



Valerio Basile, Danilo Croce, Maria Maro and Lucia C. Passaro (dir.)

**EVALITA Evaluation of NLP and Speech Tools for Italian
- December 17th, 2020**
Proceedings of the Seventh Evaluation Campaign of Natural
Language Processing and Speech Tools for Italian Final Workshop

Accademia University Press

UNITOR @ Sardistance2020: Combining Transformer-based Architectures and Transfer Learning for Robust Stance Detection

Simone Giorgioni, Marcello Politi, Samir Salman, Danilo Croce and Roberto Basili

DOI: 10.4000/books.aaccademia.7092
Publisher: Accademia University Press
Place of publication: Torino
Year of publication: 2020
Published on OpenEdition Books: 11 May 2021
Serie: Collana dell'Associazione Italiana di Linguistica Computazionale
Electronic ISBN: 9791280136329



<http://books.openedition.org>

Electronic reference

GIORGIONI, Simone ; et al. *UNITOR @ Sardistance2020: Combining Transformer-based Architectures and Transfer Learning for Robust Stance Detection* In: *EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020: Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian Final Workshop* [online]. Torino: Accademia University Press, 2020 (generated 17 mai 2021). Available on the Internet: <<http://books.openedition.org/aaccademia/7092>>. ISBN: 9791280136329. DOI: <https://doi.org/10.4000/books.aaccademia.7092>.

UNITOR @ Sardistance2020: Combining Transformer-based Architectures and Transfer Learning for Robust Stance Detection

Simone Giorgioni, Marcello Politi, Samir Salman, Danilo Croce and Roberto Basili

Department of Enterprise Engineering, University of Roma, Tor Vergata

Via del Politecnico 1, 00133 Roma, Italy

{simone.giorgioni,marcello.politi,samir.salman}@alumni.uniroma2.eu

{croce,basili}@info.uniroma2.it

Abstract

English. This paper describes the UNITOR system that participated to the Stance Detection in Italian tweets (Sardistance) task within the context of EVALITA 2020. UNITOR implements a transformer-based architecture whose accuracy is improved by adopting a Transfer Learning technique. In particular, this work investigates the possible contribution of three auxiliary tasks related to Stance Detection, i.e., Sentiment Detection, Hate Speech Detection and Irony Detection. Moreover, UNITOR relies on an additional dataset automatically downloaded and labeled through distant supervision. The UNITOR system ranked first in Task A within the competition. This confirms the effectiveness of Transformer-based architectures and the beneficial impact of the adopted strategies.

Italiano. *Questo lavoro descrive UNITOR, uno dei sistemi partecipanti allo Stance Detection in Italian tweet (SardiStance) task. UNITOR implementa un'architettura neurale basata su Transformer; la cui accuratezza viene migliorata applicando un metodo di Transfer Learning, che sfrutta le informazioni di tre task ausiliari, ovvero Sentiment Detection, Hate Speech Detection e Irony Detection. Inoltre, l'addestramento di UNITOR può contare su un insieme di dati scaricati ed etichettati automaticamente applicando un semplice metodo di Distant Supervision. Il sistema si è classificato al primo posto nella competizione, confermando l'efficacia delle architetture basate su Transformer e il contributo delle strategie adottate.*

1 Introduction

Stance detection aims at detecting if the author of a text is in favor of a target topic, or against it (Krejzl et al., 2017). In this task, a text pair is generally considered: one text expresses the topic, while the other one reflects the author's judgments. In a possible variant to such a setting, the topic is implicit within an entire document collection over which the stance detection is applied.

In this work, we will consider this last setting, as defined in the in the Stance Detection in Italian Tweets (SardiStance) task (Cignarella et al., 2020) within the EVALITA 2020 (Basile et al., 2020). A set of texts (here tweets) is provided, almost all concerning the same topic, i.e., the Sardines Movement¹. The goal is to recognize if each tweet is for or against (or neither) such target, only exploiting textual information. According to the task definition, this corresponds to the so-called Task A. This is quite challenging problem, since it requires at the same time to discover if a text refers to the target topic and the author's orientation, only relying on short messages written in a very conversational style.

We thus present the UNITOR system participating to the SardiStance task A. The system is based on a Transformer-based architecture for text classification (Devlin et al., 2019) that is directly pre-trained over a large-scale document collection written in Italian, namely UmBERTo. In a nutshell, the adopted architecture, which has been demonstrated achieving state-of-the-art results in many NLP tasks (Devlin et al., 2019), takes in input a message and associates it to one of the target classes indicating the stance. Moreover, due to the task complexity and the small size of the dataset, in order to improve the generalization capabilities of the neural network, we adopted a Transfer Learning approach (Pan and Yang, 2010). Our main assumption is that Stance Detection is tied to other tasks involving emotion and subjectivity

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹https://en.wikipedia.org/wiki/Sardines_movement

analysis (such as Sentiment Analysis or Irony Detection) even though important differences do exist among them. As a simplified example, let us consider a message such as “*I like the Sardines Movement*”: it clearly expresses a positive sentiment, also being in favour of the target topic. However, a message such as “*I like the EVALITA campaign.*” is positive as well but it does not express any support or opposition to the Sardines (and it should be associated to the `None` class). We thus speculate that an automatic system trained over an auxiliary task (e.g., Sentiment Classification) is beneficial, but the transfer process must be carefully designed in order to avoid catastrophic forgetting or interference problems (McCloskey and Cohen, 1989).

In this work, we investigate the possible contribution of three auxiliary tasks involving the recognition of emotions according to different settings, i.e., Sentiment Detection and Classification, Hate Speech Detection and Irony Detection. We adopt three different classifiers (one for each auxiliary task) and use them to add additional information to the tweets provided in the SardiStance dataset. As an example, when considering the auxiliary task involving Hate Detection, the corresponding classifier will augment each input tweet by expressing if this expresses hate or not. After this step, the final classifier is expected to learn the association between messages and the stance categories, “being aware” (with some unavoidable noise) if the message expresses some sort of hate, irony and more generally, sentiment. Finally, we investigate the possibility of augmenting the training material by automatically downloading messages and labeling them through distant supervision (Go et al., 2009). We first selected few hashtags clearly in favour (or not) of the target topic to download and label a set of set of messages. Then, in order to add a set of neutral messages, we selected a set of news titles concerning the Sardines Movement.

The UNITOR system ranked first in the competition, suggesting that the combination of the Transformer-based learning with the adopted strategies of Transfer Learning and Data Augmentation is beneficial. In the rest of the paper, Sec. 2 describes UNITOR. In Sec. 3, the evaluations are reported while Sec. 4 derives the conclusions.

2 Transformer-based architectures and Transfer Learning for Stance Detection

The UNITOR system implements a Transformer-based architecture described in Section 2.1. The

adopted auxiliary tasks are described in Section 2.2, while our Transfer learning strategy is in Section 2.3. Finally, an automatic strategy for Data Augmentation is presented in Section 2.4.

2.1 UNITOR as a Transformer-based Architecture

The approach proposed in (Devlin et al., 2019), namely Bidirectional Encoder Representations from Transformers (BERT) provides a very effective model to pre-train a deep and complex neural network over large scale collections of non annotated texts and to apply it to a large variety of NLP tasks. The building block of BERT is the Transformer element (Vaswani et al., 2017), an attention-based mechanism that learns contextual relations between words in a text. BERT provides a sentence embedding (as well as the contextualized lexical embeddings of words in the sentence) through a pre-training stage aiming at the acquisition of an expressive and robust language and text model. The Transformer reads the entire input sequence of words at once and is optimized through two pre-training tasks. The first pre-training objective is the (*masked language modeling*) (Devlin et al., 2019). In addition, a *Next Sentence Prediction* task is used to jointly pre-train text embeddings able to soundly represent discourse level information. This last objective operates on text-pair representations and aims at modeling relational information, e.g. between the consecutive sentences in a text. On top of the produced embeddings, BERT applies a *fine-tuning* stage devoted to adapt the entire architecture to the targeted task.

The fine-tuning process of BERT for sentence classification (here adopted) operates on a single texts or text pairs, which can be given in input to BERT, in analogy with a next sentence prediction task. The special token `[CLS]` is used as first element of each input sequence and the embedding produced by BERT are used in input to a linear classifier customized for the target classification task. While the BERT architecture is pre-trained on large-scale corpora, its application to new tasks is generally obtained by customizing the final classifier to the targeted problem and fine-tuning all the network parameters for few epochs, to avoid catastrophic forgetting. In (Liu et al., 2019b) RoBERTa is proposed as a variant of BERT which modifies some key hyperparameters, including removing the next-sentence pre-training objective, and training on more data, with much larger mini-

batches and learning rates. This allows RoBERTa to improve on the masked language modeling objective compared with BERT and leads to better downstream task performances.

UNITOR is based on a RoBERTa architecture pre-trained over Italian texts: we adopted UmBERTo² which is pre-trained over a subset of the OSCAR corpus, made of 11 billion tokens. These architectures achieved state-of-the-art results in a wide range of NLP tasks. However, they also rely on large scale annotated datasets composed of (possibly hundreds) thousands of examples. In order to improve the quality of this architecture in the SardiStance Task with a quite limited dataset, we adopted a simple Transfer Learning strategy by relying on the following three auxiliary tasks.

2.2 Supporting UNITOR through Auxiliary tasks

In this work, we speculate that the complexity of the Stance detection task can be simplified whenever the system to be trained is already aware if input messages express some sort of Sentiment, Irony or Hate. In order to expose UNITOR to such information, we trained specific classifiers over dedicated corpora made available in the previous editions of EVALITA, as it follows:

Sentiment Detection and Classification. This task consists in the automatic detection of subjectivity (and the eventual positive or negative polarity) in texts (Pang and Lee, 2008). Even though the Stance Detection is clearly different from a traditional task of Sentiment Analysis, we speculate that they are nevertheless related. As an example, we can suppose that the presence of stance is more probable in messages expressing subjectivity. We thus considered the setting proposed in SENTIPOLC 2016 (Barbieri et al., 2016) where a dataset of 8,000 tweets is made available. For each message, the presence of subjectivity is made explicit and, eventually, the positive and negative polarity. The labeling provided in the dataset was slightly modified and mapped to a classification problem over three classes: all objective tweets were labeled with the special tag <neutrale>, the subjective and positive messages with <positivo> while the negative ones with <negativo>³.

²<https://huggingface.co/Musixmatch/umberto-commoncrawl-cased-v1>

³We discarded the few available messages with mixed polarity, to simplify the final classification task.

Irony Detection. We speculate that a robust detection of stance requires the recognition of irony, which can even reverse the output of the classification task. For example a false stance can be expressed through an ironic message, such as “*Le Sardine sono il futuro passato dell’Italia*”⁴. The objective of Irony Detection is to detect whether a given message is ironic or not. We used the dataset provided IronITA 2018 (Cignarella et al., 2018), where a dataset of 4,800 labeled messages is made available. We adopted the original binary classification task, mapping ironic messages to the <ironico> and <non ironico> labels.

Hate Speech Detection. Being against a topic can be often expressed through messages expressing also hate. We thus introduce also the Hate Speech Detection task, which involves the automatic recognition of hateful contents. We considered the setting proposed in HaSpeeDe 2018 (Bosco et al., 2018), where a dataset of 3,000 messages is made available. We adopted the original binary classification task: we mapped messages expressing hate with the <odio> label and <non odio> in the other case.

2.3 Transferring auxiliary tasks in the Transformer-based learning

In order to transfer the information from each auxiliary task into UNITOR, we first trained a specific UmBERTo-based sentence classifier on each of the datasets described in the previous section. In each case, the standard parameters proposed in (Devlin et al., 2019) are used to fine-tune the model⁵. After these three training steps, the entire SardiStance dataset is processed by each of the three classifiers and the resulting labels are used to “augment” the input messages. In particular, these labels generated a sort of new sentence, which is paired with the corresponding message. The following example shows how a tweet⁶ against the movement is used in input to UNITOR:

“*[CLS] negativo ironico odio [SEP] #elezioniregionali Le Sardine aiuteranno a salvare il Paese! #mafammilpiacere Sono proprio dei bei perdigiorno falliti! [SEP]*”

Consistently with (Devlin et al., 2019), the first

⁴In English: “*Sardines are the future past of Italy*”

⁵The number of epochs was tuned over a development set made of 10% of the corresponding dataset and the best epoch was selected by maximizing the classification accuracy.

⁶In English: “*#regionalelections The Sardines will help to save the country! #please They’re just a bunch of losers!*”

pseudo-token [CLS] is added to generate the embedding used in input in the final linear classifier. Then, the pseudo-sentence “*negativo ironico odio*” suggests that the message expresses negative polarity and hate through the adoption of irony. Finally, between the [SEP] pseudo-tokens, the original message is reported. This particular schema resembles the classification of text pairs used in relational learning tasks, such as in Textual Entailment (Devlin et al., 2019). The output of the auxiliary classifiers defines a sort of hypothesis, i.e., *the authors aims at expressing a negative sentiment through an ironic message which also expresses hate*, while the original message is the direct consequence, i.e., the “implied” message⁷. The UNITOR model is thus an UmBERTo-based classifier trained over text pairs, where the first element encodes the information derived from the auxiliary tasks and the second one is the original message. Even though the quality of this labeling process can introduce noise (due to incorrectly classified messages) this augmented input is expected to simplify the final training process, by explicitly providing information about sentiment, hate and irony.

2.4 Distant Supervision for Stance Detection

In order to balance the limited amount of available data (especially considering the complexity of the task) we augmented the training material by labeling additional messages via Distant Supervision (Go et al., 2009). We speculate that a tweet containing an hashtag such as *#vivalessardine* (in English: *#ILikeSardine*) is in favour to Sardines instead of a tweet containing for example *#sardinefritte* (in English: *#friedSardine*) is against to our target. Hence, we downloaded from the TWITA corpus (Basile and Nissim, 2013) 3,200 tweets and labeled them via Distant Supervision. In particular, the following subset are derived: 1,500 tweets against the movement since containing *#gatticonsalvini* and 1,000 tweets in favour, since containing *#nessunotochilesardine*, *#ios-toconlesardine*, *#unmaredisardine*, *#vivalessardine* or *#forzasardine*. Finally, to enlarge the subset of messages without stance, 700 neutral statements were downloaded, which are actually titles from news, derived by querying “*sardine*” in Google

⁷We investigate different ways to encode this information, even using complex sentences, but negligible differences in the tuning process were measured, so we applied the simplest schema.

news. In the experimental evaluations discussed in the next section, this dataset of “silver” data is simply added to the training material. To avoid over-fitting, we removed 90% of the occurrences of the hashtags used as query in the new data.

3 Results and Discussion

UNITOR participated to Task A - Textual Stance Detection (Cignarella et al., 2020) where the available dataset is composed by 2,132 tweets concerning the Sardines Movement: 1,028 tweets are against the movement (label *Against*), 589 tweets in favour of it (label *Favour*) and 515 tweets do not express any stance about the target topic (label *None*).

As discussed in Section 2, UNITOR is based on the UmBERTo pre-trained model, which relies on the RoBERTa architecture. For parameter tuning, we adopted a 10-cross fold validation, so that the training material is divided in 10 folds, each split according to 90%-10% proportion. The model is trained using a standard Cross-entropy Loss and an ADAM optimizer initialized with a learning rate set to $2 \cdot 10^{-5}$ and linearly decreased during the training process. We trained the model for 5 epochs, using a batch size of 32 elements. At test time, an Ensemble of such classifiers is used: each message is in fact classified using all 10 models trained in the different folds and the label suggested by the highest number of classifiers is selected. In the Task A, we submitted two constrained runs, i.e., system considering only tweets from the competition, and two unconstrained ones, where additional tweets were acquired and labeled by applying the approach presented in Section 2.2. All models are implemented using Pytorch⁸ and experiments were run over Google Colab⁹.

Results are reported in Table 1 in terms of Precision, Recall and F1 scores obtained by the different models with respect to each label. The final rank considers the average F1 (F1-avg) between the *Favour* and *Against* classes.

First of all, the high complexity of this task is confirmed by the results obtained by the strong Baseline method (the last row). It is a Support Vector Machine trained over a simple Bag-of-Word model (Cignarella et al., 2020) and achieves an average F1 of 57.84%, being competitive with many systems participating to the task and ranking 13th over 22 submissions. One important re-

⁸<https://pytorch.org/>

⁹<http://colab.research.google.com/>

Rk	System	F1			Rec			Prec			
		avg	Against	Favor	None	Against	Favor	None	Against	Favor	None
1	UNITOR_u_1	68.53%	78.66%	58.40%	39.10%	76.01%	57.65%	45.35%	81.50%	59.16%	34.36%
2	UNITOR_c_1	68.01%	78.81%	57.21%	39.79%	74.66%	63.78%	43.60%	83.43%	51.87%	36.59%
3	UNITOR_c_2	67.93%	79.39%	56.47%	36.72%	77.09%	61.22%	37.79%	81.83%	52.40%	35.71%
4	Opponent_c_1	66.21%	75.80%	56.63%	42.13%	68.60%	64.29%	52.91%	84.69%	50.60%	35.00%
5	UNITOR_u_2	66.06%	76.89%	55.22%	37.02%	72.64%	56.63%	44.77%	81.67%	53.88%	31.56%
6	UmBERTo	65.69%	77.41%	53.97%	35.93%	74.12%	57.14%	40.11%	81.00%	51.14%	32.54%
13	Baseline	57.84%	71.58%	44.09%	27.64%	68.06%	49.49%	29.65%	75.49%	39.75%	25.89%

Table 1: Results obtained by UNITOR at the SardiStance task. In bold best results for each measure. In the system name "c" and "u" refer to constrained and unconstrained runs.

sult is obtained by the straight application of the UmBERTo model over the original messages (next to last row in Table 1). In fact, this Transformer-based architecture, empowered with the Ensemble technique, achieves an average F1 of 65.69%: a system which directly applies an Ensemble of UmBERTo-based models would have ranked 6th in the competition.

We thus trained UmBERTo by adopting the Transfer Learning approach presented in Section 2.3 in the constrained setting. The adoption of all the three auxiliary tasks led to the constrained submission called UNITOR_c_2. Moreover, we considered the training of UmBERTo by considering one auxiliary task at a time. When considering only the Hate Speech Detection task, better results were obtained over the development set, with respect to the adoption of the other tasks taken individually, i.e., Sentiment Detection and Irony Detection¹⁰. Such a variant, called UNITOR_c_1, considers tweets enriched only with information derived by the hate classifier and it generally shows higher precision with respect to the Against class. This suggests that a tweet expressing hate is more likely in opposition to the Sardines Movement. Both constrained models ranked 3rd and 2nd in the competition, respectively. These results are impressive as they both outperformed of about 2% of absolute F1 the standard UmBERTo. Moreover, they confirm the beneficial impact of Hate Speech Detection as an auxiliary task. Finally, we augmented the training dataset by using the additional data presented in Section 2.2. We extended the training material used to train UNITOR_c_2 in order to obtain the unconstrained submission called UNITOR_u_2. It is worth noticing that all three auxiliary tasks were used in this submission. This led to a performance drop, i.e. a 66.06% of average F1, which is lower

¹⁰The results of this tuning stage were not reported here for lack of space.

with respect to the best opponent system, which achieved a 66.21% of F1. It seems that the noise added both from the auxiliary tasks and the additional data, negatively impacted the overall quality. On the contrary, when only the Hate Speech Detection task is considered (i.e., UNITOR_u_1) additional data are positively capitalized by the model, achieving the best average F1 score in the competition, i.e. 68.53%. These results suggest that the combination of the Transformer-based learning with the adopted strategies of Transfer Learning and Data Augmentation is highly beneficial, when only Hate is considered.

From an error analysis, it seems that a significant number of incorrect classifications occurred in longer and complex messages, where the topic of the stance is not clearly explicit nor captured by the UmBERTo model, such as in “#carfagna: “io per i liberali che non si affidano a Salvini” e “dalle sardine buone idee”. Auto-scacco in due mosse. Con la Polverini poi...”¹¹. This message is considered to be Against while the system assigns the label None. Here, it is very challenging to understand the connection between the “good ideas of the sardines” and the very colloquial expression “Auto-scacco” which can be translated as “She messed herself”. The same appears in the tweet “Ho finalmente capito chi mi ricordava Mattia Santori, quello delle sardine: Lodo Guenzi. (e infatti in quanto a democristianità stiamo lá)”¹² which again labeled Against but classified as None. Clearly the system is not able to link the movement to its leader nor to the negative opinion about belonging to the Christian Democrat Party. Another example is the tweet “Dopo

¹¹In English: “#carfagna: “come with me liberals who do not rely on Salvini” and “from Sardines movement good ideas.” She messed herself up with two moves. Not to mention Polverini...”

¹²In English: “I finally understood who reminded me of Mattia Santori, the one with the Sardines movement: Lodo Guenzi. (in fact as far as Christian Democrats are concerned they are pretty the same).”

*avere ascoltato @luigidimaio mi viene in mente una sola parola:grazie. Fiducia nelle sue scelte e immenso rispetto per i grandi risultati ottenuti. Ora un nuovo inizio, con un nuovo entusiasmo. Andiamo versogli #statigenerali con serietà e maturità. Forza@mov5stelle!*¹³. Here the system incorrectly assigns the Favour label because the tweet is in favour of a different movement.

4 Conclusion

In this work we present the results obtained by the UNITOR system, which participated to the SardiStance task. UNITOR ranked first in Task A, both for constrained and unconstrained runs. These results confirm the beneficial impact of Transformer based architecture for text classification also in the Stance Detection task. Moreover, we demonstrate the beneficial impact of Hate Speech Detection as an auxiliary task in a Transfer Learning setting. Finally, we empirically demonstrate that the adoption of Distance Supervision is useful to reduce data sparseness. Future work will apply the above approaches to task B within SardiStance. Moreover, we will investigate multi-task learning approaches (Liu et al., 2019a) to capitalize information from auxiliary tasks in a more principled way.

References

- Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, and Viviana Patti. 2016. Overview of the evalita 2016 sentiment polarity classification task. In *Proceedings of EVALITA 2016, Napoli, Italy, December 5-7, 2016*, volume 1749 of *CEUR Workshop Proceedings*.
- Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, Atlanta.
- Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro. 2020. Evalita 2020: Overview of the 7th evaluation campaign of natural language processing and speech tools for italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*. CEUR-WS.org.
- ¹³In English: “After listening to @luigidimaio only one expression came to my mind: thank you. I have trust in his choices and a huge respect for the great results obtained. Now it’s a new start, with new enthusiasm. Let’s move towards the #statigenerali with seriousness and maturity.Forza@mov5stars”
- Cristina Bosco, Felice Dell’Orletta, Fabio Poletto, M. Sanguinetti, and M. Tesconi. 2018. Overview of the evalita 2018 hate speech detection task. In *EVALITA@CLiC-it*.
- Alessandra Teresa Cignarella, Simona Frenda, Valerio Basile, Cristina Bosco, Viviana Patti, Paolo Rosso, et al. 2018. Overview of the EVALITA 2018 task on irony detection in Italian tweets (IronITA). In *Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*, volume 2263, pages 1–6.
- Alessandra Teresa Cignarella, Mirko Lai, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. SardiStance@EVALITA2020: Overview of the Task on Stance Detection in Italian Tweets. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. CEUR-WS.org.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL 2019*, pages 4171–4186, Minneapolis, Minnesota, June.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report.
- Peter Krejzl, Barbora Hourová, and Josef Steinberger. 2017. Stance detection in online discussions.
- Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019a. Multi-task deep neural networks for natural language understanding. In *Proceedings of ACL*, pages 4487–4496, Florence, Italy, July.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.
- Michael McCloskey and Neil J. Cohen. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. *The Psychology of Learning and Motivation*, 24:104–169.
- S.J. Pan and Q. Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008.