Exploring Amharic Sentiment Analysis from Social Media Texts
Building Annotation Tools and Classification Models

Seid Muhie Yimam and Hizkiel Mitiku Alemayehu and Abinew Ali Ayele and Chris Biemann

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Language Technology

The 28th International Conference on Computational Linguistics COLING 2020, 8-13 December 2020

ABSTRACT

This work deals with sentiment analysis for Amharic social media texts. Social media users are ever-increasing, however, low-resource languages such as Amharic have received less attention due to 1) lack of **well-annotated datasets**, 2) unavailability of **computing resources**, and 3) fewer or no **expert researchers** in the area.

Research questions and approaches:

- Explore the suitability of existing tools for the sentiment analysis task. We build a social-network-friendly annotation tool called 'ASAB' using the Telegram bot and collect 9.4k tweets.
- Explore the suitability of machine learning approaches for Amharic sentiment analysis. The FLAIR deep learning text classifier, based on network embeddings that are computed from a distributional thesaurus, outperforms other supervised classifiers.
- Investigate the challenges in building a sentiment analysis system for Amharic. We found that the widespread usage of sarcasm and figurative speech are the main issues in dealing with the problem.

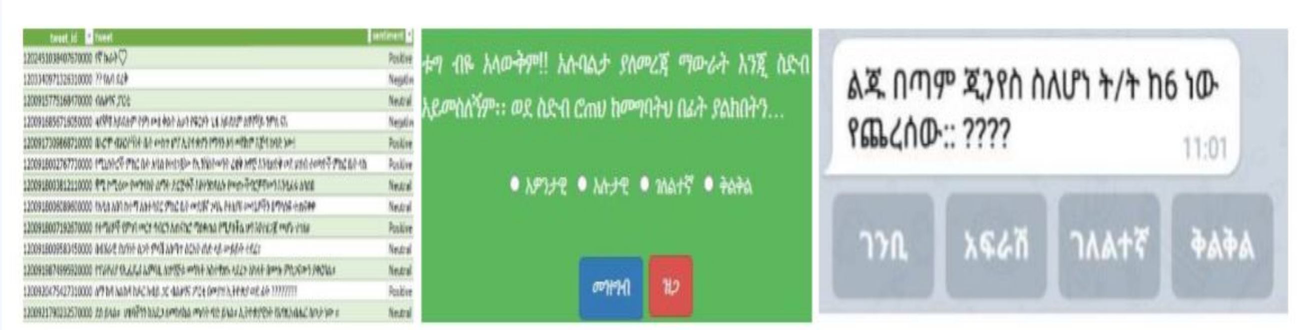
MOTIVATION

- Social media data is increasing but:
  - Lack of standard datasets.
  - Lack of basic NLP tools.
  - Lack of annotation tools and platforms.
  - Complex nature of Amharic .
- Hence, we need to build an **annotation tool**, annotate the data, and build models, enable applications to capture opinions from a social media text.

OBJECTIVE

- Exploring different annotation strategies and tools for low-resource languages.
- Annotating a large dataset.
- Build different machine learning models.

DATA COLLECTION TOOLS



(a) Excel-sheet for annotation (b) Web annotator interface (c) ASAB interface

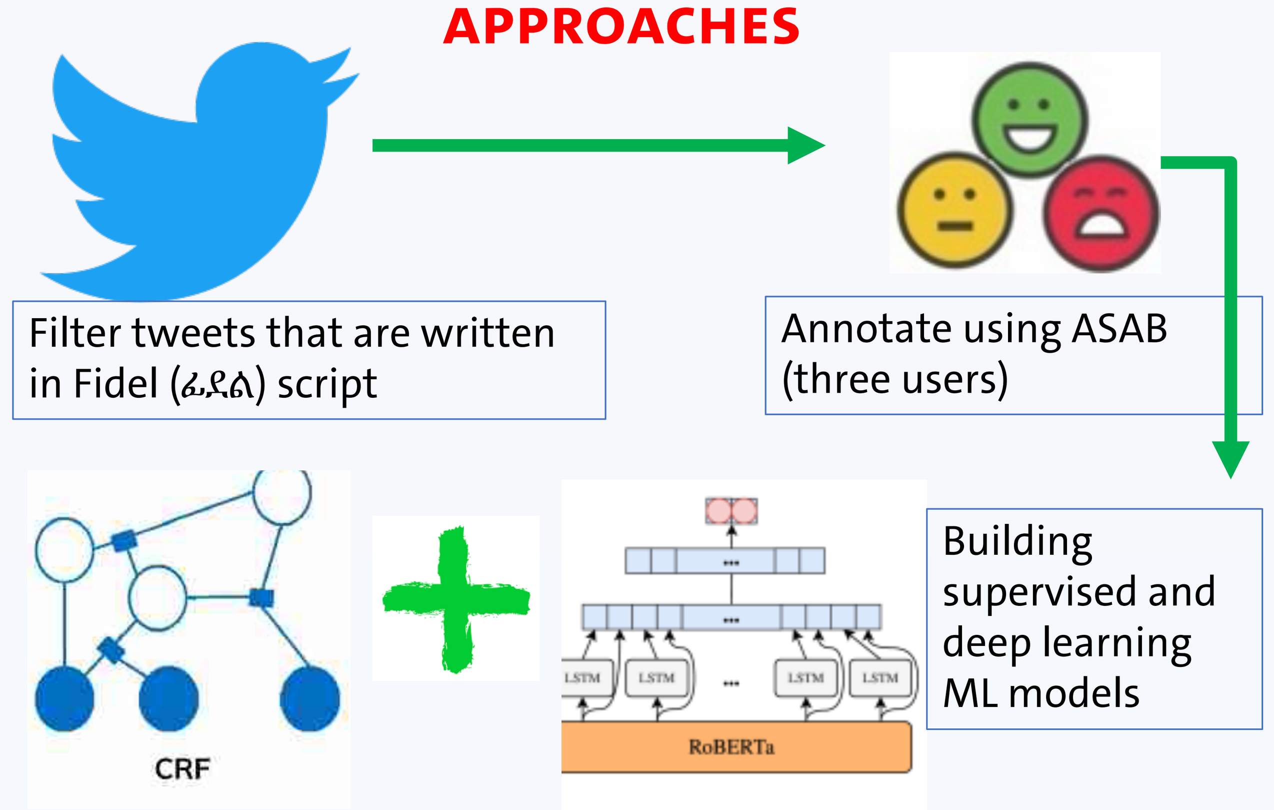
AMHARIC SENTIMENT ANNOTATOR BOT (ASAB)

- ASAB support mobile card vouchers rewards for annotators.
- Reward given when a user annotates 50 tweets.
- ASAB integrates a controlling control questions for every 6 tweets.
- A users with 3 consecutive mistakes will receive a warning message.
- User blocked after the fourth wrong attempt.

MACHINE LEARNING MODELS

- Baseline methods:
  - Stratified, Uniform, and Most frequent.
- Supervised approaches:
  - SVM, KNN, Logistic regression, Nearest centroid
  - Features: **TF-IDF** with the **CountVectorizer** and **TFIDFTransformer** methods from scikit-learn.
- Deep learning approaches:
  - Models based on FLAIR deep learning text classifier.
  - Features: Word2Vec, network embeddings, contextual embeddings (**RoBERTa** and **FLAIR** embeddings)

APPROACHES



DATA ACQUISITION AND DATASET CHARACTERISTICS

- Data Source: Ethiopic Twitter Dataset for Amharic (ETD-AM) Yimam et al. (2019).
- Data collected: December 2019 - January 2020.
- Political and social events happening:
  - The current Ethiopian Prime Minister Dr. Abiy Ahmed has received the 100th Nobel peace prize.
  - Around 17 university students were kidnapped.
  - The ruling party EPRDF was resolved and transformed itself to 'prosperity party'.
  - Religious and ethnic conflicts reached climax.

RESEARCH OUTPUTS

- Dataset
- Annotation tool
- Pre-trained models
- Source code

CONTACT/RESOURCE

yimam@informatik.uni-hamburg.de

https://github.com/uhh-lt/ASAB

RESULTS AND ANALYSIS OF ANNOTATED DATA

- 9.4k tweets annotated (143,848 words and 45,525 types), each tweet three annotators.
- A total of 92 Telegram users visited ASAB.
- 58% of users completed at least 50 tweets and got rewarded.
- 4 users blocked for consecutive mistakes.

Model	All				Cleaned			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
	TF-IDF representation							
Logistic-Regression	53.04	34.76	39.99	37.19	58.42	46.80	60.88	52.92
Random-Forest	50.48	33.97	37.75	35.76	54.36	44.59	52.17	48.09
K-Nearest-Neighbor	46.75	34.99	33.53	34.25	51.68	49.85	50.22	50.03
NearestCentroid	42.07	35.96	37.61	36.77	51.92	47.36	49.20	48.26
Support-vector-machine	52.40	33.54	41.81	37.22	53.19	35.44	46.42	40.20
	Word Embeddings							
fastText	49.63	36.26	35.08	35.66	54.12	48.35	51.33	49.79
F-Word2Vec	53.57	**41.76**	43.01	**42.38**	55.40	55.54	54.91	55.22
F-fastText	51.22	31.96	51.80	39.53	56.68	43.50	67.81	53.00
	Contextual Embeddings							
F-AmharicFlair	53.04	38.66	45.85	41.95	58.65	53.24	59.25	56.09
F-Multi-flair	53.67	36.41	41.89	38.96	55.75	46.96	55.05	50.68
F-Multi-flair-Finetuned	54.42	38.10	**45.87**	41.63	59.81	54.49	59.58	56.92
Amharic-Roberta	53.89	40.05	39.71	39.88	56.33	46.62	56.39	51.04
	Graph Embeddings							
F-DeepWalk	**54.53**	38.49	41.08	39.74	58.65	55.89	57.71	56.78
F-Role2Vec	52.40	39.45	41.16	40.29	**60.51**	**56.26**	**60.89**	**58.48**
	Baselines							
Stratified	34.72	26.59	26.71	26.65	39.02	33.79	33.80	33.80
Uniform	23.54	24.19	23.69	23.94	30.89	29.72	30.96	30.33
Mostfrequent	47.60	25.00	11.90	16.13	51.92	33.33	17.31	22.78

ERROR ANALYSIS

- We randomly select tweets where the model prediction and the user annotations differ.
- Possible source of errors:
  - Users press the wrong button by mistake.
  - Some users might not understand the tweet.
  - Due to slow internet connection, some users reported that there was a delay between the first and the second tweet.
  - Sarcasm, figurative speech, mixed scripts, incomplete phrases and sentences, and spelling and grammar errors cause most of the model errors.