

Computational Analysis of the Historical Changes in Poetry and Prose

Amitha Gopidi: amithareddygopidi@gmail.com
Aniket Alam: aniket.alam@iiit.ac.in



INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY

HYDERABAD

Poetry, a rapidly changing art

- Fine arts developed, and their types and uses defined, in times very different from ours.
- The esoteric definitions of poetry are insufficient in enveloping the changes in poetry with little verifiability
- They are also insufficient in distinguishing between poetry and prose

“The general term poetry, for example, now encompasses so many diverse and often irreconcilable artistic enterprises that it often proves insufficient to distinguish the critical issues at stake.”

- Michael Dana Gioia, Poetry At The End of Print Culture

Problem Statement

A quest to understand what has changed in poetry with respect to prose over the last 150 years within ‘the age of mechanical reproduction’.

Comparisons between poetry and prose in the early and late stages of this age.

1. Early stage, 1870-1920, wherein romantic poetry flourished
2. Late stage, 1970-2019, which saw art occurring through various digital media forms.

Experiments

1. Prose vs Poetry of Each Period
2. Poetry: 1870-1920 vs 1970-2019
3. Prose: 1870-1920 vs 1970-2019
4. Poetry vs Prose Both Periods Combined

Poetry and Prose - Differentiating features?

1. **Semantic features** commonly attributed to poetry:



Imagery, metaphors, sentiment, choice of words, themes, topics and associations

The notion of 'poetic' in prose as corroborated by Terry Eagleton. Toni Morrison, for instance, is called a highly 'poetic' writer.

Poetry and Prose - Differentiating features?

2. **Visual difference** - Line breaks



"A poem is a fictional, verbally inventive moral statement in which it is the author, rather than the printer or word processor, who decides where the lines should end." - Terry Eagleton

Can a quote be converted to a poem by an individual's decisions as to where to split the sentences into new lines?

Poetry and Prose – Differentiating Features?

3. Grammar : 'Poetic license'

Poetry tends to break free from the limitations of grammar. **'Poetic license'**.

- Emily Dickinson
- E.E. Cummings
- Rupi Kaur

Feature	Example
Adjective Inversion	"I sing the body electric"
Subject Verb Inversion	"Ten thousand saw I at a glance"
Prepositional Phrase Inversion	"Until we meet again, to be counted as bliss"
Yoda Construction	"Whose woods these are I think I know"
Dependent Clause	"What man cannot imagine, he cannot create"
Question	"Shall I compare thee to a summer's day?"
Beginning with a conjunction	<p>"Two roads diverged in a yellow wood, And sorry I could not travel both And be one traveller, long I stood And looked down one as far as I could To where it bent in the undergrowth"</p>

Poetry and Prose – Differentiating Features?

4. Meter

- Meter is a subsection of rhythm and consists of the most identifiable rhythms.
- Meter defined as a the arrangement of stressed(s) and unstressed(w) syllables in a certain manner.
- Poets use inversion to fit their material into a meter

5. Rhyme (Only for poetry)

Quotes as Prose

Quotes from popular novels chosen

- They possess all semantic features that may be attributed to poetry
- They're the touchstones of books
- Contextually independent of the situation in the book
- Make sense in a stand alone manner

Curation of Data

Poetry:

- PoetryFoundation and PoemHunter websites
- List of popular poets of that time period manually chosen
- Their works were collated in the form of pdf files, which were converted computationally.

Prose:

Computationally created dataset of 30 top liked quotes (lesser if 30 weren't available) from 500 most popular books of the time segment as listed in Goodreads website.

Dataset Overview

Type	Time Period	Count
Prose	1870-1920	7838 quotes
Prose	1970-2019	12623 quotes
Poetry	1870-2019	13635 poems
Poetry	1970-2019	7917 poems

Implementation - Grammar

- Used Stanford CoreNLP tools of tokenization, POS tags, dependency parse trees and OpenIE triples.
- Simple heuristics were used to decide which kind of inversions exist in a given sentence using POS tags and OpenIE triples in the sentence.
 - For instance, if the subject given by OpenIE triple of a line is a noun or pronoun, and it is preceded by a verb, the line would be marked as having subject-verb inversion.
 - The OpenIE tool trained on prose, doesn't always fetch results for lines in poetry and in these cases, we use POS tags as they are accurate for poetry as well.
- The various grammatical feature counts were normalized by the number of lines in poetry datasets and the number of sentences in prose datasets, so as to remove dependency on the length of the poem or quote.

Implementation - Meter

- In order to implement meter, we used Stanford Literary Lab's Poesy, a python module for poetic processing.
- The module gives information of a base meter among four types of base meters:
 - **Iambic [ws]**
 - **Trochaic [sw]**
 - **Anapestic [wws]**
 - **Dactylic [sww]**
- Also gives information of number of repetitions of meter in a given line, such as pentameter, hexameter etc.

Implementation - Rhyme

- Inversion also used to fit into a rhyme scheme along with meter.
- Used the Poesy module.
- This feature was only used for comparison between the two poetry datasets.
- Not been applied to prose datasets as the values were null.

Classifications

Feature vectors consisted of 9 features discussed in the previous sections.

- 7 are grammar related
- Base meter
- Number of feet
- Rhyme Type (only used for classification between the two poetry datasets)

Grammar feature counts normalized by length of the poem/quote.

Classifications

- Feature data trained through a Random Forest Classifier and KNN classifier
- 70:30 split for the training and testing data

- Optimal value of the number of trees for random forest classifier was found to be 100
- Value for k was taken as 3 for the KNN classifier

- Dealing with class imbalance:
 - Adjusted weights inversely proportional to class frequencies in the data.
- Four experiments of different classifications between prose and poetry were conducted.

Results

Random Forest Classifier performed better than KNN Classifier in all of the below experiments

- Prose vs Poetry of Each Period
- Poetry: 1870-1920 vs 1970-2019
- Prose: 1870-1920 vs 1970-2019
- Poetry vs Prose Both Periods Combined

Poetry vs Prose classification results for 1870-1920

Classifier	Feature	Accuracy	F1 Score	ROC AUC Score
Random Forest	All	98.4	98.4	98.02
Random Forest	Meter	93.7	93.7	93.8
Random Forest	Inversion	91.8	91.8	91.3
kNN	All	97.4	97.4	97.02
kNN	Meter	93.6	93.6	93.8
kNN	Inversion	91.8	91.8	91.2

Poetry vs Prose classification results for 1970-2019

Classifier	Feature	Accuracy	F1 Score	ROC AUC Score
Random Forest	All	91.5	91.3	89.7
Random Forest	Meter	83.4	82.8	80.06
Random Forest	Inversion	85.1	84.5	82.06
kNN	All	90.0	89.8	88.2
kNN	Meter	80.6	79.2	75
kNN	Inversion	81.4	83.6	81.07

Poetry: 1870-1920 vs 1970-2019

- Also considered rhyme as a feature.
- Poetry has undergone a significant change with an accuracy of 77% in classification.

Classifier	Accuracy	F1 Score	ROC AUC Score
Random Forest	77.0	76.3	73.11
kNN	72.6	72.5	70

Prose: 1870-1920 vs 1970-2019

- Classification between prose datasets done to see if prose changing could have contributed to prose and poetry converging.
- Nothing much as changed in prose over the two periods because evaluation scores are close to a random guess (59%).

Classifier	Accuracy	F1 Score	ROC AUC Score
Random Forest	59.2	50.7	50
kNN	51.4	52.09	51.1

Poetry and Prose Both Periods Combined

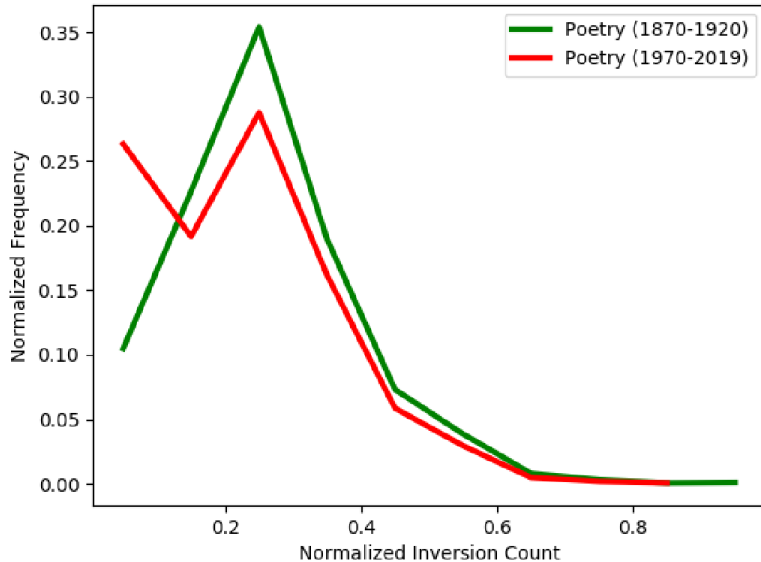
- This classification done with combined datasets of poetry against combined datasets of prose.
- Given an input, this classifier would differentiate between it being a poem or prose with 94.7% accuracy.
- This result is important, as we do not consider line breaks.

Classifier	Accuracy	F1 Score	ROC AUC Score
Random Forest	94.7	94.7	94.7
kNN	94.04	94.0	94.01

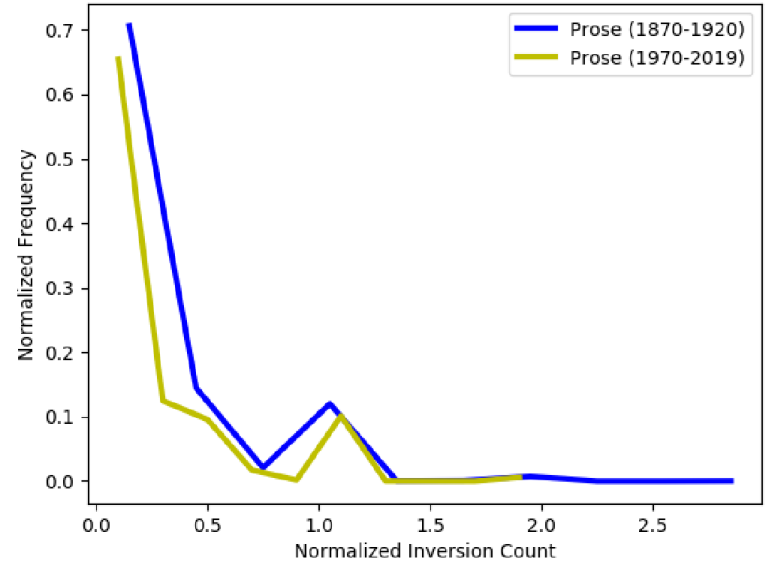
Analysis - Inversion

Figures 1 and 2 are plotted between the normalized inversion count(so as to remove any dependency on the length of the poem/prose), and the normalized frequency of the datasets(so as to remove dependency on the number of data points).

Poetry (1870-1920) vs Poetry (1970-2019) - Inversion Count



Prose (1870-1920) vs Prose (1970-2019) - Inversion Count



Analysis - Inversion

- Figure 1 indicates a fall in the inversion count in the second time period.
- Figure 2 shows that the inversion counts of the two periods of prose are more or less the same.

Analysis – Meter Base Type

The figures 3 and 4 represent the historical change in meter over the two periods for both prose and poetry. The y axis represents the percentage of dataset, which is a normalized indicator and does not skew the graph towards the period with higher data points

Figure 3: Meter Type - Poetry (1870-2019) vs Poetry (1970-2019)

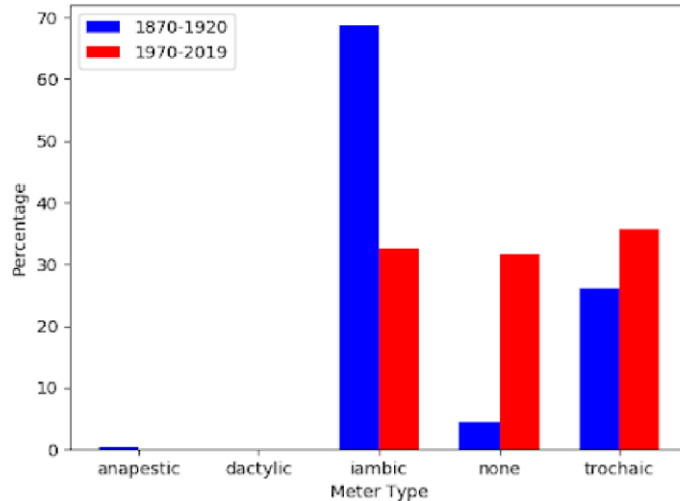
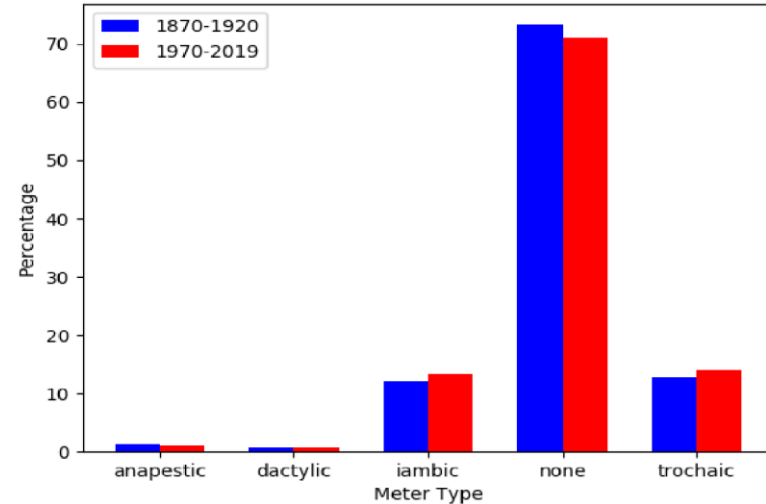


Figure 4: Meter Type – Prose (1870-1920) vs Prose (1970-2019)



Analysis – Meter Base Type

- Figure 3 clearly shows the dominance of the iambic base meter in poetry datasets, and its fall from 1870-1920 to 1970-2019. It also shows that in 1970-2019, the number of poems with no distinguishable meter has risen considerably with no significant change in anapestic or dactylic base meters.
- Figure 4 proves that there is no significant difference in the base meter of prose over the two chosen periods, with none value as the most dominant.

Popular Meters

- Figure 5 and Figure 6 were drawn so as to show the change in meter. Meter plotted is a combination of the base meter and its feet in the poem. The top 7 meters were chosen for the plots.
- Figure 5 shows the significant fall of some of the popular meters in the second time period as compared to the first. The increase in the none values also suggests that the second time period consists of poetry with no recognizable meter.

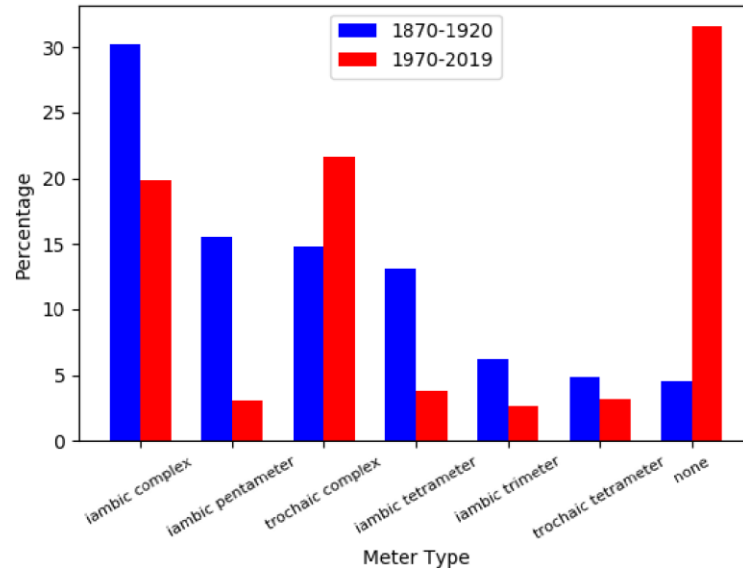


Figure 5:
Poetry

Popular Meters

Figure 6, on the other hand, shows the lack of recognizable meter in prose which is expected

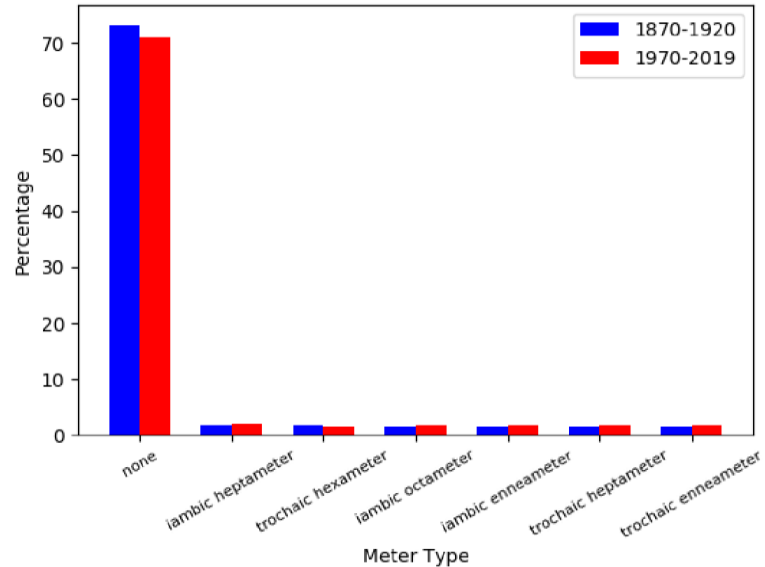
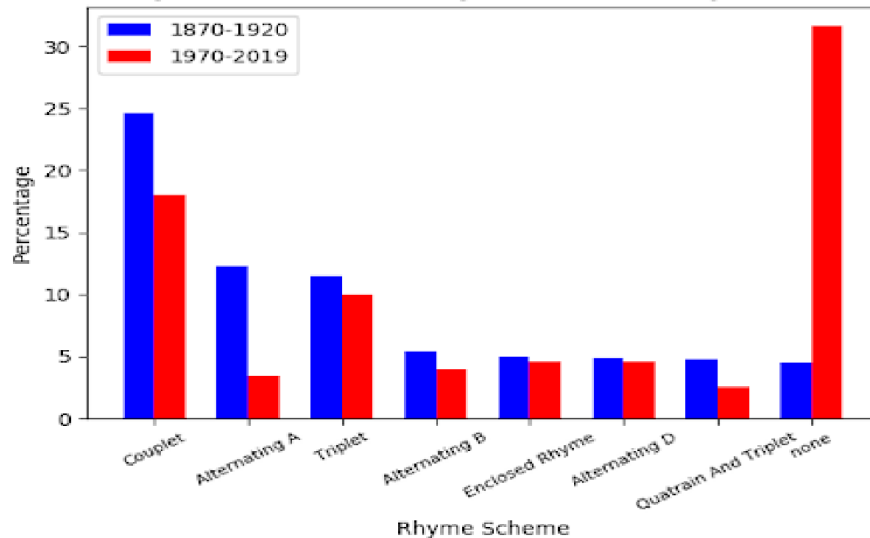


Figure 6: Prose

Rhyme

The rhyme feature used in figure 7, shows that a large percentage of 1970-2019 poetry has no rhyme scheme, while also showing that the prevalence of the other rhyme schemes has also come down.



Conclusion

- From the experiments conducted, it has been proved that the poetry of 1970-2019 is more similar to prose of its period than the poetry of 1870-1920 was to the prose of the same period.
- The changes in prose of the two periods with respect to stylistic features are minimal, but those in poetry are significant.
- The convergence of poetry and prose and lack of change in prose, proves that poetry does not possess the liminal boundaries that prose enjoys.
- The importance of a new age definition of poetry is thus established considering the changes in poetry as an artform.

Conclusion

Apart from justifying the historical changes in poetry and prose, this paper also achieves high accuracy in the classification of poem and prose using no semantic features.

- This is an important indicator that semantic content of poetry and prose can be very alike and that they can still be differentiated using stylistic features without considering the obvious visual difference of line breaks.

Future Work

The future work of this paper is to use these features in constructing a personalized poetry assistant that learns the stylistic preferences of the user in inversions and meter, based on user input of creative text.

Questions?

“Our fine arts were developed, their types and uses were established, in times very different from the present, by men whose power of action upon things was insignificant in comparison with ours. But the amazing growth of our techniques, the adaptability and precision they have attained, the ideas and habits they are creating, make it a certainty that profound changes are impending in the ancient craft of the Beautiful. In all the arts there is a physical component which can no longer be considered or treated as it used to be, which cannot remain unaffected by our modern knowledge and power. For the last twenty years neither matter nor space nor time has been what it was from time immemorial. We must expect great innovations to transform the entire technique of the arts, thereby affecting artistic invention itself and perhaps even bringing about an amazing change in our very notion of art.”

- Paul Valery