End-to-End Time-Sensitive Fact Check

Wee Yong Lim
National University of Singapore
weeyong.lim@u.nus.edu

Mong Li Lee
National University of Singapore
leeml@comp.nus.edu.sg

Wynne Hsu
National University of Singapore
whsu@comp.nus.edu.sg

ABSTRACT

Online social networks have become a prime target for the spread of fake news and misinformation. Despite advances in tools and techniques for the detection and verification of fake news, assessing the credibility of information remains a challenge. Previous work have extracted claims from social media post for credibility assessment. These claims typically describe the object, subject and predicate, and have no notion of time. We observe that the credibility of claims depends not just on the content, but also on the time period that the claim is purported to be valid for. We develop an end-to-end framework for evaluating time-sensitive claims. The framework generates alternate claims and takes into consideration relationships between these alternate claims and the target claim to performs a joint credibility assessment. Experiments results on two datasets shows the effectiveness of the proposed framework to increase the accuracy of claim assessment.

1 INTRODUCTION

The spread of misinformation and allegations of fake news on online platforms is detrimental to society when readers are unable to confidently discern the credibility of the news. While platforms such as Twitter are particularly useful for posting timely update of events on the ground, misinformation can also arise from malicious intents or inaccuracies due to the dynamic nature of evolving events.

Research to identify rumours has focused on identifying features based on tweet contents, user profiles and propagation patterns to train classifiers for discriminating the veracity of individual tweets [4] or tweet clusters [2, 11, 17, 21]. More recently, data-driven models have been explored to learn hidden features from the text content of social posts [12, 22]. All these works do not consider the time information of the posts.

The work in [9] extracts claims from tweet contents, and represents a claim as a triplet comprising of <subject, predicate, object> to depict a unit of information for which a credibility label can be attributed unambiguously. However, we observe that this representation is insufficient as the credibility of a claim depends not just on the content, but also on the time period and location that the claim is purported to be valid for.

For example, the claim <*Romelu Lukaku, played for, Chelsea*> indicating that the football player Romelu Lukaku has played for the club Chelsea. This claim is valid only between 2011 and 2012[1]. This motivates us to extend the definition of a claim to capture temporal and spatial information. The spatio-temporal information affects the validity of the claims and can be used to guide their credibility assessment.

At the same time, given a target claim, we can generate alternate claims that share the same common subject and predicate, but have different object. These alternate claims may have some *exclusive relationship* with the target claim and would allow us to jointly review their credibility. For example, suppose we have a target claim that <*Romelu Lukaku, played for, Manchester United, 2016*>, and we have the following two alternate claims:

(1) <*Romelu Lukaku, played for, Everton, 2016*>
(2) <*Romelu Lukaku, played for, Chelsea, 2016*>

If we can confirm that Romelu played for *Everton* in 2016, then it would allow us to debunk the target claim since the same player cannot simultaneously play for multiple clubs.

In this work, we design an end-to-end framework that considers spatio-temporal information in the verification of claims in social media. Our framework will utilize the web search results of a target claim to generate possible alternate claims. We identify potential relationships among the target and alternate claims, and utilize these relationship to perform joint assessment of their credibilities using probabilistic soft logic. Experiments on two datasets of time-sensitive claims demonstrate the effectiveness of the use of alternate claims in assessing the credibility of the target claims.

2 PROPOSED FRAMEWORK

Our proposed framework takes as input a time-sensitive claim which is a quintuple $(s, p, o, g, [t_1, t_2])$, where s is the clause for the subject, o is the clause for the object, p is a predicate or property between the subject and object, g is the geographical location, time period $t = [t_1, t_2]$ refers to the begin and end time interval for the validity of the claim.

An interactive framework was proposed in [10] to gather evidence from web search results for the credibility assessment of a target claim. A claim is determined to be *credible, not credible* or *inconclusive* depending on the level of supporting evidence. Here, we extend the framework to include the following key components:

(1) Generate alternate claims from web search results
(2) Identify relationships among the target and alternate claims
(3) Assess credibility of the target claim taking into account the relationships between the target and alternate claims

Figure 1 gives an overview of proposed framework. The following subsections give the details of the key components.

2.1 Generate Alternate Claims

A claim is represented as a relation tuple with temporal information. We start with a target claim and search for alternate claims that differ from the target claim in terms of their subject/object as well as the time period.

[1]https://en.wikipedia.org/wiki/Romelu_Lukaku

ROME 2019, July 25, 2019, Paris, France

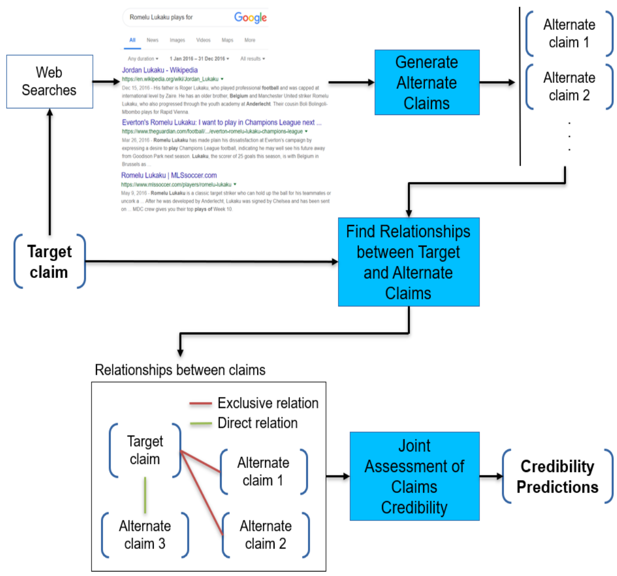


Figure 1: Overall Framework for Time-Sensitive Claims

The work in [8] designed a method called T-verifier to find alternative terms in a statement with a specified doubt unit. Here, we use the subject or object in a target claim as the doubt unit. We carry out a web search using a query that contains the words and time information in the claim, less the doubt unit. For example, given a target claim <Romelu Lukaku, plays for, Manchester United, UK, [2016, 2017]>, and suppose the object football club is the doubt unit, we can construct a web query *"Romelu Lukaku plays for"* and limit the web search results to the time period [2016, 2017].

Each web search result (WSR) has a title and snippet. We use a named entity recognition tool [13] to extract named entities which have the same type as the doubt unit, e.g., Person, Organization, Location. We also extract the associated dates by resolving for possible date[2] in the WSR. Entities that are not associated with dates within the search time period are discarded.

We rank the extracted entities based on their coverage in the set of WSR retrieved. The coverage of e is given by the fraction of WSR that contains e. Let W be the set of WSR retrieved and $contains(e, w)$ returns 1 if the WSR $w \in W$ contains the entity e. Then we can define coverage of an entity e in the set W as

$$Coverage(e, W) = \sum_{w \in W} \frac{contains(e, w)}{|W|} \quad (1)$$

The top-ranked entities are used to substitute the doubt unit in the target claim to form alternate claims. Note that the alternate claims have the same predicate as the target claim.

[2]http://datefinder.readthedocs.io/en/latest/

2.2 Relationship between Target and Alternate Claims

After generating alternate claims for a target claim, the next step is to identify whether they have some mutually exclusive or direct relationship with the target claim.

We say that a pair of claims has a *mutually exclusive relationship* if they refer to situations that cannot occur simultaneously. For example, if the same person cannot be physically present in two different places at the same time.

On the other hand, a pair of claims has a *direct relationship* if they refer to situations that reinforce one another. For example, two claims about the same subject/object at nearby location in consecutive time periods increases the likelihood that these claims are true.

Identifying these relationships is not a trivial task for the following reasons. First, knowing whether two claims refer to the same person requires entity resolution since a person's name can have multiple variations or the same person can be referenced by his action, title or office. For example "Donald Trump", "President Trump" and "United States President" all refer to the same person. We address this issue by querying the claims' subjects/objects against the Wikidata knowledge base to retrieve their corresponding entries in the knowledge base. If the results retrieved are the same, we conclude that the two claims refer to the same person.

Second, location entities may have different granularity in a pair of claims, e.g. one claim may consider the city "Ottawa" while the other claim may mention the country "Canada". Since location entities are organized in a hierarchy, we augment the extracted location entities in the claims by their least common ancestor. In our example, we augment "Canada" to the first claim and determine that these two claims are likely to refer to the same location.

Algorithm 1 finds a set of potential relationships between claims based on the spatial-temporal exclusivity of the entities in the claims. The input is a target claim c and its set of alternate claims A. For each $c' \in A$, the algorithm checks whether c and c' refer to the same entity (Lines 2-3). If they refer to the same entity, we then check whether there is temporal overlap in c and c' (Line 4). If c and c' have overlapping time periods, then the algorithm will check if the claims share any common location (Line 5). If so, an *exclusive* relationship between c and c' is added to the result (Line 6). Otherwise, if c and c' have consecutive time periods, then a *direct* relationship between c and c' is added to the result (Lines 8-9). The algorithm returns the set of pairwise relationships found between the target claim and its alternate claims.

2.3 Joint Assessment of Claims Credibility

Having identified the relationships between the target and alternate claims, we use probabilistic soft logic (PSL) to jointly assess their credibilities [1, 18].

We perform a web search using the target claim as query and obtain an initial estimate of the credibility of the target claim. This estimate is derived from the features of the web search results such as fraction of relevant WSR from reputable sites, fraction of supporting WSR. Similarly, for each alternate claim, we also obtain an initial estimate of its credibility from its web search result. Then we run PSL to take into account the exclusive and direct

Algorithm 1 Find Relationship between Claims

Input: Target claim c and its set of alternate claims A
Output: Set of pairwise relationships R

$R = \emptyset$
for each claim $c' \in A$ **do**
  if $resolvePersons(c) \cap resolvePersons(c') \neq \emptyset$ **then**
    if $temporalOverlap(c, c')$ **then**
      if $getLocations(c_i) \cap getLocations(c_j) \neq \emptyset$ **then**
        $R \leftarrow R \cup \{exclusive(c, c')\}$
      end if
    else if $temporalConsecutive(c, c')$
      $R \leftarrow R \cup \{direct(c, c')\}$
    end if
  end if
end for

relationships between the target and alternate claims to adjust their initial estimates.

For a claim c, the initial estimates for the classes $\texttt{CR}$ (Credible), $\texttt{NC}$ (Not Credible) or $\texttt{IC}$ (InConclusive) are generated based on features extracted from its WSRs evidence. These are expressed as soft observations, $evidence(c, class)$, with values in the range [0,1]. The estimate of a claim, $evidence(c, class)$, directly affects its final estimate $likely(c, class)$ via the following first order logic rule:

$$evidence(c, class) \rightarrow likely(c, class) \tag{2}$$

Next, we use logic rules to express the relationships between claims. If a pair of claims (c_1, c_2) has a direct relationship, denoted by $direct(c_1, c_2)$, then c_2 is likely to be credible when c_1 is deemed credible, and vice versa. Similarly, c_2 is likely to be not credible (or inconclusive) when c_1 is deemed to be not credible (or inconclusive). This is captured by the following rule:

$$direct(c_1, c_2) \wedge evidence(c_1, class) \rightarrow likely(c_2, class) \tag{3}$$

On the other hand, if a pair of claims (c_1, c_2) has an exclusive relationship, denoted by $exclusive(c_1, c_2)$, then it is not possible for both c_1 and c_2 to be $\texttt{credible}$ at the same time. The claim that has weaker evidence for being credible will be penalised by having its likelihood for the classs $\texttt{CR}$ lowered. This is summarized by:

$$\begin{aligned} exclusive(c_1, c_2) \wedge (evidence(c_1, \texttt{CR}) > evidence(c_2, \texttt{CR})) \\ \rightarrow \neg likely(c_2, \texttt{CR}) \end{aligned} \tag{4}$$

For each claim, the $likely$ soft truth values for all the classes must sum to 1. In order to infer the truth value for the $likely$ atoms, rules are grounded with candidate values for the $likely$ atoms and observations for the rest of the atoms. A ground rule is satisfied only if the truth value of the head of the rule is more than or equal to the truth value for the body of the rule. The probabilistic soft logic uses the Most Probable Explanation (MPE) inference algorithm to maximise the number of ground rules satisfied, taking into account the exclusive and direct relationships between claims to generate a final credibility estimation for the claims.

3 EXPERIMENT EVALUATION

We carry out experiments to evaluate the effectiveness of the proposed framework to debunk time-sensitive rumours. we use the Kaggle dataset of football transfers from 2000 to 2018 given that there are usually a lot of rumours surrounding football players signing on various football clubs.

3.1 Effectiveness of Alternate Claim Generation

The objective of generating alternate claims is to identify claims that may have some exclusive relationship with the target claim, which will facilitate the verification of the target claim credibility.

We first generate 142 target claims which are rumours of football players who have signed with some football clubs between 2014 to 2018. For each target claim, we assume that the object football club is the doubt unit and issue a web query consisting of the footballer name as the object, and "plays for" as the predicate.

We rank the alternative units based on their coverage (recall Equation (1)) and construct alternate claims. We match the alternate claims generated with the ones in the Kaggle football transfer dataset. Figure 2 shows the number of correct matches as we vary the number of top-k alternate units used to construct the alternate claims. For comparison, we also generate alternate claims by simply using the first k alternate units returned by the web search, in other words, there is no ranking involved. We observe that our coverage-based ranking is more effective in generating a list of alternate units that leads to alternate claims are are true. Generating alternate claims that are true increases the ability of the proposed framework to classify if a target claim is true or not, as shown in the subsequent experiments.

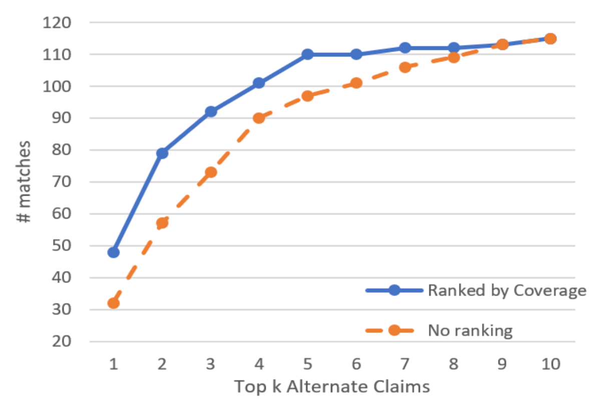


Figure 2: Correct matches for top-k alternate claims.

3.2 Effectiveness of Credibility Assessment

In this section, we evaluate the effectiveness of the framework to classify the credibility of time-sensitive claims.

We create 20 target claims (10 credible and 10 not credible) from the Kaggle football dataset (see Table 1). For each target claim, we generate alternate claims using the predicates "will play for" and "play for". Alternate claim with the predicate "play for" has the same time period as the target claim, thus establishing an exclusive relationship with the target claim since the same football player cannot play for different clubs in same time period. On the other hand, the alternate claim with predicate "will play for" uses the

	Target claim	Credible?
F1	<Tiemoue Bakayoko, will play for, Chelsea, 2016-07-08, 2017-07-08>	Y
F2	<Fernando Torres, will play for, Chelsea, 2010-01-24, 2011-01-24>	Y
F3	<Roque Mesa, will play for, Sevilla, 2017-01-23, 2018-01-23>	Y
F4	<Philippe Coutinho, will play for, Barcelona, 2017-01-01, 2018-01-01>	Y
F5	<Iago Aspas, will play for, Sevilla, 2013-07-07, 2014-07-07>	Y
F6	<Diego Costa, will play for, Atletico Madrid, 2016-09-14, 2017-09-14>	Y
F7	<Dimitar Berbatov, will play for, Fulham, 2011-08-24, 2012-08-24>	Y
F8	<Steven Nzonzi, will play for, Sevilla, 2014-07-02, 2015-07-02>	Y
F9	<Kyle Walker, will play for, Manchester City, 2016-07-07, 2017-07-07>	Y
F10	<Daley Blind, will play for, AFC Ajax, 2017-07-10, 2018-07-10>	Y
F11	<Kevin Gameiro, will play for, Newcastle, 2010-06-03, 2011-06-03>	N
F12	<Aleksandar Mitrovic, will play for, FC Porto, 2014-07-14, 2015-07-14>	N
F13	<Esteban Granero, will play for, Fiorentina, 2012-08-08, 2013-08-08>	N
F14	<Gaston Ramirez, will play for, Benfica, 2016-07-28, 2017-07-28>	N
F15	<Jorginho, will play for, Manchester City, 2017-07-07, 2018-07-07>	N
F16	<Romelu Lukaku, will play for, Chelsea, 2016-07-03, 2017-07-03>	N
F17	<Benjamin Mendy, will play for, Chelsea, 2016-07-17, 2017-07-17>	N
F18	<Leroy Fer, will play for, Sunderland, 2015-01-25, 2016-01-25>	N
F19	<Neymar, will play for, Manchester United, 2016-07-27, 2017-07-27>	N
F20	<Jannik Vestergaard, will play for, Tottenham, 2017-07-06, 2018-07-06>	N

Table 1: Target claims created from the Kaggle Football dataset.

previous year as its time period. This creates a direct relationship with the target claim as any speculation in the previous year that a player will play for a club increases the likelihood that he will actually play for that club the following year.

We use the target and alternate claims as queries to retrieve web search results, and obtain their initial credibility estimates. Then we run PSL to take into account the exclusive and direct relationships to adjust these credibility estimates.

We first analyze the results for target claims which are credible. Table 2 shows the initial classifications based on web search results, and the changes in these classifications after applying our proposed framework. We observe that the web evidence could only give the correct classification for 3 out of the 10 claims. Half of the target claims were deemed inconclusive, with another 2 wrongly classified to be not credible. In contrast, our framework is able to correctly classify 9 out of 10 claims after we take into consideration the relationships between the target and alternate claims.

	Credible	Not Credible	Inconclusive
Initial classification	3	2	5
Our framework	**9**	1	0

Table 2: Classification of credible Football claims.

Next, we examine the results for target claims which are not credible. Table 3 shows that the majority of these claims (8 out of 10) are initially classified as inconclusive, suggesting that the web search results do not provide sufficient evidence to determine the credibility of these claims. However, the additional evidence from the alternate claims enables our framework to adjust the credibility assessment, and correctly classify 6 of them to be not credible.

	Credible	Not Credible	Inconclusive
Initial Classification	1	1	8
Our framework	1	**6**	3

Table 3: Classification of not credible Football claims.

3.3 Generalizability Evaluation

Finally, we demonstrate the generalizability of our framework on a second dataset consisting of various company acquisition actions, mainly in the technology and pharmaceutical industries.

We create 20 target claims (10 credible and 10 not credible). For each target claim, we generate alternate claims using the predicates "to acquire" and "acquired". For example, for the target claim *(Facebook, to acquire, WhatsApp [2013-02-12, 2014-02-12])*, the alternate claim about another potential buyer is generated:

<Google, to acquire, WhatsApp [2013-02-12, 2014-02-12]>

Alternate claims after the acquisition event are also generated:

<Facebook, acquired, WhatsApp [2014-02-26, 2015-02-26]>

<Google, acquired, WhatsApp [2014-02-26, 2015-02-26]> Given that a company cannot be acquired by two different companies at the time period, alternate claims with the predicate "acquired" will have an exclusive relationship with the target claim. Alternate claims with the predicate "to acquire" in the previous time period will have a direct relationship with the target claim.

We obtain an initial credibility estimates for the target and alternate claims from their web search results. Then we run PSL to take into account the exclusive and direct relationships. Table 5 shows the initial classification for the credible target claims, and the results after applying our framework. We see that the alternate claims in our framework are able to help verify that 3 of the inconclusive claims are in fact credible.

	Target claim	Credible?
A1	<Microsoft, to acquire, LinkedIn, 2015-12-01, 2016-12-01>	Y
A2	<Microsoft, to acquire, Github, 2017-05-28, 2018-05-28>	Y
A3	<Facebook, to acquire, WhatsApp, 2013-02-12, 2014-02-12>	Y
A4	<Google, to acquire, Motorola Mobiity, 2010-08-08, 2011-08-08>	Y
A5	<Microsoft, to acquire, Skype, 2010-05-03, 2011-05-03>	Y
A6	<Eric Carreel, to acquire, Nokia Health, 2017-05-24, 2018-05-24>	Y
A7	<IBM, to acquire, Red Hat, 2017-10-21, 2018-10-21>	Y
A8	<IBM, to acquire, SoftLayer, 2012-05-28, 2013-05-28>	Y
A9	<ON Semiconductor, to acquire, Fairchild Semiconductor, 2014-11-11, 2015-11-11>	Y
A10	<Google, to acquire, Waze, 2012-06-04, 2013-06-04>	Y
A11	<Allergan, to acquire, Salix, 2014-02-15, 2015-02-15>	N
A12	<Comcast, to acquire, Sprint, 2017-04-21, 2018-04-21>	N
A13	<Intel, to acquire, Mellanox, 2018-03-11, 2019-03-11>	N
A14	<Johnson Johnson, to acquire, Pharmacyclics, 2014-02-26, 2015-02-26>	N
A15	<Pfizer, to acquire, Onyx, 2012-08-18, 2013-08-18>	N
A16	<Sanofi, to acquire, Medivation, 2015-08-15, 2016-08-15>	N
A17	<Valeant, to acquire, Allergan, 2013-11-10, 2014-11-10>	N
A18	<Roche, to acquire, Tesaro, 2017-11-26, 2018-11-26>	N
A19	<Sanofi, to acquire, Actelion, 2016-01-19, 2017-01-19>	N
A20	<Reckitt Benckiser, to acquire, Merck Consumer Health, 2013-04-30, 2014-04-30>	N

Table 4: Target claims created from the Acquisition dataset.

	Credible	Not Credible	Inconclusive
Initial classification	6	0	4
Our framework	**9**	0	1

Table 5: Classification of credible Acquisition claims.

Table 6 shows the results for the not credible Acquisition claims. We see that 9 of the 10 claims are initially deemed to be inconclusive, again suggesting that using only web search results is not sufficient to determine the claims' credibility. However, after we take into consideration the relationships between the target and alternate claims, we are able to correctly classify 5 of these inconclusive claims to be not credible.

	Credible	Not Credible	Inconclusive
Initial classification	1	0	9
Our framework	1	**5**	4

Table 6: Classification of not credible Acquisition claims.

We observe consistent experiment results for the claims derived from both the Football and Acquisition data, despite the different domains. This indicates that the use of alternate claims can be an effective means to assess the credibility of time-sensitive claims.

3.4 Case Study

In this section, we discuss two cases from the Acquisition data. The first case shows how an alternate claim is helpful in enriching the evidence and enables the framework to arrive at the correct credibility assessment for a target claim. The second case provides some insights into why an alternate claim did not help the framework to rectify a wrongly classified target claim.

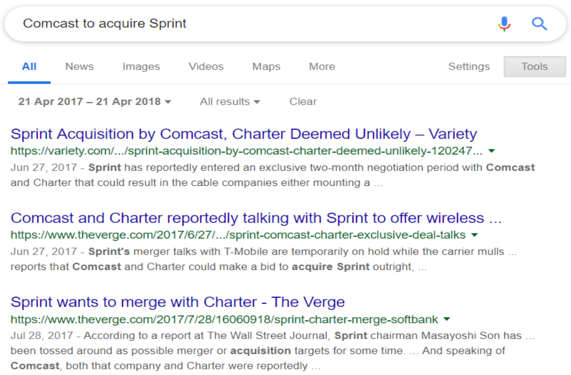
Fig. 3(a) shows a sample of the web search results for the target claim *(Comcast, acquire, Sprint)*. It appears that Sprint was sought after not only by Comcast, but also Charter. One web result reportedly mention that Sprint wants to merge with Charter. Given the conflicting evidence, this target claim was initially deemed to be inconclusive. Fig. 3(b) shows the web search results for the alternate claim *(T-mobile, acquire, Sprint)* which provides consistent evidence that supports T-mobile's intention to acquire Sprint during the same time period. In light of the exclusive relationship between these two claims, and the evidence supporting the alternate claim is stronger compared to the evidence for the target claim, our framework is able to correctly classify the target claim as not credible.

Fig. 4(a) shows a sample of the web search results for the target claim *(Roche, acquire, Tesaro)*. There is large number of articles that speculated Roche's intention to acquire Tesaro. Given this overwhelming number of WSR that supports the target claim, it is wrongly classified as credible. In contrast, there is hardly any article for the alternate claim *(GSK, acquire, Tesaro)* (see Fig. 4(b)) even though GSK has actually acquired Tesaro. One possible explanation is that GSK has successfully kept secret this acquisition until the deal has been completed. As a result, the framework is not able to debunk that rumour *(Roche, acquire, Tesaro)*.

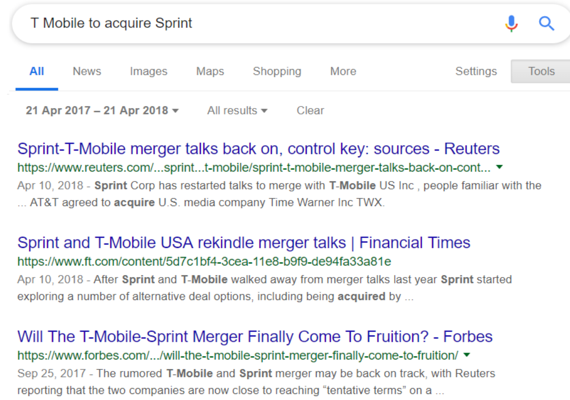
4 RELATED WORK

There has been much research on the credibility assessment of claims [3]. [16] introduces a crowd sourcing platform called Verily that incentivises participants to provide evidence for the credibility of claims. [19] combats fake news by recommending fact-checking URLs to users, while [6] uses an interactive platform to assist users determine the veracity of Twitter news accounts.

Web search results have also been used to help assess the credibility of claims. Multiverifier determines the truthfulness of a statement by utilizing the top-n search results that are most related to

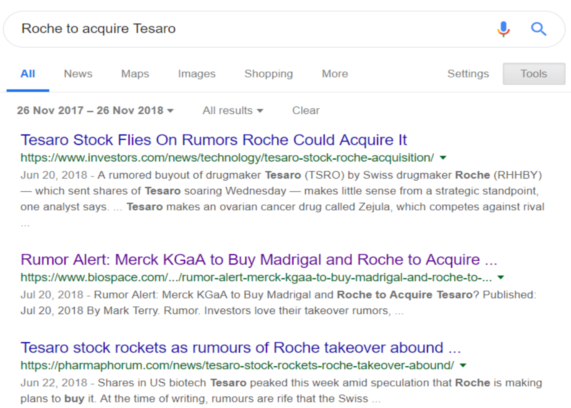


(a) WSR for target claim *(Comcast, acquire, Sprint)*

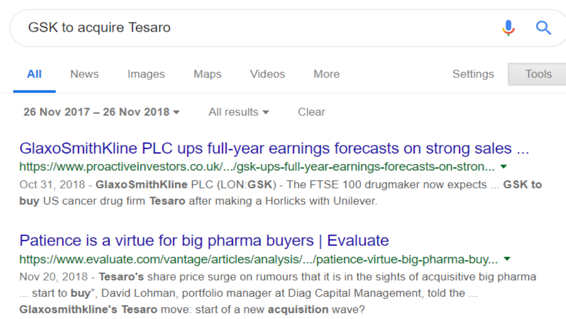


(b) WSR for alternate claim *(T-mobile, acquire, Sprint)*

Figure 3: An inconclusive claim correctly classified to be not credible.



(a) WSR for target claim *(Roche, acquire, Tesaro)*



(b) WSR for alternate claim *(GSK, acquire, Tesaro)*

Figure 4: An incorrectly classified claim that remains uncorrected due to overwhelming rumours on the web.

the statement [20]. ClaimBuster [5] tries to match claims against previously verified claims before carrying out a web search. An interactive framework called iFact collects evidence from web search results for the verification of claims, and utilizes direct and inverse relationships between claims to obtain a more consistent credibility assessment of claims [10]. CredEye [15] provides users with evidence about the credibility of a claim by classifying the stance of web articles snippets containing overlapping words with the claim. All these works do not consider temporal information.

Several works have focused on the generation of alternate claims [8, 14]. T-verifier [8] considers a statement as consisting of a topic unit and a doubt unit. The topic unit is used as query to a web search engine, and the top-5 alternate units are used to substitute the doubt unit to form alternative statements. In contrast, [14] takes as input a tuple *(subject, verb, object)* and uses a knowledge base of relation tuples to heuristically decide whether the subject or object is the doubt unit. These works do not extract the date and geographical location to construct alternate claims.

The use of contextual information has been shown to be useful for assessing claims' credibility. [7] integrates temporal validity and provenance information to process claims in the form of conjunctive queries. The authors do not use relationships between claims to influence the credibility assessment of claims. In contrast, our work makes use of spatial-temporal exclusivity of entities to identify relationships between claims about the same entities. Such relations are expressed in logic rules, to infer credibility predictions of the claims.

5 CONCLUSION

In this work, we described an end-to-end framework for assessing the credibility of time-sensitive claims. We designed a method for generating alternate claims that differs from a target claim in some doubt unit such as the subject, object or time period. With this, we can leverage on the exclusive and direct relationships between a target claim and its alternate claims to perform a joint credibility assessment. Experiment results demonstrated the effectiveness of the proposed framework to generate alternate claims and enhance the web evidence for the verification of a target claim. Future work includes exploring other relationships between claims such as consistency, aligned, containment and succession.

REFERENCES

[1] S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. Hinge-loss markov random fields and probabilistic soft logic. *Journal of Machine Learning Research*, 2015.

[2] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web (WWW)*, 2011.

[3] S. Cazalens, P. Lamarre, J. Leblay, I. Manolescu, and X. Tannier. A content management perspective on fact-checking. In *Companion Proceedings of The Web Conference*, 2018.

[4] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier. Tweetcred: Real-time credibility assessment of content on twitter. In *Social Informatics*, 2014.

[5] N. Hassan, G. Zhang, F. Arslan, J. Caraballo, D. Jimenez, S. Gawsane, S. Hasan, M. Joseph, A. Kulkarni, A. K. Nayak, V. Sable, C. Li, and M. Tremayne. Claimbuster: The first-ever end-to-end fact-checking system. *Proceedings of the VLDB Endowment*, 2017.

[6] A. Karduni, R. Wesslen, S. Santhanam, I. Cho, S. Volkova, D. Arendt, S. Shaikh, and W. Dou. Can you verifi this? Studying uncertainty and decision-making about misinformation using visual analytics. In *International AAAI Conference on Web and Social Media (ICWSM)*, 2018.

[7] J. Leblay. A declarative approach to data-driven fact checking. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

[8] X. Li, W. Meng, and C. Yu. T-verifier: Verifying truthfulness of fact statements. In *Proceedings of the 27th IEEE International Conference on Data Engineering (ICDE)*, 2011.

[9] W. Y. Lim, M. L. Lee, and W. Hsu. Claimfinder: A framework for identifying claims in microblogs. In *6th International Workshop on Making Sense of Microposts (#Microposts), in conjunction with WWW Conference*, 2016.

[10] W. Y. Lim, M. L. Lee, and W. Hsu. iFact: An interactive framework to assess claims from tweets. In *Proceedings of the ACM on Conference on Information and Knowledge Management (CIKM)*, 2017.

[11] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM)*, 2015.

[12] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.

[13] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, 2014.

[14] N. Nakashole and T. M. Mitchell. Language-aware truth assessment of fact candidates. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014.

[15] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum. Credeye: A credibility lens for analyzing and explaining misinformation. In *Companion Proceedings of the The Web Conference*, 2018.

[16] A. Popoola, D. Krasnoshtan, A.-P. Toth, V. Naroditskiy, C. Castillo, P. Meier, and I. Rahwan. Information verification during natural disasters. In *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, 2013.

[17] V. Qazvinian, E. Rosengren, and D. Radev et al. Rumor has it: Identifying misinformation in microblogs. In *International Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011.

[18] M. Samadi, P. Talukda, M. M. Veloso, and M. Blum. ClaimEval: Integrated and flexible framework for claim evaluation using credibility of sources. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

[19] N. Vo and K. Lee. The rise of guardians: Fact-checking url recommendation to combat fake news. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018.

[20] T. Wang, Q. Zhu, and S. Wang. Multi-verifier: A novel method for fact statement verification. *World Wide Web*, 2015.

[21] F. Yang, Y. Liu, X. Yu, and M. Yang. Automatic detection of rumor on sina weibo. In *ACM SIGKDD Workshop on Mining Data Semantics*, 2012.

[22] F. Yu, Q. Liu, S. Wu, L. Wang, and T. Tan. A convolutional approach for misinformation identification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.