Changes in European Solidarity Before and During COVID-19: Evidence from a Large Crowd- and Expert-Annotated Twitter Dataset

Hyewon Choi

Abstract

- Social solidarity, anti-solidarity.
- How European solidarity discourses changed before and after the COVID-19 outbreak was declared a global pandemic.
- Annotated data : 2.3k Eng & Ger tweets
- Multiple human annotators & two annotation approaches- experts vs crowds
- Model : BERT / 270k tweets (September 2019 ~ December 2020)
- Result : solidarity became increasingly salient and contested during the crisis. While the number of solidarity tweets remained on a higher level and dominated the discourse in the scrutinized time frame, anti-solidarity tweets initially spiked, then decreased to (almost) pre-COVID-19 values before rising to a stable higher level until the end of 2020.

1. Introduction

1. Introduction

- Social solidarity : modern societies integrated, functioning and cohesive.
- right-wing populists : social solidarity should be confined within the nation state
- supporters of the EU: European solidarity as a means to overcome the great challenges imposed on EU countries and its citizens today
- whether we can detect changes in the debates on European solidarity before and after the outbreak of COVID-19.

1. Introduction-contributions..

- annotated twitter data-> social solidarity & anti-solidarity(2.3k)
- automatically labeled tweets(over 270k) : ensemble of BERT classifiers trained on the expert and crowd annotations.
- Training BERT on crowd- & expert annotations -> over 25 points improvement over BERT trained on expert annotations alone.
- Finding : both solidarity and anti-solidarity escalated with the occurrence of incisive political events, such as European lockdowns.

2. Related Work

2. Related Work

Social Solidarity in the Social Science

- The European Migration Crisis /the Financial Crisis
- insight into solidarity discourses vs a strong focus on rejecting solidarity claims
- Binner and Scherschel, 2019; Dragolov et al., 2016; Gerhards et al., 2019; Lahusen and Grasso, 2018 -> rather optimistic
- Baker et al., 2020; Nicoli, 2017 -> 과소 표현 to political minority
- large volumes of longitudinal social media data that reflect potential fragmentation of political opinion (Sunstein, 2018) and its change over time.
- how contested European solidarity is and how it developed since the onset of COVID-19.

2. Related Work

Emotion and Sentiment Classification in NLP

- expressions of (anti-)solidarity
- Ex) texts containing a positive sentiment towards persons which are at their core anti-European, nationalistic and excluding reflect anti-solidarity.

3. Data and Annotations

3. Data and Annotations

- 1)Migration and the distribution of refugees among European member states (2015~)
- 2)Financial solidarity(financial support for indebted EU countries) (2010~)

3. Data and Annotations

Data.

- (2019.9.1~2020.12.31) – 271,930 tweets written in Eng, Germ
- an initial list of hashtags (e.g.,“#refugeecrisis”, “#eurobonds”)
- evaluated 456 co-occurring hashtags with at least 100 occurrences
- selected 45 hashtags
- Could keep the hashtag list associated with our 270k tweets constant over time.

3. Data and Annotations

Definition of Social Solidarity.

- “The preparedness to **share** one’s own resources with others, be that directly by donating money or time in support of others or indirectly by supporting the state to **reallocate and redistribute** some of the funds gathered through taxes or contributions”

Definition of anti-Social Solidarity.

- “contest social solidarity and/or **deny solidarity** towards vulnerable social groups and other European states, e.g. by promoting **nationalism** or the closure of national borders”

3. Data and Annotations

Expert Annotations.

- Tweet annotating guidelines – 4categories
- Solidarity, Anti-solidarity, ambivalent, not applicable
- Solidarity :
  - support for people in need
  - the willingness and/or gratitude towards others to share resources and help them
  - criticizing the EU ,not doing enough to share resources or help socially vulnerable groups
- Anit-solidarity :
  - the above-mentioned criteria are reversed.
  - tendencies of nationalism or advocates for closed borders

3. Data and Annotations

Expert Annotations.

- Ambivalent : both expressing solidarity and anti-solidarity statements
- Not applicable :
  - not contain the topic of (anti-)solidarity at all
  - not concerned with discourses on refugee or financial solidarity.

3. Data and Annotations

Expert Annotations.

- Annotation process (I-VI)
- kohen’s kappa value.

0~0.2: Slight (약간 일치)
0.2~0.4: Fair (어느 정도 일치)
0.4~0.6: Moderate (적당한 정도로 일치)
0.6~0.8: Substantial (상당히 일치)
0.8~1: Almost perfect (거의 완벽히 일치)

- I-III : Cohen’s kappa values of 0.51
- IV : two groups of two annotators annotated 339 tweets with hashtags not included before : 0.49
- V,VI : one group of two students annotated overall 588 tweets, with a resulting kappa value of 0.79 and 0.77 respectively

3. Data and Annotations

Expert Annotations.

- To raise Kappa value..
- discussions with the social science experts
- extension of the guidelines
- introduced a gold-standard for annotations
  -majority voting and discussions among the annotators
  -gold-standard label could not be reached->a social science expert decided
  -some hard cases were left undecided (not included in the dataset)

3. Data and Annotations

Expert Annotations.

- human reference performance

	4 classes	3classes(ambivalent + not applicable)
Kappa	0.64	0.69
F1 score	69%	78.5
Final stage(6) – F1 score	Above 80%	85.4%~89.7%

3. Data and Annotations

Crowd annotations

- students in an introductory course to NLP
- students - the guidelines, 100 expert annotated tweets
- 1) reading the guidelines
- 2) looking at 30 random expert annotations
- 3) they were asked to annotate 20 tweets themselves
- 4) self-report their kappa agreement with the experts
- 5) repeated this with another 30 tweets for annotator training 20 tweets for annotator testing.
- 6) training set : 30 expert-annotated tweets /30 entirely novel tweets

3. Data and Annotations

Crowd annotations

- 125 students.
- Each novel tweet was annotated by up to 3 students (2.7 on average).
- simple strategy:
  1) the majority label
  2) the annotation of the most reliable annotator

3. Data and Annotations

Crowd annotations

- 0.6 ~ 0.7 – majority of students
- 4 class : 0.5 ~ 0.6

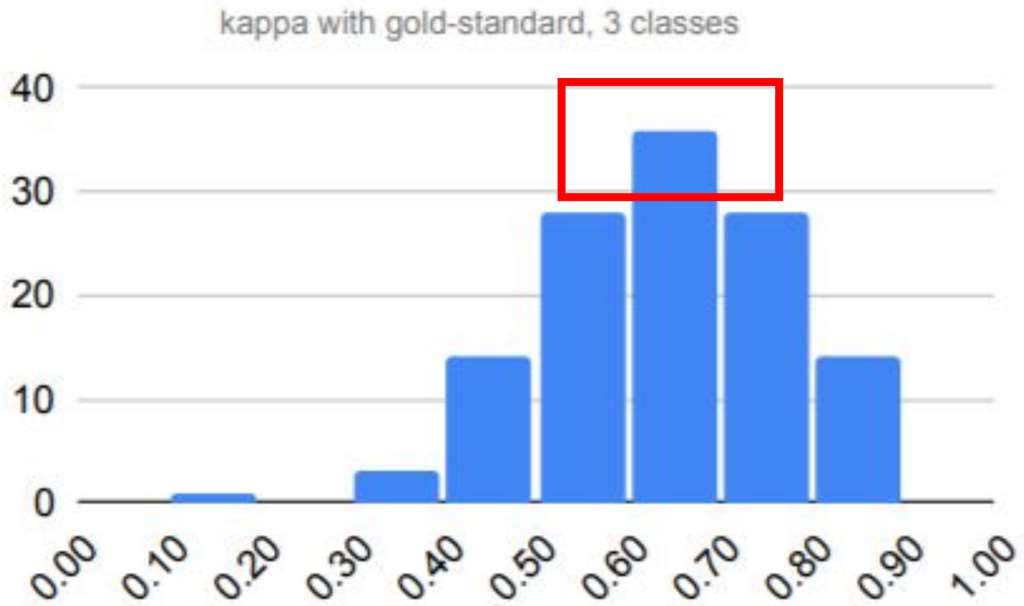

Figure 1: Distribution of kappa agreements of crowd workers with expert annotated gold-standard, 3 classes.

3. Data and Annotations

Crowd annotations

- 1196 : Eng, 1103 : Ger
- Crowds(S) : 50% of total
- Experts(S) : 30% of total

	S	A	AMB	NA	Total
Experts	$\boxed{386}$	246	113	174	919
Crowds	$\boxed{768}$	209	186	217	1380

Table 2: Number of annotated tweets (after geofiltering) for the four classes `solidarity` (S), `anti-solidarity` (A), `ambivalent` (AMB), and `not applicable` (NA).

4. Methods

4. Methods

- Multilingual BERT(MBERT)
- XLM-R
- 1) data augmentation 2) transfer learning techniques

- Oversampling of minority classes
- Back-translation
- Fine-tuning
- Auto-labeled data
- Ensembling

4. Methods

- Oversampling of minority classes
  :randomly duplicate (expert and crowd annotated) tweets
- Back-translation
  : the Google Translate API
  Eng-> Ger (pivot language)
  Ger-> Eng(for annotator, expert&crowd)
  To oversample.
- Fine-tuning
  :MBERT / XLM-R with masked language model and next sentence prediction tasks on domain-specific data
- Auto-labeled data
  : 270k tweet, training set : 35k -> 105k
- Ensembling: k = 15

5. Experiments

5. Experiments

- **5.1 Experimental setup**
  - only hashtages
  - only texts
  - combining expert and crowd annotations for training
  - examining the augmentation and transfer learning strategies
  - ensembling various models

5. Experiments

5.2 Results

Condition	Train size	MBERT Dev	MBERT Test	XLM-R Dev	XLM-R Test
E, Hashtag only	572	51.7±0.5	49.0±1.1	48.0±0.9	44.0±0.8
E	579	64.2±1.2	57.7±0.4	64.0	63.3
E+C	1959	66.4±0.5	64.0±1.5	65.0	64.8
E+C, No Hashtags	1959	64.0±0.3	58.0±0.5	62.0	60.0
E+C, Hashtag Only	1567	55.8±2.0	49.5±2.1	47.8	42.2
E+C+Auto label	106959	76.4	78.3	77.5	78.4
E+C+Auto label+Oversample	108048	76.4	76.3	77.4	76.9
E+C+Auto label+Backtranslation	108918	76.0	77.1	77.5	78.7
E+C+Auto label+Pretraining	106959	78.4	78.8	78.6	79.0
E+C+ALL	110007	78.8±1.3	78.6±0.8	78.9	79.7

Table 3: Macro-F1 scores (in %) for different conditions. Entries with ± give averages and standard deviations over 3 different runs with different test and dev sets. 'E' stands for Experts, 'C' for Crowds. 'ALL' refers to all data augmentation and transfer learning techniques.

5. Experiments

5.3 Model Analysis

		Predicted		
		S	**A**	**O**
	S	63	3	2
Gold	**A**	5	37	4
	O	5	6	45

Table 4: Confusion matrix for best ensemble with macro-F1 score of 84.5% on the test set (for one specific train, dev, test split).

5. Experiments

LIME 알고리즘에 대한 설명 :
https://blogs.sas.com/content/saskorea/2018/12/10/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-%ED%95%B4%EC%84%9D%EB%A0%A5-
%9C%EB%A6%AC%EC%A6%88-4%ED%83%84-%EB%9D%BC%EC%9E%84lime%EC%9C%BC%EB%A1%9C-%EB%AA%A8%EB%8D%B8-%ED%95%E
4%9D%EB%A0%A5/

5.3 Model Analysis

Figure 2: Our best-performing model (macro-F1 of 84.5%) predicts `anti-solidarity` for the current example because of the hashtag #Remigration (according to LIME). The tweet, also given as translation in Table 5 (2) below, is overall classified as `other` in the gold standard, as it may be considered as expressing no determinate stance. Here, we hide identity revealing information in the tweet, but our classifier sees it.

Text	Gold	Pred.
(1) You can drink a toast with the AFD misanthropists #seenotrettung #NieMehrCDU	S	A
(2) Why is an open discussion about #Remigration (not) yet possible?	O	A
(3) Raped and Beaten, Lesbian #AsylumSeeker Faces #Deportation	A	O

Table 5: Selected misclassifications of best performing ensemble model. We consider the bottom tweet misclassified in the expert annotated data (correct would be `solidarity`). Tweets are paraphrased and/or translated.

5. Experiments

5.3 Model Analysis

		Predicted		
		S	**A**	**O**
Gold	**S**	63	3	2
	A	5	37	4
	O	5	6	45

Table 4: Confusion matrix for best ensemble with macro-F1 score of 84.5% on the test set (for one specific train, dev, test split).

5. Experiments

5.4 temporal analysis

$$S/A := \frac{\#\text{Solidarity tweets}}{\#\text{Anti-Solidarity tweets}}$$

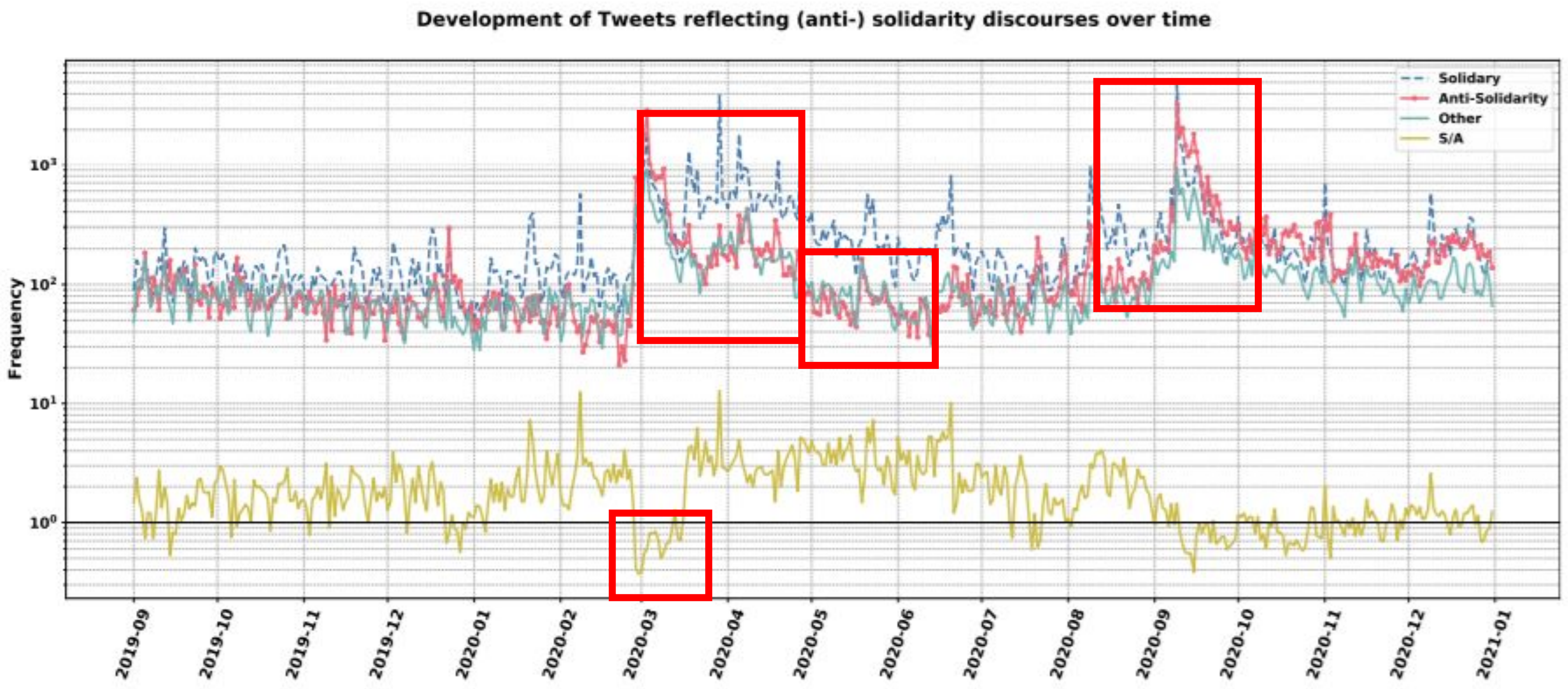

Figure 3: Frequency of `solidarity` (S), `anti-solidarity` (A) and `other` (O) tweets over time as well as the ratio S/A. The constant line 1 indicates where $S = A$. Y-axis is in log-scale.

Moria migrants: Fire destroys Greek camp leaving 13,000 without shelter

9 September 2020

Europe migrant crisis

6. Conclusion

6. Conclusion

- The textual material in social media
- (anti-)solidarity statements climaxed momentarily with the first lockdown
- The predominance of solidarity expressions was quickly restored at higher levels than before.
- (Anti-)Solidarity statements were balanced by the end of the year 2020, when infection rates were rising.
- The COVID-19 pandemic has not severely negatively impacted the willingness of European Twitter users to take responsibility for refugees, while financial solidarity with other EU countries remained low.
- These findings are in line with survey-based, quantitative research and its rather optimistic overall
- Severe strains during crises coincide with increased levels of anti-solidarity statements.

Thank
You.