generally written for people who are interested in literature. It has been shown that the typical number one hits, on average, are more clichéd [13].

Lyrics	Poetry
proud, arrogant, cocky	haughty, imperious
sexy, hot, beautiful, cute	gorgeous, handsome
merry, ecstatic, elated	happy, blissful, joyous
heartbroken, brokenhearted	sad, sorrowful, dismal
real	genuine
smart	wise, intelligent
bad, shady	lousy, immoral, dishonest
mad, outrageous	wrathful, furious
royal	noble, aristocratic, regal
pissed	angry, bitter
greedy	selfish
cheesy	poor, worthless
lethal, dangerous, fatal	mortal, harmful, destructive
afraid, nervous	frightened, cowardly, timid
jealous	envious, covetous
lax, sloppy	lenient, indifferent
weak, fragile	feeble, powerless
black	ebon
naïve, ignorant	innocent, guileless, callow
corny	dull, stale

Table 6. For twenty different concepts, we compare adjectives which are more likely to be used in lyrics rather than poetry and vice versa.

6. APPLICATIONS

The algorithm developed has many practical applications in Music Information Retrieval (MIR). They could be used for automatic poetry/lyrics generation to identify adjectives more likely to be used in a particular type of document. As we have shown we can analyze documents, analyze how lyrical, poetic or article-like a document is. For lyricists or poets we can come up with alternate better adjectives to make a document fit its genre better. Using the word distributions we can come up with a better measure of distance between documents where the weights are assigned to a word depending on its probability of usage in a particular type of document. And, of course, our work here can be extended to different genres of writings like prose or fiction.

7. CONCLUSION

Our key finding is that the choice of synonym for even a small number of adjectives are sufficient to reliably identify genre of documents. In accordance with our hypothesis, we show that there exist differences in the kind of adjectives used in different genres of writing. We calculate the probability distribution of adjectives over the three kinds of documents and using this distribution and a simple algorithm we are able to distinguish among lyrics, poetry and article with an accuracy of 67%, 57% and 80% respectively.

Adjectives likely to be used in lyrics are more rhymable than the ones used in poetry. This might be because lyrics are written keeping in mind the melody, rhythm, instrumentation, quality of the singer's voice and other qualities of the recording while poetry is without such concerns. There is no significant difference in the semantic orientation of adjectives which are more likely to be used in lyrics and those which are more likely to be used in poetry. Using the probability distributions, obtained from training data, we present adjectives more likely to be used in lyrics rather than poetry and vice versa for twenty common concepts.

Using the probability distributions and our algorithm we show that we can discern poetic lyricists from non-poetic ones. Our algorithm consistently misclassifies a majority of the lyrics of such poetic lyricists as poetry while the percentage of misclassified lyrics as poetry for the non-poetic lyricists is significantly much less.

Calculating the probability distribution of adjectives over the various document types is a vital step in our method which in turn depends on the synonyms extracted for an adjective. Synonym extraction is still an open problem and with improvements in it our algorithm will give better accuracy levels. We extract synonyms from three different sources – Wikipeia, WordNet and an online Thesaurus, and prune the results based on the semantic similarity between the adjectives and the obtained synonyms.

We use a simple naïve algorithm, which gives us better result than Naïve Bayes. An extension to the work can be coming up with an improved version of the algorithm with better accuracy levels. Future works can use a larger dataset for lyrics and poetry (we have an enormous dataset for articles) to come up with better probability distribution for the two document types or to identify parts of speech that effectively separates genres of writing. Our work here can be extended to different genres of writings like prose, fiction etc. to analyze the adjective usage in those writings. It would be interesting to do similar work for verbs and discern if different words, representing the same action, are used in different genres of writings.

8. ACKNOWLEDGEMENTS

Our research is supported by a grant from the Natural Sciences and Engineering Research Council of Canada to DGB.

9. REFERENCES

[1] P. Chesley, B. Vincent., L. Xu, and R. Srihari, "Using Verbs and Adjectives to Automatically Classify Blog Sentiment", *Training*, volume 580, number 263, pages 233, 2006.

[2] D.R. Entwisle and C. Garvey, "Verbal productivity and adjective usage", *Language and Speech*, volume 15, number 3, pages 288-298, 1972.

[3] H. Wu and M. Zhou, "Optimizing Synonym Extraction Using Monolingual and Bilingual Resources", in *Proceedings of the 2nd International Workshop on Paraphrasing*, volume 16, pages 72-79, 2003.

[4] C. Bohn and K. Norvag, "Extracting Named Entities and Synonyms from Wikipedia", in *Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications, (AINA* '10), pages 1300–1307.

[5] J. Niemi, K. Linden and M. Hyvarinen, "Using a bilingual resource to add synonyms to a wordnet:FinnWordNet and Wikipedia as an example", in *Proceedings of the Global WordNet Association* , pages 227–231, 2012.

[6] L. Jiang and M. Zhou, "Generating Chinese couplets using a statistical MT approach", in *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 377–384, 2008.

[7] D. Genzel, J. Uszkoreit and F. Och, "Poetic statistical machine translation: rhyme and meter", in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 158–166, 2010.

[8] G. Pirro and J. Euzenat, "A Feature and Information Theoretic Framework for Semantic Similarity and Relatedness", in *Proceedings of the 9th International Semantic Web Conference (ISWC '10)*, pages 615-630, 2010.

[9] G. Miller, "WordNet: A Lexical Database for English", *Communications of the ACM*, volume 38, number 11, pages 39-41, 1995.

[10] G. Miller and F. Christiane, *WordNet: An Electronic Lexical Database*, 1998.

[11] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", in *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC '10)* , pages 2200–2204, 2010.

[12] A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining", in *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC '06)*, pages 417–422, 2006.

[13] A.G. Smith, C. X. S. Zee and A. L. Uitdenbogerd, "In your eyes: Identifying cliché in song lyrics", in *Proceedings of the Australasian Language Technology Association Workshop*, pages 88–96, 2012.

[14] E. Nichols, D. Morris, S. Basu, S. Christopher, "Relationships between lyrics and melody in popular music", in *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR '09)*, 2009.

[15] K. Negus, *Bob Dylan*, Equinox London, 2008.

[16] A. DiFranco, *Verses*, Seven Stories, 2007.

[17] S. Sondheim, *Look, I Made a Hat!* New York: Knopf, 2011.

[18] S. Sondheim, *Finishing the Hat*, New York: Knopf, 2010.

[19] http://www.poemhunter.com.