

Unfamiliar Analysis of Orhan Pamuk

Başak Ünsal
Handan Kulan

Table of Content

- ◉ Introduction & Motivation
- ◉ Problem Definition
- ◉ Literature Review
- ◉ Methodology
- ◉ Results
- ◉ References

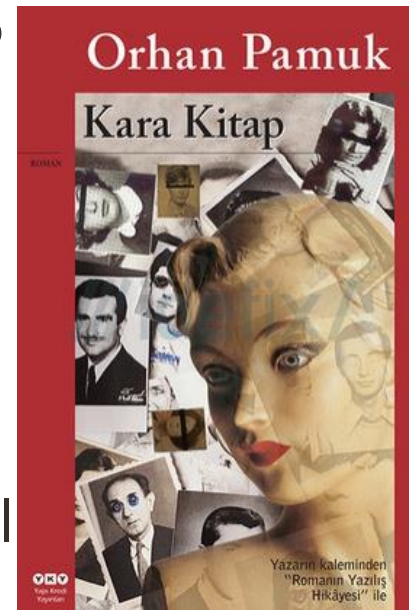
Introduction & Motivation

- Born in 1952
- Recipient of the 2006 Nobel Prize in Literature
- Kara Kitap written in 1990
(one of the most famous book)
- Understanding writing style of him



Problem Definition-Kara Kitap

- Two main characters : Celal and Galip
- Galip : husband of Rüya, brother of Celal
- Celal : columnist in Milliyet
- Rüya & Celal left
- Galip admired Celal
- Celal & Rüya murdered by fan of Celal
- Galip returns law career & continue writing



Problem Definition-Kara Kitap

- Two main section;
 - 19 subsection : Celal's
 - 17 subsection : Galip's
- Analysis of writing style of Orhan Pamuk
 - Using newspaper column data

Problem Definition-Kara Kitap

- Same tool applicable to English version
- It has two different translations by two different translators
- Does features in Turkish version remain same in translated version?

Problem Definition-Benim Adım Kırmızı

- Has multiple parts
- «Kara» tries to find murderer
- «Katil Diyecekler Bana» written by murderer
- «Ben...» written by suspicious characters
- According to writing styles, murderer will be defined.

Importance

- Identifying writing style of Orhan Pamuk by comparing
 - Multiple books
 - Different characters
 - Different authors
- Applicable to other author verification problems

Literature Review

- Support Vector Machine [1]
 - For n training sample : $O(n^2)$
- Bayesian Classification [2]
 - Learning network structure and conditional probabilities has exponential cost

Methodology

- ◉ Feature Selection
 - ◉ Length of sentences, length of words, part of speech tagging
- ◉ Distance and Similarity Calculation
 - ◉ Cosine, Dice Similarities
 - ◉ Jaccard Coefficient
 - ◉ Euclidian, Manhattan Distances

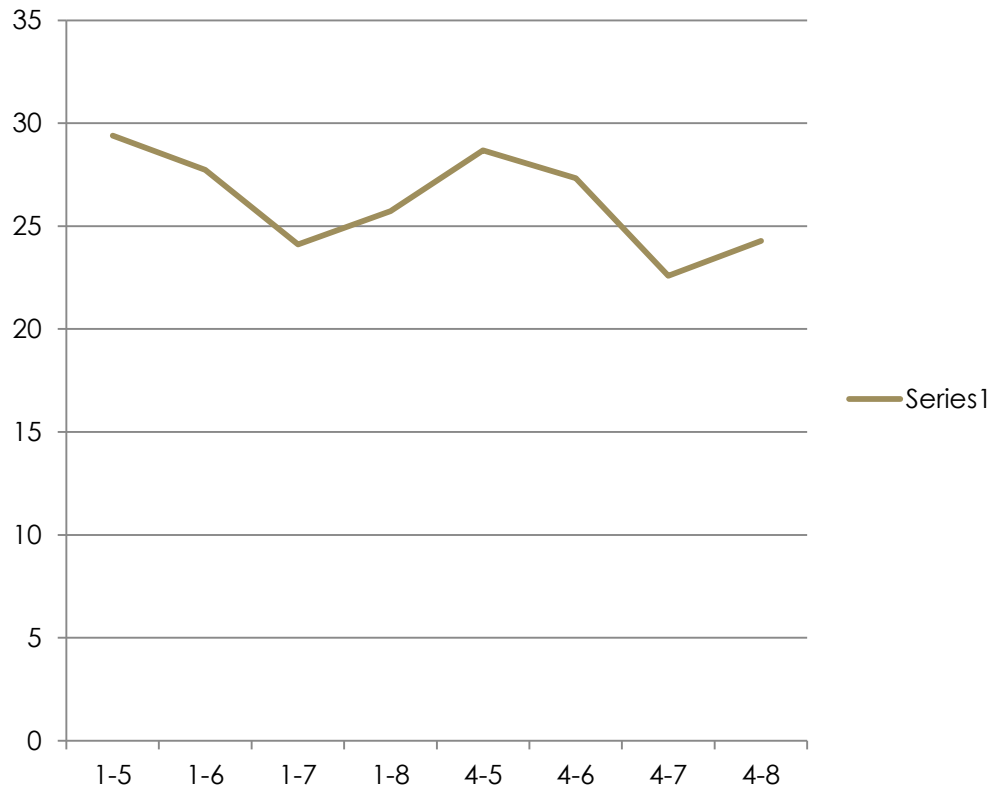
Methodology

- Used Rapid Miner and bag of words
- All data collected in .txt
- Data divided into smaller parts (subsampling)
- Stemming is applied (with java code and Zemberek [3])

Methodology

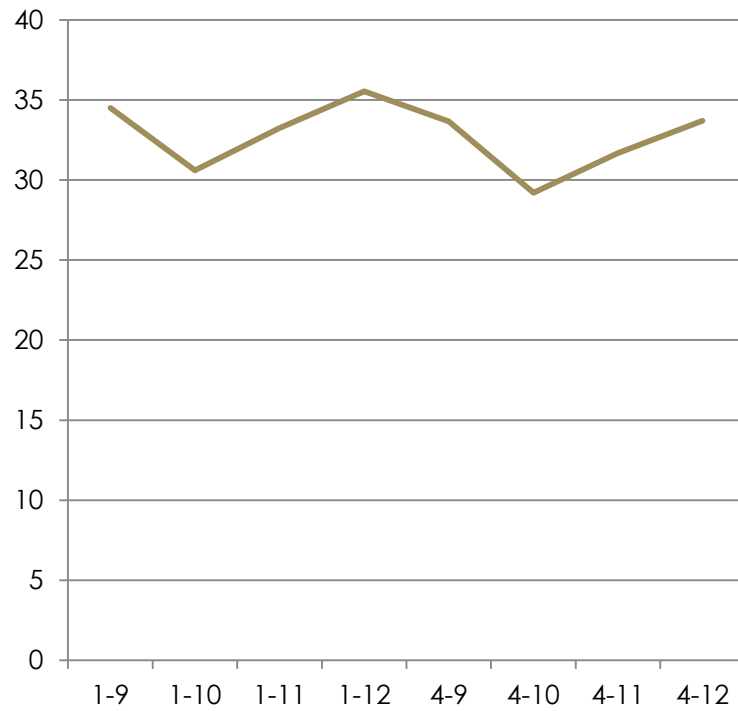
- Similarity Calculations
 - Cosine similarity : ineffective distribution
 - Dice similarity : low sensitivity
- Jaccard Coefficient
 - Low sensitivity (similar values)
- Distance Measures
 - Euclidian : small values
 - Manhattan : most suitable

Results – Kara Kitap



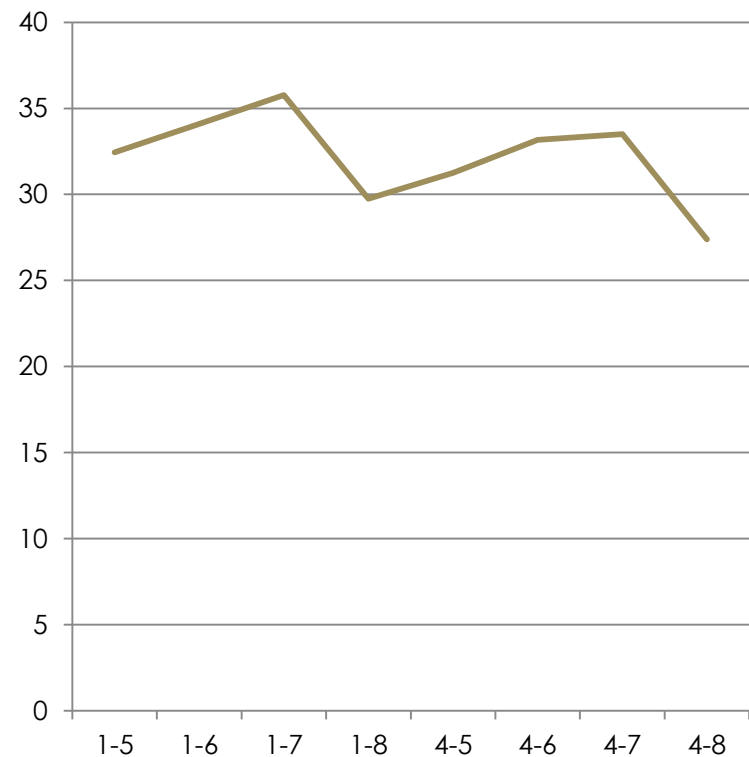
galip-celal	manhattan
1-5	29.3999124
1-6	27.7274519
1-7	24.11137018
1-8	25.72703303
4-5	28.67811298
4-6	27.32748886
4-7	22.59313429
4-8	24.27393961

Results - Kara Kitap



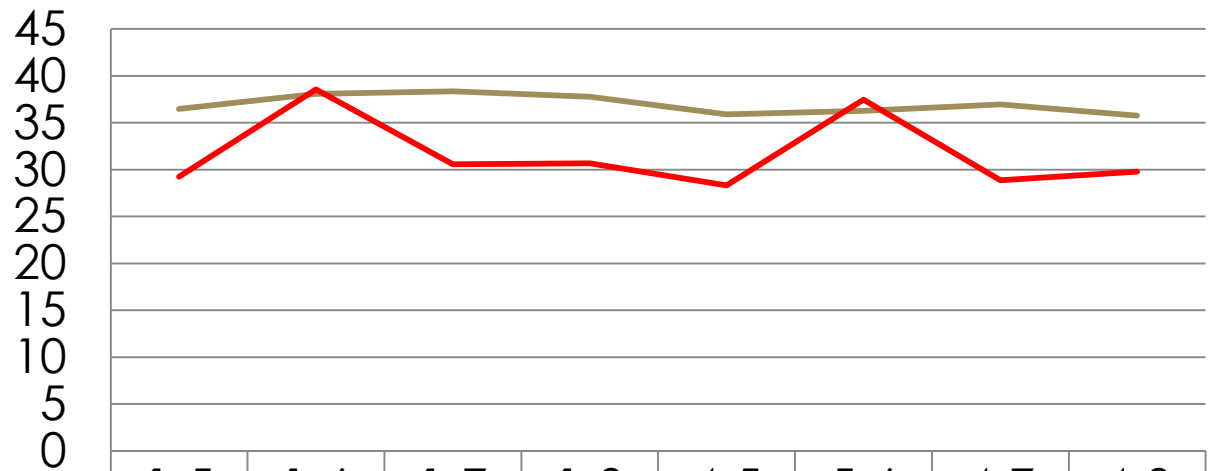
	manhattan
Orhan Pamuk - Celal	
1-9	34.5024028
1-10	30.5981067
1-11	33.230467
1-12	35.5340153
4-9	33.663248
4-10	29.2016482
4-11	31.687706
4-12	33.6984854

Results - Kara Kitap



Orhan Pamuk - Galip	
	manhattan
1-5	32.4509054
1-6	34.1008121
1-7	35.7661548
1-8	29.742384
4-5	31.2467704
4-6	33.1667873
4-7	33.5055025
4-8	27.3810673

Results – Orhan Pamuk



— Orhan Pamuk - Kaan Arslanoğlu	1-5	1-6	1-7	1-8	4-5	5-6	4-7	4-8
	36.47	38.06	38.34	37.76	35.9	36.27	36.93	35.76
— Orhan Pamuk - Oğuz Atay	1-5	1-6	1-7	1-8	4-5	5-6	4-7	4-8
	29.24	38.56	30.55	30.66	28.32	37.45	28.85	29.77

Future Works

- Other features will be applied
- English version will be analyzed
- Benim Adım Kırmızı will be reviewed

The End

- Thanks for your participation

References

- [1] - Koppel, Moshe, and Jonathan Schler.
"Authorship verification as a one-class classification problem." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.
- [2] - Zhao, Ying, Justin Zobel, and Phil Vines.
"Using relative entropy for authorship attribution." Information Retrieval Technology. Springer Berlin Heidelberg, 2006. 92-105.
- [3] - [zemberek.googlecode.com](https://github.com/zemberek)