Contextual Temporal Profiles for Time Scoping of Knowledge Base Facts

Abstract

Methods for information extraction (IE) and knowledge base (KB) construction have been widely studied in recent years. However, maintenance of knowledge once it has been acquired is still largely under-explored. In real-life, changes happen to entities in the knowledge base. Propagating these changes to the KB is crucial for reducing inaccuracies in the state of the KB. In this paper, we present a method for detecting state changes of KB entities. We introduce a semi-supervised method that leverages Contextual Temporal Profiles (CTPs) of entities to detect change inducing contexts. Our experiments show how our method can be used to successful detect the time when an entity undergoes a certain state change.

1 Introduction

Knowledge Base (KB) construction is at a point where problems related to maintenance can no longer be avoided. One such problem is the detection of state changes that KB entities undergo overtime. When the relation value of a given entity is no longer true, the KB needs to be updated. This can occur for example if a person gets divorced or a company undergoes management changes. A consequence of the state change person fired or resigns from their job, is that the person is no longer an employee of their current employer. Currently, most IE methods detect patterns for learning attributes and mix them together with those involving state changes in the attribute. It is not unusual for an IE system to learn that the phrases: "is married to" and "is divorced from" are both high precision patterns to indicate the *hasSpouse* relation. In reality, one of those phrases marks the beginning and the other marks the end of an relation value for the *hasSpouse* relation. To DO: give specific defn of state change in terms of predicates.

State change is related to KB base update in general. However, prevalent approaches to IE extract knowledge from static Web snapshots (Fader 2011; Nakashole 2011), with no mention of how to perform updates. Other approaches periodically extract knowledge from a corpus such as Wikipedia, every time re-applying the extractor to all the documents even those that did not change (Suchanek 2007). The NELL system (Carlson 2010) follows a never-ending extraction model with the extraction process going on 24 hours a day. However NELL's focus is on language learning to self improve on its reading ability over time. In contrast, here we focus on detecting updates to specific attributes of entities. Detecting these changes has unique challenges not seen in IE methods: i) The state of an entity in the knowledge base can change in two ways: a) the entity acquires a new attribute value (e.g, person is newly married or has a new job) b) the entity loses an attribute value (e.g, person is divorced or was fired). Knowledge acquisition methods focus on, by design, acquiring new attribute values but not on detecting loss of relation values. In contrast, the problem of state change detection includes detection of *both* acquisition and loss of relation values. ii) In state change detection, we would like to detect the state change at the right time. This means that state change detection requires accurate and timely temporal scoping of relation values.

In this paper we propose a method for state change detection, based on Contextual Temporal Profiles (CTPs) of KB entities.

2 Related work

Temporal scoping of KB facts is still largely under-explored. Prior work on fall into two categories: i) those that attempt to extract temporal scope from text, at the time of fact extraction; ii) those that attempt to extract temporal scope by leveraging statistics in large Web corpora. Early methods fall under the category i); Timely YAGO [?] and PRAVDA [?] are two such methods. Timely YAGO applies regular expressions to Wikipedia infoboxes to extract time scopes. It is therefore not applicable to any other corpora but Wikipedia. The technique employed by PRAVDA it to use textual patterns along with a graph-based re-ranking method. Methods falling under category 1) have the downside that it is not clear how they can be applied to facts that are already in the knowledge base. Before this paper, only one other approach fell under category ii), which is a recent system called CoTs(?). Like our approach, Cots leverages corpus statistics to infer temporal scope information. CoTs models facts by temporal profiles to monitor how its mention rises and fall over time. Unlike our approach, CoTs is exclusively based on simple counts of facts in documents, ignoring the context of those documents.

TARSQI [?] is a tool for automatically annotate events and time-expressions found in natural language text. TARSQI uses the TimeML[?] markup language in its annotations. The TempEval[?] challenge has led to a number of works on temporal relation extraction [??]. However, These approaches based on TimeML and TempEval targeted at events which just verbs, different from the precise relational facts this paper is focused on. Furthermore, the focus appears to be on deciding which event happened before or after which one. This is also different from the specific dates we seek to expose in our work.

3 Method

Our main idea is that given an entity, and its Contextual Temporal Profile (CTP), we can learn when such an entity undergoes specific state changes. The CTP at a given time point t contains the context within which the entity is mentioned at that time . Our method is based on two related insights: i) the context of the entity at time t reflects the state change that the entity undergoes. ii) the *difference in context* before, at time $t-1$, and after, at time $t+1$, reflects the context responsible for the state change at time t. However naively applying these two insights does not result in a good solution. This is because the entity can undergo a multiplicity of changes at the same time t. Thus both the contexts and the differences in contexts can contain information pertaining to several state changes. We therefore need a way of determining which part of the context is relevant to a given state change sc_i. To this end, we generate what we refer to as an $aggregate$ CTP, $CTP(\hat{e}, sc_i)$ for a hypothetical average entity $\hat{e}$ undergoing state change sc_i. We generate $\hat{e}$ and its $CTP(\hat{e}, sc_i)$ for from the CTPs of a seed set of entities that undergo state change sc_i. Once aggregate CTPs are computed, we can then use them to detect state changes of previously unseen entities by their similarity to the aggregate CTPs.

3.1 Learning State Change Vectors

To build CTPs for entities, we use the Google Books Ngram corpus. It contains n-grams for up to $n = 5$, along with occurrence statistics from over about 5 million digitized books (Michel 2011). Of importance is the fact that this corpus is time-stamped, with granularity of year.

Definition 1 (Contextual Temporal Profile) *The Contextual Temporal Profile (CTP) of an entity e, at time t consists of the context, $C_e(t)$, within which e is mentioned. Specifically $C_e(t)$ is generated from the 5-grams that mention e at time t. Computing all contexts $C_e(t)$, $\forall t_i$ generates the CTP of e.*

The CTPs can contain contextual unit (unigrams or bi-grams) that are not related to state change (s) occurring at time t. We therefore would like to ensure that the context is free of such words. The first thing we do is to disregard stop words. The second thing is to compute $tf-idf$ statistics for each contextual unit and only retain the top k ranking units in the context $C_e(t)$. We compute tf-idf by treating each time unit t as a document containing words that occur in the context of e.

Additionally, notice that the CTPs may contain contexts attributed to several state changes. We therefore tease apart the CTPs to isolate contexts specific to a given state change. For this, our method takes as input a small set of seed entities,

$\mathcal{S}(sc_i)$, for each type of state change . Thus for US presidency state change, where an entity enters or leaves the presidency, we would have seeds as follows: for entering office, *(Richard Nixon, 1969),(Bill Clinton, 1993)* and for leaving office *(Richard Nixon, 1974),(Bill Clinton, 2001)*. From the CTPs of the seeds for state change sc_i, we generate an aggregate CTP, $CTP(\overline{e}, sc_i)$.

Definition 2 (Aggregate CTP for $\hat{e}$) *The CTP of a mean entity $\hat{e}$, at time t, for state change sc_i is made up of the contexts of the CTPs of all entities in the seed set $\mathcal{S}(sc_i)$ that undergo state change sc_i at time t. Thus, the aggregate context $C_{\overline{e}}(t)$ = $\bigcap_{e:C_e(t)\neq\emptyset} C_e(t)$.*

Notice that so far the aggregate CTPs $CTP(\overline{e}, sc_i)$, a non-empty aggregate contexts only appears at time points $ti, tx, tw, ...tz$ where at least one entity in the seed set, $\mathcal{S}(sc_i)$, undergoes a state change. For reasons that will become clear later, we also generate aggregate contexts for times $t-1$ and $t+1$, in the same way we generate $C_{\overline{e}}(t)$. A distinction is made between contexts at times $t-1$, t, $t+1$ and the way we treat them.

One might think that the aggregate CTP $CTP(\overline{e}, sc_i)$ is likely to be sparse in terms of populated time points, depending on the number of seeds. However, the way we use the profiles is such that we generate vectors from these CTPs. In particular we generate three vectors v_c, v_{bc}, v_{ac}.

Definition 3 (Aggregate Change Vectors) *From the aggregate CTP, $CTP(\overline{e}, sc_i)$, we generate three change vectors, v_c, v_{bc}, v_{ac}: v_c reflects the context during the state change, $v_c = \cup_{w \in C_{\overline{e}}(t)}$; v_{bc} is a difference vector containing the difference in context before and the context during change: $v_{bc} = C_{\overline{e}}(t) - C_{\overline{e}}(t-1)$. v_{ac} is a difference vector containing the difference in context after and the context during change: $v_{ac} = C_{\overline{e}}(t) - C_{\overline{e}}(t-1)$.*

The aggregate change vectors are then used detecting state changes.

3.2 Detecting State Changes

To detect state changes for a previously unseen entity, we generate its CTP, its change vectors, compare every these info for each candidate time t for state change sc_i to happen. Rank the candidate time points based on their similarity to the aggregate change vectors. TO DO: expand this sections

4 Experiments

We carried out experiments to answer the following question: do Contextual Temporal Profiles help improve temporal scope extraction over plain temporal profiles? We answer this questions by comparing to the CoTs system which leverages temporal profiles but only relies on mention counts and does not attempt to leverage the context of the mention.

Our evaluation procedure is as follows: given a fact, its temporal scope. For comparison purposes, we evaluate on the relations by CoTs. CoTs was evaluated on five relations: three from the US Administration domain (relations US President, US Vice President, and US Secretary of State) ; and two from the Academy Awards domain (Best Director and Best Movie)

5 Conclusion

We conclude as follows:

References

A. Angel, N. Koudas, N. Sarkas, D. Srivastava: Dense Subgraph Maintenance under Streaming Edge Weight Updates for Real-time Story Identification. In *Proceedings of the VLDB Endowment*, PVLDB 5(10):574–585, 2012.

S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z.G. Ives: DBpedia: A Nucleus for a Web of Open Data. In *Proceedings of the 6th International Semantic Web Conference (ISWC)*, pages 722–735, Busan, Korea, 2007.

M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, O. Etzioni: Open Information Extraction from the Web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2670–2676, Hyderabad, India, 2007.

K. D. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor: Freebase: a Collaboratively Created Graph Database for Structuring Human Knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages, 1247-1250, Vancouver, BC, Canada, 2008.

A. Carlson, J. Betteridge, R.C. Wang, E.R. Hruschka, T.M. Mitchell: Coupled Semi-supervised Learning for Information Extraction. In *Proceedings of the Third International Conference on Web Search and Web Data Mining (WSDM)*, pages 101–110, New York, NY, USA, 2010.

A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr., T. M. Mitchell: Toward an Architecture for Never-Ending Language Learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI)* 2010.

L. Del Corro, R. Gemulla: ClausIE: clause-based open information extraction. In *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, pages 355-366. 2013.

A. Das Sarma, A. Jain, C. Yu: Dynamic Relationship and Event Discovery. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining (WSDM)*, pages 207–216, Hong Kong, China, 2011.

A. Fader, S. Soderland, O. Etzioni: Identifying Relations for Open Information Extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1535–1545, Edinburgh, UK, 2011.

C. Havasi, R. Speer, J. Alonso. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. In *Proceedings of the Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria, 2007.

J. Hoffart, F. Suchanek, K. Berberich, E. Lewis-Kelham, G. de Melo, G. Weikum: YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages. In *Proceedings of the 20th International Conference on World Wide Web (WWW)*, pages 229–232, Hyderabad, India. 2011.

Karin Kipper, Anna Korhonen, Neville Ryant, Martha Palmer, A Large-scale Classification of English Verbs, Language Resources and Evaluation Journal, 42(1): 21-40, 2008, data available at `http://verbs.colorado.edu/~mpalmer/projects/verbnet/downloads.html`

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Holberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, Erez Lieberman Aiden: Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014):176182.

N. Nakashole, M. Theobald, G. Weikum: Scalable Knowledge Harvesting with High Precision and High Recall. In *Proceedings of the 4th International Conference on Web Search and Web Data Mining (WSDM)*, pages 227–326, Hong Kong, China, 2011.

N. Nakashole, T. Tylenda, G. Weikum Fine-grained Semantic Typing of Emerging Entities. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1488-1497, 2013.

N. Nakashole, G. Weikum, F. Suchanek: PATTY: A Taxonomy of Relational Patterns with Semantic Types. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1135 - 1145, Jeju, South Korea, 2012.

Feng Niu, Ce Zhang, Christopher Re, Jude W. Shavlik: DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference. In the VLDS Workshop, pages 25-28, 2012.

A. Ritter, Mausam, O. Etzioni, S. Clark: Open Domain Event Extraction from Twitter. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1104-1112, Beijing, China, 2012.

D. Shahaf, Carlos Guestrin: Connecting the dots between news articles. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 623-632, 2010.

T. Sakaki, M. Okazaki, Y. Matsuo: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, pages 851-860, Raleigh, North Carolina, USA, 2010.

F. M. Suchanek, G. Kasneci, G. Weikum: Yago: a Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web (WWW)* pages, 697-706, Banff, Alberta, Canada, 2007.

F. M. Suchanek, M. Sozio, G. Weikum: SOFIE: A Self-organizing Framework for Information Extraction. In*Proceedings of the 18th International Conference on World Wide Web (WWW)*, pages 631–640, Madrid, Spain, 2009.

W. Wu, H. Li, H. Wang, K. Zhu: Probase: A Probabilistic Taxonomy for Text Understanding. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 481–492, Scottsdale, AZ, USA, 2012.

Word Wide Web Consortium (W3C): RDF Primer. `http://www.w3.org/TR/rdf-primer/`, Accessed 2013.