



MEDNARODNA
PODIPLOMSKA ŠOLA
JOŽEFA STEFANA

JOŽEF STEFAN
INTERNATIONAL
POSTGRADUATE SCHOOL

Classification of Cross-cultural News Events

SiKDD 2021

Ljubljana, October 4th

Abdul Sittar, Dunja Mladenic

Outline

- Introduction
- Data description
- Methodology
- Dataset annotation and statistics
- Results
- Conclusions and future work

Introduction

- Collective programming of the mind which distinguishes the members of one category of people from another.
- Cultural values influence selection, analysis and news propagation.
- Culturally similar countries observe heavier news flow for similar events.

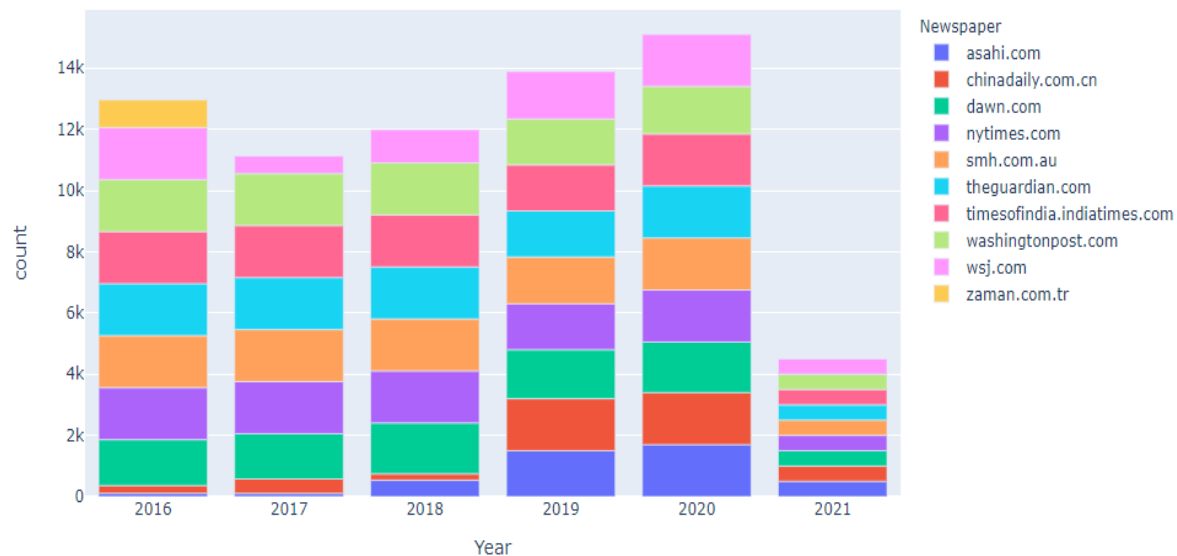
Contributions

- A novel perspective of aligning news events across different cultures through categorizing countries and news events.
- A cross-cultural automatically annotated dataset for several domains.
- Comparison of several classification models adopting different set of features (character ngrams, GLOVE embeddings, and word ngrams).



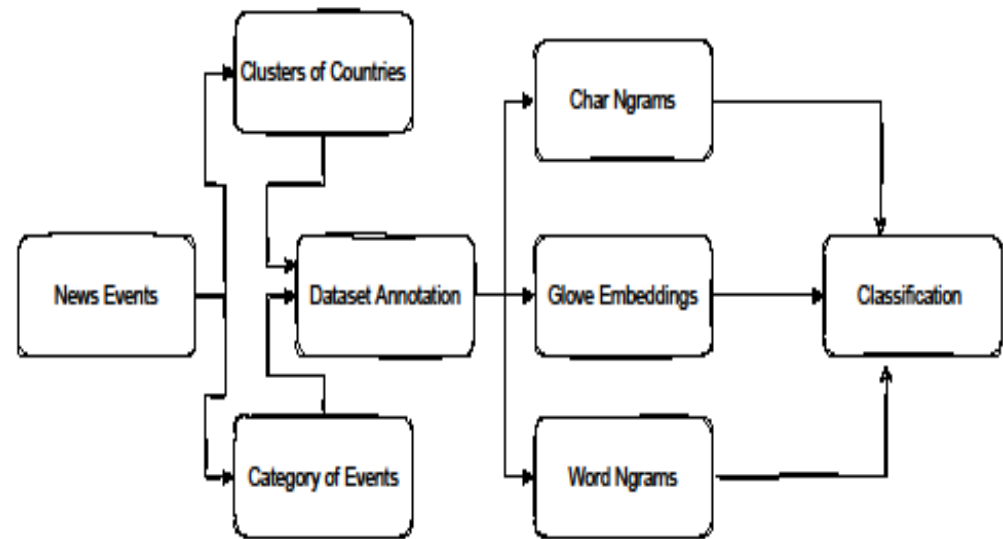
Data Description

- Approximately 8000 events
- Except Zaman.com, almost 1000 events



Methodology

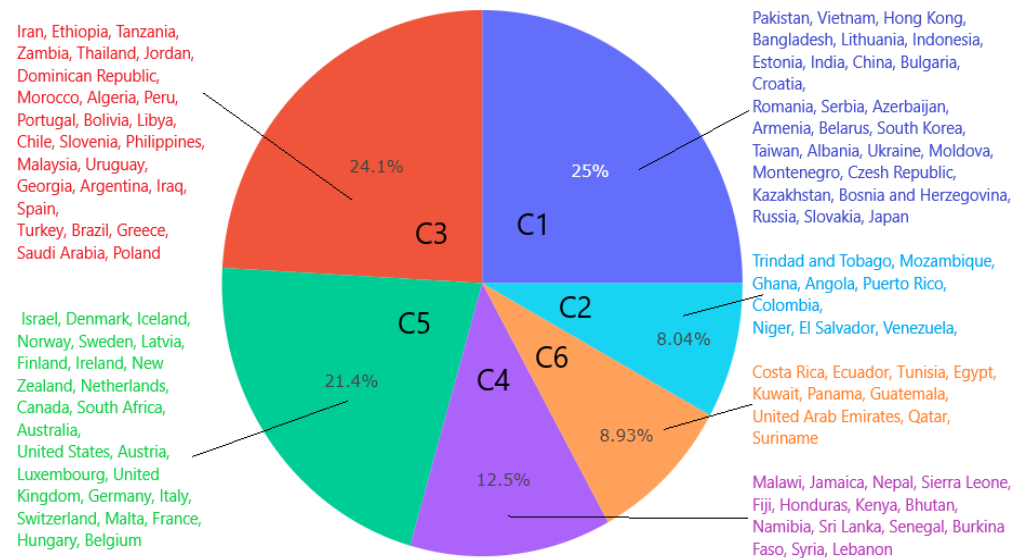
- Clusters of countries
- Clusters of events
- Annotation



- Char-ngrams
- Word-ngrams
- Glove Embeddings

Dataset Annotation

- Clusters of countries
- Hofstede's national cultural dimensions
- Power distance index
- Uncertainty avoidance by individuals
- Non-individualistic cultures
- Masculinity vs. Femininity
- Long-term orientation
- Indulgence vs. Restraint



Dataset Annotation

- Colour of category
- More categories in culture
C5



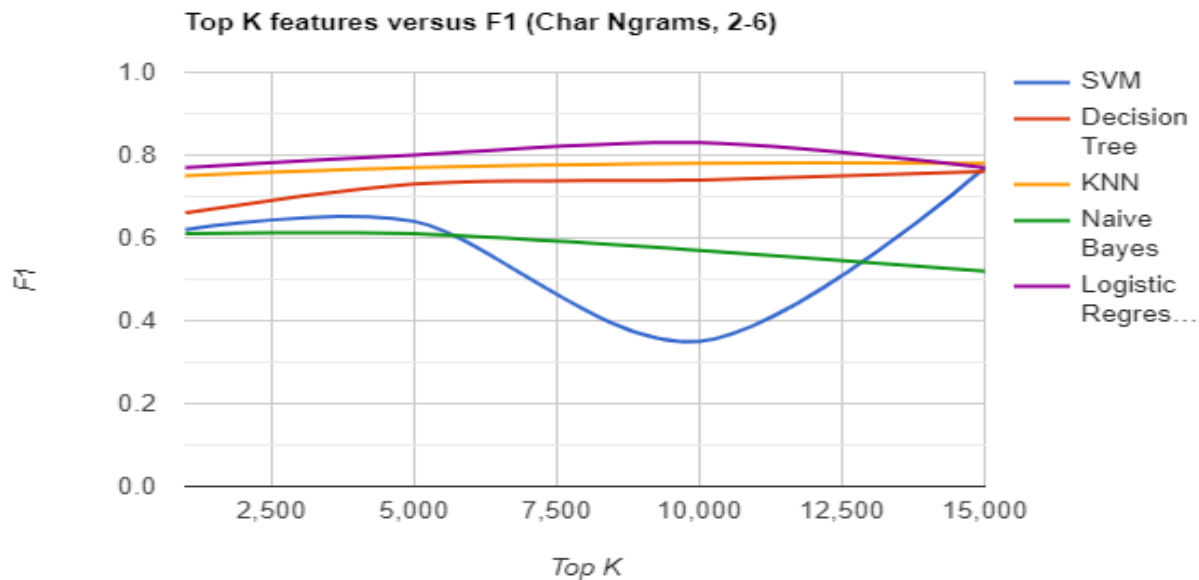
Dataset annotation

- Shared categories of events between the pair of six clusters
- 50% news events belongs to red and blue clusters
- Each cluster do not lies in a specific geographic area
- Purpose: observe cultural differences in same type of news events for two places



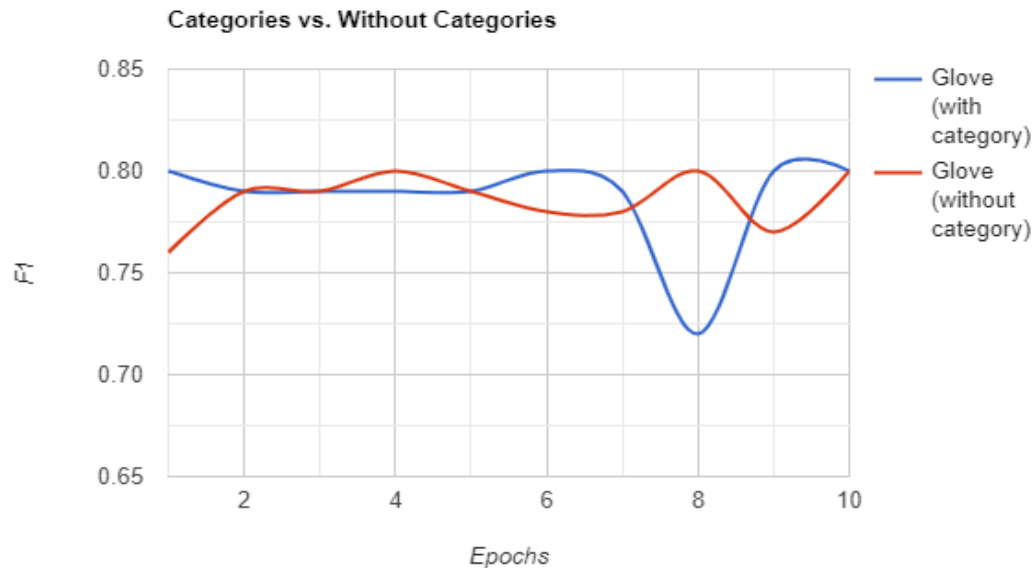
Results and Analysis

- Top 15K characters, using logistic Regression, 0.85



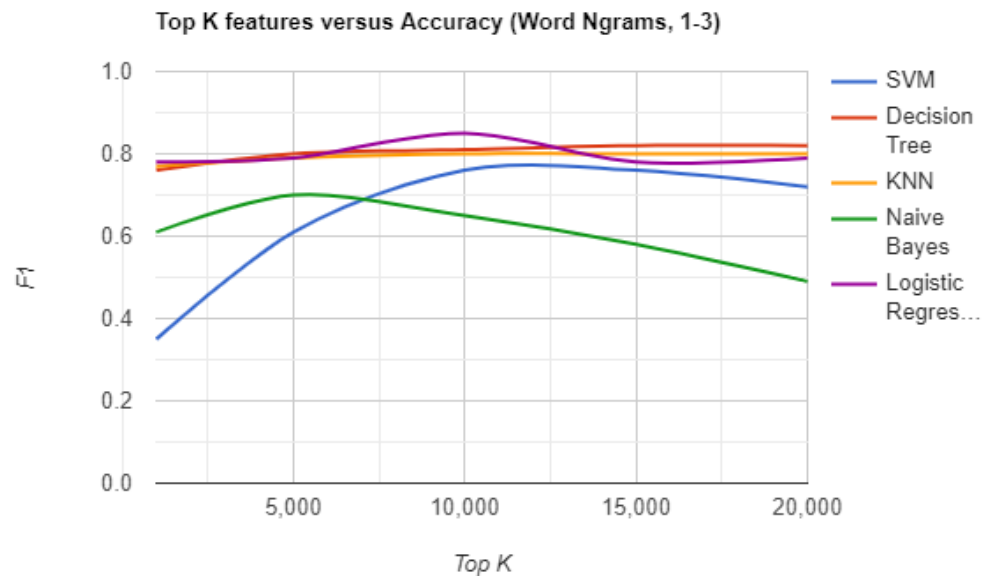
Results and Analysis

- With and without the category 0.80 and 0.79 f1 respectively



Results and Analysis

- Top 10K or 20K, 1 to 3 word ngrams



Conclusions and Future Work

- Helpful in examining the impact of specific socio-cultural factors on news events.
- Part of speech (POS) tagging, other embeddings.

<https://zenodo.org/record/5225053#.YUiQFZ0zZZc>

<https://dmoz-odp.org/>

