Resources and Experiments on Sentiment Classification for Georgian

Nicolas Stefanovitch[1], Jakub Piskorski[2], Sopho Kharazi[3]

[1]European Commission Joint Research Centre, Ispra, Italy

[2]Polish Academy of Sciences, Warsaw, Poland

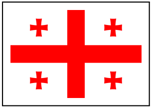
[3]Piksel SRL, Ispra, Italy

LREC 2022

Motivation

Georgian language (kat)

- language isolate
- $\sim$ 4 million speakers
- unicameral script (Geor)

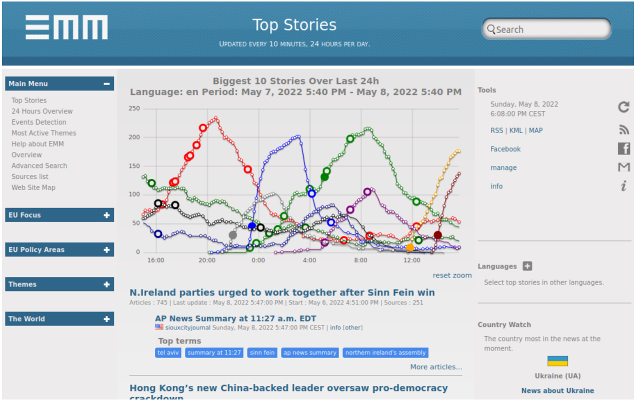


ყოველი ადამიანი იბადება თავისუფალი და
თანასწორი თავისი ღირსებითა და უფლებებით

under-resourced language!

Motivation

Europe Media Monitor (EMM)



https://emm.newsbrief.eu

Motivation

Sentiment analysis

NEGATIVE NEUTRAL POSITIVE

The former prime minister talks about Begashvili and says that he was the most sympathetic minister to him
$\rightarrow$ Positive

Motivation

Sentiment analysis

NEGATIVE NEUTRAL POSITIVE

ყოფილი პრემიერი ბეგაშვილზე საუბრობს და აცხადებს
რომ მისთივს ის ყველაზე სიმპატიური მინისტრი იყო

$\rightarrow$???

Content

New resources

- Tonality dictionary
- Annotated dataset: Georgian Sentiment Snippets

Experiments

- Different ML-based models: LR, SVM, Transformers
- Different settings: 3-class vs. 4-class
- Different training and test datasets
- Different classifiers: Georgian trained, transfer learning based, translation based
- Different dataset wrt. inter annotator agreement
- Perturbation of the dataset

Georgian Tonality Dictionary

Georgian Verbal Morphology:

PREVERB + **agreement prefix** + **version vowel** + ROOT + passive/causative suffix + thematic suffix + imperfect marker + tense vowel + agreement suffix + plural suffix

VB:შე+ყვარ (exclude:ყვარელ%, ყვარყვარ%)

$\Rightarrow$

შევაყვარ%; შევიყვარ%; შევუყვარ%; შევეყვარ%; შევყვარ%; შემაყვარ%; შემიყვარ%; შემუყვარ%; შემეყვარ%; შემყვარ%; შეგაყვარ%; შეგიყვარ%; შეგუყვარ%; შეგეყვარ%; შეგყვარ%; შეგვაყვარ%; შეგვიყვარ%; შეგვუყვარ%; შეგვეყვარ%; შეგვყვარ%; შეჰაყვარ%; შეჰიყვარ%; შეჰუყვარ%; შეჰეყვარ%; შეჰყვარ%; შესაყვარ%; შესიყვარ%; შესუყვარ%; შესეყვარ%; შესყვარ%; შეაყვარ%; შეიყვარ%; შეუყვარ%; შეეყვარ%; შეყვარ%; ვაყვარ%; ვიყვარ%; ვუყვარ%; ვეყვარ%; ვყვარ%; მაყვარ%; მიყვარ%; მუყვარ%; მეყვარ %; მყვარ%; გაყვარ%; გიყვარ%; გუყვარ%; გეყვარ%; გყვარ%; გვაყვარ%; გვიყვარ%; გვუყვარ%; გვეყვარ%; გვყვარ%; ჰაყვარ%; ჰიყვარ%; ჰუყვარ%; ჰეყვარ%; ჰყვარ%; საყვარ %; სიყვარ%; სუყვარ%; სეყვარ%; სყვარ%; აყვარ%; იყვარ%; უყვარ%; ეყვარ%; ყვარ%

	Very Positive	Positive	Negative	Very Negative	Total
Raw	84	721	831	350	1986
Expanded	342	4220	6989	2572	14123
Final	10630	32176	23869	3940	70615

Georgian Sentiment Snippets (GSS)

Negative	Neutral	Positive	Mixed	Total
1417	1734	765	307	4223
33.5%	41.0%	18.1%	7.2%	100%

- text snippets sampled from news articles gathered by EMM over 5 years
- average length of the text snippets is 114 characters
- 3 sampling strategies (random, polarising entities, tonality words)
- 12 native-speaker annotators, out of which 1 expert annotator
- annotation for objective and subjective sentiments (83.6% similarity)
- 15.7% of non-neutral snippets, do not contain dictionary words

Data Perturbation (methodology)

GOAL: reduce political bias and add variation

- randomly **change numerical/temporal expressions**
- **replace** the most frequent **person names** detected in the snippets, by randomly selecting a replacement from a pool of names computed over the whole data gathered before sampling snippets
- randomly **replace** a limited set of frequent **country names** with other country names
- **manual perturbation** if the above do not apply: changing adjectives, named-entities, replacing words with synonyms, etc.

RESULT: **56.7%** of the text snippets modified

Data Perturbation (effect on bias)

10-L: ten most common last names

100-F: hundred most common first names

Experiment	Negative	Neutral	Positive	JS div.
none	36.2%	44.3%	19.5%	0.0
10-L in Pert.	38.9%	42.9%	18.2%	4.1e-4
10-L in Orig.	33.0%	47.6%	19.4%	6.6e-4
100-F in Pert.	36.6%	47.4%	16.0%	1.1e-3
100-F in Orig.	35.2%	48.6%	16.2%	1.3e-3

- Jensen-Shannon divergence shows that perturbation reduces the bias associated with 10-L and 100-F
- overall positive bias towards 10-L: perturbations increase their proportion of *negative* labels by 5.9% and decrease the proportion of *positive* labels by 1.2%

Experiment: different approaches + different training and test datasets

Model: XLM-T (xlm-roberta-base pretrained additional data)

Georgian classifier:

- train and test on GSS

Transfer learning:

- train on UMSAB (Unified Multilingual Sentiment Analysis Benchmark)
- test on GSS, either on Georgian text or on English translation

Translation based:

- train and test on English translation of GSS

Experiment: 3-class classification

approach	macro F1
random guess	37.0 %
Lexicon based	56.1 %
LR (based on lexicon features)	57.9 %
SVM	66.3 %
Transformers (transfer learning on Georgian text)	40.7 %
Transformers (transfer learning on English translation)	67.5 %
Transformers (train on Georgian dataset)	75.2 %
Transformers (train on English translation of dataset)	76.8 %

Conclusions

→ about 10 points increase from Lexicon to SVM

→ about 10 points increase from SVM to transformers

→ transfer learning did not work well for Georgian

→ quality translation if available provides best results

? slower convergence on Georgian due to smaller vocabulary

Experiment: 4-class classification

approach	**macro** F_1
random guess	32.0 %
L2-LR	49.8 %
SVM	49.6 %
SVM (train on English translated Georgian dataset)	38.2 %
Transformers (train on Georgian dataset)	55.2 %

Conclusions

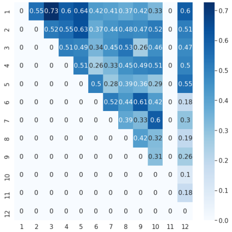
- $\rightarrow$ about 16 points decrease for SVM vs. 3-class settings
- $\rightarrow$ about 20 points decrease for transformers vs. 3-class settings
- $\rightarrow$ *mixed* class is almost never predicted, instead, for *mixed*-labelled snippets the dominant polarity gets predicted

Experiment: Influence of IAA

All: 11 annotators

Exp: 1 expert (last column)

Top5 : top 5 annotators



		Micro average			Macro average			Weighted average			
Experiment	Support	Prec.	Rec.	F_1	Prec.	Rec.	F_1	Prec.	Rec.	F_1	α
All+Exp	3916	75.9	75.9	75.9	75.8	75.5	75.6	76.1	75.9	75.9	0.543
Top5+Exp	3451	76.9	76.9	**76.9**	77.3	76.6	**76.8**	77.1	76.9	76.9	**0.622**
All	1818	71.2	71.2	71.2	42.5	46.1	43.7	61.7	71.2	65.5	0.461
Top5	1401	67.3	67.3	67.3	52.8	46.7	44.4	66.7	67.3	62.3	0.571

- α on a par with best value of [Mozetič et al., 2016] for similar settings
- Top5 correlate strongly with each other and with expert
- All+Exp has 13.4% more data than Top5+Exp, performs 1 point worse

Questions ?

The resources will be made publicly available at the time of the conference