Spatio-temporal grounding of claims made on the web, in PHEME

Leon Derczynski and Kalina Bontcheva

University of Sheffield
S1 4DP, UK
leon,kalina@dcs.shef.ac.uk

Abstract

Social media presents us with a digitally-accessible sample of all human discourse. This sample is full of claims and assertions. While the state of the art in NLP is adapting to the volume, velocity and variety of this sample and the information in it, the accuracy of claims made in social media remain largely unstudied. PHEME, a 36-month EU project started in January 2014, focuses on this fourth challenge: veracity. As a core part of establishing veracity, we need to identify the spatio-temporal context of assertions made on informal websites. This project note introduces the spatio-temporal challenges and planned semantic annotation activities that are part of the PHEME project.

1. Introduction

Social networks are rife with lies and deception, half-truths and facts. Irrespective of an assertion's truthfulness, the rapid spread of such information through social networks and other online media can have rapid and serious consequences. In such cases large amounts of user-generated content need to be analysed quickly, yet it is not currently possible to carry out such complex analyses in real time.

Social media poses three major computational challenges, dubbed the 3Vs of big data: volume, velocity, and variety (Laney, 2001). Content analysis methods have faced additional difficulties, arising from the short, noisy, and strongly contextualised nature of social media. In order to address the 3Vs of social media, new language technologies have emerged, e.g. using locality sensitive hashing to detect breaking news stories from media streams (volume), predicting stock market movements from microblog sentiment (velocity), and recommending blogs and news articles based on user content (variety).

PHEME[1] focuses on a fourth crucial challenge: veracity. It will model, identify, and verify **phemes** – internet memes annotated for truthfulness or deception – as they spread across media, languages, and social networks.

One of the many challenges in determining veracity is the automatic extraction of a claim's context. As well as understanding complex social context, it is critical to know when and where each claim was made, or to when and where it was intended to apply. This project note discusses the role of spatio-temporal information extraction and reasoning in solving this challenge.

2. Motivation

PHEME addresses the spatio-temporal validity of information and historical content to assess contradictions, through means of regional and longitudinal models of users, networks, trust, and influence.

The temporal delimitation of any assertion is of great importance, because the assertion is true only inside these bounds. Specifically, it is possible to extract two truths that seem to contradict (e.g. "The president of the USA is George W Bush" and "The president of the USA is Barack Obama") but are in fact both accurate when the appropriate temporal information is added. In other words, there is something like temporal validity of facts, which needs to be taken into account when detecting contradictions.

Similarly, assertions have spatial constraints, especially when they discuss underspecified entities. For example, we may say "The president is Obama" and "The president is Hollande"; these assertions seem to conflict, but are in fact both true simultaneously – just in distinct spatial regions.

It may not always be possible to ground assertions using single mentions of relations. Assertions may be spread over multiple documents, each mentioning different constraints. However, failing to determine the bounds of assertions – or assigning incorrect dates and places to claims – potentially leads to the rejection of correct information, reducing our overall ability to detect and ground/refute rumours in real-time. Spatio-temporal reasoning and inference offer solutions to these problems, and PHEME seeks to advance spatio-temporal relationship extraction to support measurement of veracity on the web.

3. Background

Unlike traditional news, a notable proportion of social media content posted online is explicitly geotagged (Sadilek et al., 2012), and studies suggest that it is possible to infer the geo-locations of about half of the remaining such content (Rout et al., 2013). Social media messages also have at least a creation time as temporal context. This implicit spatio-temporal (ST) metadata is not currently heavily exploited by modern NLP methods.

Given the constraint that a single entity can only be in any one place at a time, these forms of ST information give a means of determining the truthfulness of statements (Ji and Grishman, 2011; Derczynski and Gaizauskas, 2013).

Temporally, current systems are capable of detecting the publication date of documents (Chambers, 2012) and of grounding some of the time expressions contained therein (Strötgen and Gertz, 2010). Detecting events and assertions and temporally ordering these with regard to times is critical to ST grounding of facts and rumours; the state of the art in event detection is good (Kolya et al., 2012), but ordering events and times relative to each other

[1] The project is named after the Greek goddess Pheme, who was the personification of fame and renown; her favour being notability, her wrath being scandalous rumours.

or across documents remains an active area of novel research with some progress to be made. Fortunately, linking events to times – the most important type of temporal association for PHEME – is the task at which automated systems perform best (Derczynski, 2013).
Spatially, the challenge of grounding the locations in document content is critical to accurate bounding of claims. The state of the art is somewhat less mature than that of temporal context; while many tools can identify a range of named entities, recognition of new families of spatial entities (especially when general nouns are used in a spatial sense) is a subject of active research, e.g. Gaizauskas et al. (2012).
Spatio-temporal annotation in PHEME serves as one component in a complex system, linked together with longitudinal user behaviour modelling, information provenance, network structure, a-priori knowledge, and cross-media links.

4. Digital journalism

Journalists are currently using a plethora of social media applications in order to meet their diverse needs, e.g. Tweetdeck[2] for monitoring the social web; Storify[3] for news aggregation; crowdsourcing tools like Ushahidi's Swiftriver platform,[4] and online content filtering sites like Storyful.[5] The focus of all these tools is on getting the right content to journalists, but not on helping them with interpretation and verification of the authenticity and credibility of that content. Methods and tools vary according to the nature of the journalistic task, however. For example, observations of the Guardian newsroom (Procter et al., 2012) revealed that journalists prefer simple Twitter clients rather than more sophisticated tools such as Tweetdeck in activities such as live blogging. For reliability's sake, journalists prefer to rely on sources that their experience suggests they can trust. This solves the problem of reliability but limits their capacity to exploit social media to its full potential.
Spatio-temporal knowledge plays also an important role in this use case. A key challenge is to identify the regionality of events (e.g., neighbourhood, city, or country level) (Xu et al., 2012). Regionality is important because different events are relevant at varying scales, which impacts their newsworthiness and interestingness to digital journalists and users interested in local content.

5. Content Annotation

The project involves the creation of new language resources. These in turn helps create and evaluate general-purpose tools for projecting spatio-temporal annotations across languages, given parallel texts and re-using existing corpora (e.g. TimeBank, the multilingual TempEval, ACE2 temporal annotations, WikiwarsDE). The resources will be used to develop multilingual temporal annotation tools, based on their state-of-the-art techniques, developed for longer texts.
The project also addresses the problem of geo-locating events mentioned in documents. We intend to go beyond features based on words in the document, and use disambiguated URIs (e.g. against GeoNames[6]) and additional knowledge from the LOD resource (e.g. NUTS subdivisions, latitude/longitude, neighbouring locations).
Regarding annotation schemata, the de facto standards of ISO-TimeML (Pustejovsky et al., 2010) and ISO-Space (Pustejovsky et al., 2011) will be adopted and experimented with. Following Pustejovsky and Stubbs (2011), we intend to use temporal narrative containers for annotating events. In addition, recent adaptations of narrative containers to spatial annotation will be tried (Pustejovsky and Yocum, 2013). Narrative containers promise to lighten human annotator load while still capturing expressive representations of spatio-temporal information.
Standoff annotation may be required in some scenarios, as social media data typically has strict licensing constraints.
Existing standards provide a framework for annotating the factuality of assertions (e.g. Saurí and Pustejovsky (2009)), which can be applied over social media text in order to formalise the strength of assertions made there.
In the scope of rumour detection and analysis, Qazvinian et al. (2011) annotated messages for whether or not they related to a pre-determined rumour. PHEME involves two additional challenges: identifying the rumours in the first place, and then identifying the type of rumour from one of four classes: misinformation, disinformation, controversy and speculation. For the project, a "code frame" system is under development (Procter et al., 2013) for annotating topics and actor types. Code frames are specific to a research question that embodies information demand. Rumour messages are subdivided into categories, which may be related to claims, counter-claims or appeals for information; be with or without evidence; or simply rumour-relevant comments. Streams of related messages are categorised using code frames and annotated accordingly.

6. Project Contribution

PHEME aims to further the state of the art in spatio-temporal annotation and reasoning. In order to spatio-temporally ground asserions, PHEME will adapt existing annotation tools to social media data, through the creation of new training data in this genre. The project will also cover new target languages through lightly-curated annotation porting, taking advantage of the language-independent nature of grounded spatial and temporal data.
Another important benefit of storing and analysing "traditional" and social media content over space time is that these archives enable longitudinal analyses (Derczynski et al., 2013). For instance, longitudinal analyses on the online social graphs can reveal the evolution of social relationships and thus build models of trustworthiness and authority. It is also possible to start building user profiles over time, including previously spread rumours and, in general, what users talked about in the past. Focused on specific events, longitudinal analysis reveals discourse around events, arising from both social and traditional media. Similarly, in journalism and brand and reputation management applications, there is also demand for retrospective analyses of

[2] See http://www.tweetdeck.com/
[3] See http://www.storify.com/
[4] See http://www.ushahidi.com/products/swiftriver-platform
[5] See http://storyful.com/
[6] See http://www.geonames.org/

media content after a significant incident (e.g. to establish whether social media was used to entice more riots).

6.1. Dataset collection

The first phase is a human pilot annotation, of events, times and places in the target genres and languages. This includes annotation of web and social media text for events and times, in order to later temporally bound assertions. It also includes the annotation of locations (both formal and informal), and identification of document creation locations.

Corpora are then extended using cross-linguistic projection. PHEME will develop tools for projecting ST annotations across language (Spreyer and Frank, 2008; Costa and Branco, 2012). This allows the creation of new resources for English, German, Bulgarian, and possibly also the project's minor languages (French, Italian, Swahili).

Following the construction of a dataset, we will build spatial and temporal IE systems in multiple languages. These are aimed at ST grounding. As mentioned in Section 5., we intend to follow the narrative containers scheme. This centres on finding spaces and times within which groups of events are collected, before trying to resolve the specific, hard-to-annotate and potentially low-information individual relations. Finally, for grounding, while documents often come with a document creation date, and document creation location is harder to come by. To address this, the project investigates spatial grounding at both document level (creation location) and at assertion/event level.

Having found spatial and temporal entities in documents, it becomes possible to reason about bounds of assertions. We will apply and extend temporal reasoning and assertion bounding tools, which brings interesting challenges, particularly in the social media domain where one may be faced with many short documents describing different facets of a claim. In particular, cross-document spatio-temporal reasoning is a novel and unexplored research area. The output of these reasoning and bound-finding tools will be used as inputs to trustworthiness assessment systems.

6.2. Spatio-Temporal Information Extraction

Building upon existing resources is important to the advanced, complex tasks that PHEME addresses. Fully-featured ST information extraction pipelines can be built from state-of-the-art tools.

Regarding temporal annotation, we begin with annotation primitives: timexes, events and the relations between them. These are reasonably well-researched problems in newswire, but adapting to short messages which are pushed over networks by humans – i.e. social media messages – presents challenges in terms of the large linguistic variety, and interesting opportunities, from extra information and structure in personal profiles and network connections.

There are existing tools that may provide initial insights into the problem. For timexes, GATE and HeidelTime (Strötgen and Gertz, 2012) offer excellent entity extraction; TIMEN provides an open-source normalisation resource, and the state of the art leads to flexible parsing tools for handling previously-unseen timex formulations (Angeli et al., 2012; Bethard, 2013). Regarding event extraction, while older systems like EVITA are available, fast newer systems like TIPSem and the outcomes of the ARCOMEM project (Demidova et al., 2013) offer better performance. For linking times to events, systems like TIPSem, ClearTK-TimeML and NavyTime may be helpful.

Fewer tools are available for spatial annotation. In terms of locally-accessible location annotation systems, there is ANNIE, LODIE (Damljanovic and Bontcheva, 2012) and tools resulting from SemEval exercises. Developing spatial grounding and annotation systems involves more pioneering work here, beyond adapting existing tools.

Importantly, the project involves the creation of new tools for social media. PHEME couples systems like the above with document grounding and temporal relation annotation systems which operate on new languages and domains. This will involve the creation of new systems and annotations for event co-reference extraction, event-based summarisation, and ST grounding of individual messages. For example, the TimeML <TLINK> tag allows expression of intra-document co-reference and full-interval ordering between events, but need to be extended to handle both uncertain relations and also cross-document links. Similarly, the ability to create ISO-Space <PATH>s, <QSLINK>s between <LOCATION>s in different documents is required – as well as the ability to define common frames of references. We anticipate cross-document co-reference being instrumental in the grounding and subsequent veracity assessment of a significant proportion of claims and messages.

Social media networks present an unconventional kind of discourse, with different uses of reference and anaphora when compared to longer, standalone documents. The investigation of this structure will inform how annotations are used, and then leveraged for reasoning. Cross-document event co-reference is critical in order to group claims together; there is no work on this in social media, but challenges such as TDT generated extensive research on the general topic, e.g. (Bagga and Baldwin, 1999), and general concepts like chains provide a starting point.

Construction of timelines from timexes in messages and events mentioned across the network can then help define temporal bounds for events. Ji and Grishman (2011) excellent work on timelines proposes temporal bounding of assertions using times mentioned in collections of newswire documents, though this is all at day-level granularity. This granularity is suitable for retrospective analysis involving certain types of assertion (e.g. lifetimes), but not sufficient for realtime filtering of all kinds of events. In addition, extracted temporal bounds are likely to be uncertain, and require e.g. a constraint-satisfaction framework to pin down, as well as probabilistic veracity reasoning.

6.3. Evaluation

There are many ways in which PHEME's ST annotation output can be evaluated.

Primarily, we can evaluate against a gold standard (ours, or external ones, e.g. from TempEval). A secondary round of pilot annotations, over non-projected data, provides an opportunity for GS-style evaluation, as well as creating new high-quality language resources. Basic P/R/F1 measures work for spatio-temporal entity extraction – but one also needs to account for entity specificity. This may lead to

adopting or creating a new, nuanced evaluation measure. It is difficult to evaluate spatio-temporal reasoning; prior shared tasks in these areas have demonstrated this.

In addition, we can perform extrinsic evaluation using unskilled humans. For assertion grounding, a common-sense check can be applied, asking whether a particular claim (in prose) is intended to apply to certain ST constraints. This could be formulated as dialogue or question answering. A high quality crowdsourcing approach is feasible for this extrinsic evaluation (Sabou et al., 2014).

7. Conclusion

PHEME involves creating the necessary computational apparatus to model, identify, and verify phemes (internet memes with added truthfulness or deception), as they spread across media, languages, and social networks. Doing this raises difficult, interesting and important issues in spatio-temporal annotation of text in a wide variety of situations. PHEME investigates these issues in the context of social media, examining digital journalism and healthcare. Furthering spatio-temporal information extraction research promises a better understanding of the ever-present context that the meaning language relies upon so heavily.

Acknowledgments

This work received funding from the European Unions Seventh Framework Programme (grant No. 611233 PHEME).

8. References

G. Angeli, C. D. Manning, and D. Jurafsky. 2012. Parsing time: Learning to interpret time expressions. In *Proc. NAACL*, pages 446–455. ACL.

A. Bagga and B. Baldwin. 1999. Cross-document event coreference: Annotations, experiments, and observations. In *Proceedings of the Workshop on Coreference and its Applications*, pages 1–8. ACL.

S. Bethard. 2013. A synchronous context free grammar for time normalization. In *Proc. EMNLP*, pages 821–826.

N. Chambers. 2012. Labeling documents with timestamps: Learning from their time expressions. In *Proc. ACL*.

F. Costa and A. Branco. 2012. TimeBankPT: A TimeML Annotated Corpus of Portuguese. In *Proc. LREC*, pages 3727–3734.

D. Damljanovic and K. Bontcheva. 2012. Named Entity Disambiguation using Linked Data. In *Proceedings of the 9th Extended Semantic Web Conference*.

E. Demidova, D. Maynard, N. Tahmasebi, Y. Stavrakas, V. Plachouras, J. Hare, D. Dupplaw, and A. Funk. 2013. Extraction and Enrichment. Deliverable D3.3, ARCOMEM.

L. Derczynski and R. Gaizauskas. 2013. Information retrieval for temporal bounding. In *Proc. ICTIR*.

L. Derczynski, B. Yang, and C. Jensen. 2013. Towards Context-Aware Search and Analysis on Social Media Data. In *Proceedings of the 16th Conference on Extending Database Technology*. ACM.

L. Derczynski. 2013. *Determining the Types of Temporal Relations in Discourse*. Ph.D. thesis, University of Sheffield.

R. Gaizauskas, E. Barker, C. Chang, L. Derczynski, M. Phiri, and C. Peng. 2012. Applying ISO-Space to Healthcare Facility Design Evaluation Reports. In *Proc. ISA*, pages 31–38.

H. Ji and R. Grishman. 2011. Knowledge base population: Successful approaches and challenges. In *Proc. of ACL'2011*, pages 1148–1158.

A. K. Kolya, D. Das, A. Ekbal, and S. Bandyaopadhyay. 2012. Roles of event actors and sentiment holders in identifying event-sentiment association. In *Computational Linguistics and Intelligent Text Processing*.

D. Laney. 2001. 3d data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6.

R. Procter, A. Voss, and P. Brooket. 2012. A study of using social media in journalism. Technical report, University of Warwick.

R. Procter, J. Crump, S. Karstedt, A. Voss, and M. Cantijoch. 2013. Reading the riots: What were the police doing on twitter? *Policing and society*, 23(4):413–436.

J. Pustejovsky and A. Stubbs. 2011. Increasing informativeness in temporal annotation. In *Proc. LAW*, pages 152–160. ACL.

J. Pustejovsky and Z. Yocum. 2013. Capturing Motion in ISO-SpaceBank. In *Proc. ISA*, pages 25–33.

J. Pustejovsky, K. Lee, H. Bunt, and L. Romary. 2010. ISO-TimeML: An International Standard for Semantic Annotation. In *Proc. LREC*.

J. Pustejovsky, J. L. Moszkowicz, and M. Verhagen. 2011. ISO-Space: The annotation of spatial information in language. In *Proc. ISA*, pages 1–9.

V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proc. EMNLP*, pages 1589–1599. ACL.

D. Rout, D. Preotiuc-Pietro, K. Bontcheva, and T. Cohn. 2013. Where's @wally? a classification approach to geolocating users based on their social ties. In *Proc. Hypertext*.

M. Sabou, K. Bontcheva, L. Derczynski, and A. Scharl. 2014. Corpus annotation through crowdsourcing: Towards best practice guidelines. In *Proc. LREC*.

A. Sadilek, H. Kautz, and V. Silenzio. 2012. Modeling spread of disease from social interactions. In *International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 322–329. AAAI.

R. Saurí and J. Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *JLRE*, 43(3):227–268.

K. Spreyer and A. Frank. 2008. Projection-based acquisition of a temporal labeller. In *Proc. IJCNLP*.

J. Strötgen and M. Gertz. 2010. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In *Proc. SemEval*, pages 321–324.

J. Strötgen and M. Gertz. 2012. Temporal tagging on different domains: Challenges, strategies, and gold standards. In *Proc. LREC*, pages 3746–3753. ELRA.

J.-M. Xu, A. Bhargava, R. Nowak, and X. Zhu. 2012. Socioscope: Spatio-temporal signal recovery from social media. In *Machine Learning and Knowledge Discovery in Databases*, pages 644–659. Springer.