TIET at CLEF CheckThat! 2020: Verified Claim Retrieval

Utsav Shukla and Aayushmaan Sharma

Thapar Institute of Engineering and Technology, India
{ushukla_be17,aaayushmaan_be17}@thapar.edu

Abstract. Internet is the all-singing, all-dancing, multi-pronged tool of modern man, but, for its many benefits it also presents us with daunting challenges. The user is bombarded with tons of information and it is becoming increasingly difficult to tell fact from fiction. Consequently, over the past few years research work has intensified in this regard to better equip the user against false information and indoctrination. In this paper, we propose a system for automatic retrieval of supporting claims given an input claim. Our system uses elasticsearch along with a transformer model to generate a similarity score in two steps. We have submitted our system to CLEF-2020 Check That! Lab's Task-2: Claim Retrieval and our primary evaluations gave promising results which we will present, analyse and improve henceforth in the paper.

Keywords: Fact checking · Information Retrieval · Claim Retrieval · Natural Language Processing

1 Introduction

The internet has come a really long way in the past couple of decades in terms of acceptability and reach. And with easy accessibility to low-cost internet, and wide-appeal of social media services like Twitter, there is an overabundance of unverified information online. Now, this glaring lack of factual validation combined with the lucrative carte blanche that the internet allows, has given way to the taxing problem of gross misinformation and disregard for factual correctness. Needless to say, this has been a green pasture for feckless rumour-mongers and ill-intentioned propagandists. So, now more than ever, there is an exigent demand for automatic fact checking to safeguard internet users from misinformation and indoctrination. To that end, multiple investigative-journalistic fact-checking organisations like Snopes, IFCN, Full Fact etc. have emerged. But, manual fact checking is a very demanding and slow process that can take one full day to research and write about a claim [1].

Looking at these problems faced in manual fact checking, automated fact checking can be very advantageous to fully automate or assist the existing pipelines. There has been a significant amount of work in the field of automated fact checking in recent years. Various tasks have been formulated related to or that come under automated fact checking [2]. There are some research works that use external knowledge sources too (Web in this case) for a fully automated fact checking pipeline [3]. A very similar methodology that aims at retrieving already fact checked claims was proposed in a very recent work [4]. To contribute to the ongoing research in this regard, we, team Trueman from TIET, decided to participate in the CLEF Check That! Lab's Task-2: Claim Retrieval [5] [6].

In this paper, we focus on one of the key components of a fact-checking pipeline, that is, claim retrieval. The principles governing our approach are simplicity, directness and scalability. Our system uses elastic search along with transformer model similarity between the tweet and claim to generate a score in two steps : exact matching of words using BM25 and then, an NLP-based model to check for semantic similarities [7]. The initial evaluation results were good with the primary task being : fact checking of tweets by finding related, corroborating claims in the corpus given for the task. However, to improve accuracy we undertook manual analysis of faulty results and subsequently designed a suitable experiment to make an attempt to troubleshoot the problems faced. Code for our experiments and our results file can be found here[1].

2 Methodology

We propose an unsupervised system for the task. Our submitted system uses elastic search BM25 score along with cosine similarity between two text pieces' BERT encodings. Figure 1 shows the working of our submitted system.

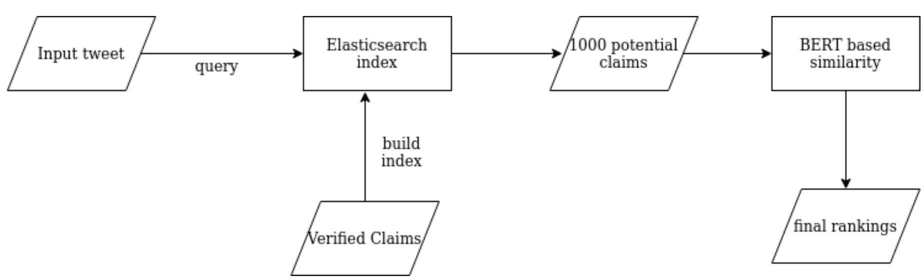

Fig. 1: Our submitted system

[1] https://github.com/us241098/checkthat2020_submission

2.1 Indexing the verified claims:

Firstly, all verified claims are indexed in elasticsearch. Ideally some preprocessing like removing URLs, hashtags and @ sign should be performed before indexing, but we skip this step as verified claims in the corpus are mostly free of these. Creating this index ensures quick retrieval of related articles at the time of query.

2.2 Ranking using BM25:

At the time of query, we retrieve top 1000 matching claims for the query/tweet along with their BM25 scores. BM scores are dependent on the exact match between the words of claim and tweet and are assigned to every claim-tweet pair. However, our initial experiments have shown that using just these scores for ranking claims give underwhelming results, hence, we use transformer-based similarity in our next step and update ranks after adding the scores assigned by it.

2.3 Transformer based model:

Since BM25 similarity relies on exact word match between the text pieces, it fails to capture the semantic similarity between claims and tweets. To compensate for this, we have used a BERT fine-tuned on NLI data to generate encodings for our text piece, we get the similarity between two encodings by computing cosine distance between them. Based on the similarity we update the scores from BM25 and subsequently update the ranks. In our system's implementation we have used the *sentence_transformer* python package and *bert_base_nli_mean_tokens* model.

3 Results and Improvements

In this section, results of our submission to Task 2 will be discussed and compared to the best performing system. In the subsequent subsection, we will discuss the cases where our system performs poorly and attempt to improve our submitted system. Table 1 shows our system's result, metrics on which systems were evaluated were MAP@1, 3, 5, all, Precision@1, 3, 5 and Reciprocal Rank@1, 3, 5. Systems were ranked on the basis of performance on MAP@5. Our submission was ranked 6th among 9 participants.

Table 1: Our results and comparison with first place.

	MAP				**Precision**			**RR**		
	@1	**@3**	**@5**	**-**	**@1**	**@3**	**@5**	**@1**	**@3**	**@5**
Buster.AI(1st)	**0.897**	**0.926**	**0.929**	**0.929**	**0.895**	**0.32**	**0.195**	**0.895**	**0.923**	**0.927**
trueman(6th)	0.743	0.768	0.773	0.782	0.74	0.267	0.164	0.74	0.766	0.771
trueman (unsubmitted)	0.757	0.797	0.80	0.808	0.759	0.283	0.173	0.759	0.798	0.802

3.1 Analysis

We manually analysed our results and found that our system made errors in the following cases:

Proper Noun Overlap: We observed that our system sometimes fails and returns claims that do not have the same proper nouns as the query and sometimes lets some proper nouns slip by. Table 2 shows some examples of the faulty results vis-à-vis proper noun overlap.

Table 2: **bold** overlapped proper nouns; *italics* missed proper nouns

Tweet Query (id)	VClaim (id)	Model's pick (id)
@HydroxCookie where are you're cookies made? *Oreo* moved to *Mexico*. Made in **USA** wins my business. (1055)	Nabisco closed their Chicago plant and moved all production of *Oreos* to *Mexico*. (6592)	Japan renamed a town 'Usa' so that it could legitimately stamp its exports 'Made in **USA**.' (5891)
Good to see the *Nishioka* shot being replayed again and again... that's what we should be talking about... @MikeCTennis (1051)	Video shows an amazing behind-the-back return by Japanese tennis player Yoshihito *Nishioka*. (10082)	The comedian and actor Tim Allen wrote a lengthy Facebook post that attacked liberals and Democratic politicians and was shared widely in August 2019. (9254)

Hyperlinks: We found that in many cases, the textual information in a tweet is very vague and crucial information regarding the tweet is contained in a hyperlink. So, to solve this problem we sought to extract article titles from the hyperlinks. Some erroneous results dealing with hyperlinks are shown in Table 3.

Hashtags (#) and At (@) sign: Upon analysis, we found out that our system was having major trouble if crucial proper nouns were contained inside a hashtag or at (@) sign. Some sample tweets showcasing this problem are tabularised below in Table 4.

3.2 Improvements to our submitted system:

Learning from the above analysis we have added a new module to our submitted system that takes removal of special symbols and noun overlap into account and accordingly updates the similarity scores. Since more often than not people put crucial (for retrieval purposes) information inside hashtags and at signs,

Table 3: Many times information needed to verify a claim is in the hyperlink

Tweet Query (id)	**VClaim (id)**	**Model's pick (id)**
Is @jacindaardern willing to denounce this legislation of child sexual abuse? **https://t.co/6YMlJiO8zr** (1013)	In August 2018, French politicians passed a law which stated that a child is capable of consenting to having sex with an adult. (5691)	The U.S.'s leading group of pediatricians issued a strong statement condemning tolerance of gender dysphoria in children. (676)
Three. million. gallons. Our story from Colorado on the Animas River disaster: **http://t.co/s40r7orNj4** pic.twitter.com/Ey26EaBEhK (1048)	A spill by the Environmental Protection Agency rendered the normally pristine blue Animas River a terrifying mustard yellow. (2558)	A new study reporting on the 2015 death of a Colorado infant claims the event was the world's first documented pot overdose. (1838)
Holy crap. I have never, in my entire career as an ant researcher, seen *anything* like this. **https://t.co/jIjTOo3fZc** (1018)	There are small islands of fire ants floating in the floodwaters from Tropical Storm Harvey. (716)	Actor Kurt Russell said that he has never seen a man as dedicated and determined as President Trump. (3125)

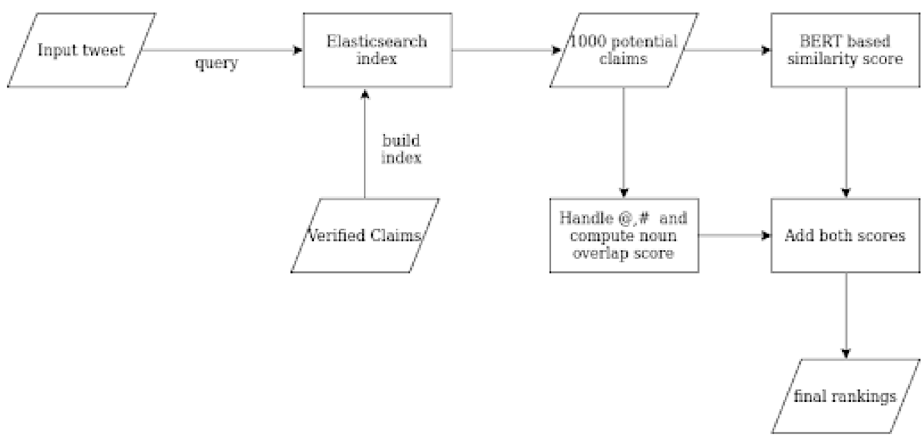

Fig. 2: Improved (unsubmitted) system

in our improved system we separated the words contained in these signs using regular expressions and prioritised them when checking for overlaps. Ideally, overlap of all proper nouns should be prioritised and to that end we added a

Table 4: Errors due to at (@) symbol and hashtags (#)

Tweet Query (id)	**VClaim (id)**	**Model's pick (id)**
@italiaricci You should get one for your house! **#PizzaVendingMachine #NowIWantOne** pic.twitter.com/3SV5Z 9bAuX (1016)	You can now order pizza from pizza vending machines. (10315)	On 30 July 2008, The Cheesecake Factory restaurants will be selling cheesecake for $1.50 per slice. Cheesecake Factory Serves Up a Delicious 30th Anniversary Celebration (6761).
We booked the one airline that doesn't give military free bags ???? **@SpiritAirlines** (1045)	A Spirit Airlines employee was rude to a soldier, charged him for a carry-on bag, and told his father that the airline 'doesn't cater to the military.' (1177)	An airline promotion allows husbands to take their wives along on business trips for free, but a survey later conducted by the airline finds that 95% of the wives were unaware of the promotion. (3391))
#Best news you'll hear all day! **#ScarfaceRemake**, starring non other than **@LeoDiCaprio** announced for 2016!! **#CantWait** (1084).	'Scarface' is being remade with Leonardo DiCaprio cast as Tony Montana. (815)	A live poll conducted by ABC News in August 2016 shows Donald Trump, Jill Stein, and Gary Johnson all well ahead of Hillary Clinton. (1638)

module to reward proper noun overlaps. We calculate the overlap using Levenshtein distance [8].The more overlapped proper nouns in tweet and claim, the higher scores reward. Table 1 shows performance of our improved system trueman(unsubmitted), we can see our new system performs slightly better than our submitted system on all evaluation metrics.

4 Conclusion and future work

In this paper we proposed a system for retrieval of verified claims given a query and submitted our results presented our results at CLEF-2020 Check That Lab!'s Claim Retrieval task. We strove to resolve some of the issues that our system faced initially by annotating and analysing faulty results, and managed to improve our performance. Yet, we strongly believe that there is a margin for future work to resolve some existing problems and to broaden the model's knowledge source in order to achieve practical application. The problem that the model was facing in regard to hyperlinks still needs some work in order to achieve full resolution. Since most of the links shared were from news or media websites, we think that title and byline retrieval from the hyperlinks will improve the system's performance considerably. Also, inculcating Image modality to analyse tweets with pictures merits future work in this regard. For the sake of this research, we considered fact-checking websites as a knowledge source in order to validate claims. We aspire to do the same and regard fact-checking and investigative journalism websites as a knowledge source but in an Indian context. This approach will improve our model's contextual sensibility and improve chances of identifying factual incorrectness and misinformation significantly. We believe that these approaches are good enough to merit significant future works and further research.

References

[1] Naeemul Hassan et al. "The quest to automate fact-checking". In: *Proceedings of the 2015 Computation+ Journalism Symposium*. 2015.

[2] James Thorne and Andreas Vlachos. "Automated Fact Checking: Task Formulations, Methods and Future Directions". In: *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 3346–3359. URL: `https://www.aclweb.org/anthology/C18-1283`.

[3] Georgi Karadzhov et al. "Fully Automated Fact Checking Using External Sources". In: *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*. Varna, Bulgaria: INCOMA Ltd., Sept. 2017, pp. 344–353. DOI: `10.26615/978-954-452-049-6_046`. URL: `https://doi.org/10.26615/978-954-452-049-6_046`.

[4] Shaden Shaar et al. "That is a Known Lie: Detecting Previously Fact-Checked Claims". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, July 2020, pp. 3607–3618. DOI: `10.18653/v1/2020.acl-main.332`. URL: `https://www.aclweb.org/anthology/2020.acl-main.332`.

[5] Alberto Barron-Cedeno et al. *Overview of CheckThat! 2020: Automatic Identification and Verification of Claims in Social Media*. 2020. arXiv: `2007.07997 [cs.CL]`.

[6] Shaden Shaar et al. "Overview of CheckThat! 2020 English: Automatic Identification and Verification of Claims in Social Media". In:

[7] Stephen Robertson and Hugo Zaragoza. "The Probabilistic Relevance Framework: BM25 and Beyond". In: *Foundations and Trends in Information Retrieval* 3 (Jan. 2009), pp. 333–389. DOI: `10.1561/1500000019`.

[8] L. Yujian and L. Bo. "A Normalized Levenshtein Distance Metric". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.6 (2007), pp. 1091–1095.