Learning Temporal Information for States and Events

Zornitsa Kozareva
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
kozareva@isi.edu

Eduard Hovy
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
hovy@isi.edu

***Abstract*—Knowing the typical duration of events (for example, hurricanes last hour or days but not seconds or years) supports a variety of tasks in automated machine reading. Recently, methods to learn these durations for a limited class have been reported. However, events are associated with several other typical times, such as initiation points and preparation intervals. In this paper we define six temporally related aspects of events. We describe an automated method to learn events from the web and patterns that signal the typical temporal characteristics of the events. Finally, we show which patterns tend to signal which aspects. This diversity of event types, temporal aspects, and time characteristics has never yet been reported.**

I. Introduction and Past Work

One of the fundamental characteristics of events and states is their ability to take temporal arguments, such as times of occurrence and durations. This information is crucial for accurate understanding of stories, which requires identification, ordering, and coreference of events and states. Robust automated multi-sentence text understanding (deep reading) is impossible without this ability.

The logical and linguistic foundations of temporal information are well understood. For example, [1] was among the first to provide a cognitive-linguistic account of the principal groupings of events and states. The temporal relation calculus of [2] is justly renowned and used. [3], [4] describe the philosophical underpinnings, while [5], [6] define and work on the logical representations. Over the years very helpful analysis of time, with accompanying annotated corpus TIME-ML [7], [8] have been developed and provided to the research community.

None of this work, however, provides a list of the typical temporal values for each event type. And for automated language processing, this information is crucial. In real text, the temporal information about instances of events or states is often not stated explicitly. It is assumed that the reader will know that a robbery takes minutes, not years, but that its preparation may take months, or that a hurricane lasts for hours or perhaps days, but not seconds or years. Little work has been done on determining the typical durations of event and state types. [9] created a list of typical durations for a small set of verbs denoting events. But given the large variety of different events and states, and the numerous verbs that indicate them, completing this list for all events and states is a daunting prospect. Surprisingly, there has been almost no work on the automated learning of such temporal information. The exception is [10], who describe a method for learning default durations of events using three manually defined text harvesting patterns.

All previous work discusses the typical duration of events. But in the real world, events are not associated with duration only. Typically, many events have a preparation stage that takes a characteristic length of time; some events are performed intermittently for a certain typical period; and some events occur during others by necessity. Some of these characteristics are studied under the rubric of linguistic Aspect; others are the interest of nave physics and cognitive models. But all of them can be useful for tasks in Information Extraction, Textual Entailment, and Machine Reading. While TIME-ML provides a great deal of useful information, there has been no investigation to date of which temporal aspects of events and states are readily available on the web for automated harvesting.

In this paper we show empirically that an automated procedure designed to acquire the temporal characteristics of events and states can learn, for a given event or state, not one but several different associated times, and that these times correspond to a set of regular (and useful) different temporal aspects, including *duration, preparation*, and so on. In Section II we define the most prominent aspects and provide in Section III a method to determine the relevant time aspects for any event, using appropriate recognition patterns. Section IV describes the data we use. In Section V we measure the general performance accuracy of the various patterns, and show the time-scale variances of the various aspect values for events.

II. Definition of Aspects

Several previous projects have manually defined extraction patterns that identify temporal information associated with events in text. These patterns generally identify event durations [10]. In the next section we introduce an automated pattern learning procedure that generates a wider-ranging set of patterns that can discover temporal information for other aspects as well. We have identified widely different time-scales associated with the events and grouped them into sets such that each set expresses a different temporal aspect of the event. We identify the following temporal aspects.

Duration (D): The temporal expression *t* specifies the length of time that *e* occurs. This is the most common aspect. It is frequently expressed by "*e* for *t*", as in "she ran for three hours". It makes no commitment whether the event occurs continuously or intermittently (see under *I* below).

Time-at (T): The timepoint or frequency *t* at which event *e* occurs, or starts occurring. It is frequently expressed by "*e* at *t*", as in "he made tea at 3 o' clock, or by "*e* is every *t*", as in "the dance is every two months".

Interval before or after (BA): The temporal expression *t* specifies the duration of the preparatory stage leading up to the event *e* or of the subsequent stage following it. It is expressed using phrases such as "*e* starts in *t*", "it is *t* until *e*", and "*t* since *e*".

Inclusion during (I): The event *e* occurs during the period defined by the temporal expression *t*. (That is, the start of *e* occurs not earlier than the start of *t* and the end of *e* not later than the end of *t*. If *t* is an unanchored temporal extent then the temporal extent of *e* is strictly less than *t*.)

We identify three subtypes of *I*, corresponding to linguistic Aspect:

- **Intermittent occurrence included (II):** *e* occurs repeatedly during *t*. There are some moments during *t* that *e* does not occur, and others that it does.
- **Continuous occurrence included (CI):** *e* occurs without interruption during *t*. There are no moments during *t* that *e* does not occur.
- **Negated occurred included (NI):** event *e* does not occur at or during the time *t* specified.

Speed (S): The expression *t* expresses the speed or rate at which the event *e*, or a part of it, occurs. This is frequently expressed as in "the shutter clicked at 30 frames a second".

Age (A): The expression *t* specifies the length of time since event *e* first occurred; that is, it provides the age of event *e* (if *e* is an often-repeated event), or the starting point of the appearance of *e*. Examples are "*e* is *t* old" or "the first *e* was *t* ago". When focusing on one or more specific instances *e* during the time period then the preferable reading is aspect *I* above.

III. Method Description

In this section we describe a semi-supervised procedure for automatically harvesting states and events from the Web as well as a pattern learning procedure that harvests patterns associated with the different temporal aspects of the events.

A. Learning States and Events from the Web

Our approach for learning states and events is inspired by Hearst's observations that sentences contain clues as to their meanings and these can be captured using lexico-syntactic patterns [11]. Among the most successful and widely used pattern based knowledge harvesting approaches are [12], [13], [14], [15].

While we could have used any of these approaches for our experiments, we choose the approach of Kozareva et al. since it (1) is minimally supervised requiring as input only one seed example and a pattern; (2) has a built-in bootstrapping procedure that retrieves larger quantities of terms; (3) has been shown to achieve higher accuracy than the methods described in [12], [13].

The algorithm is pretty straightforward to implement. It uses the so called doubly anchored lexico-syntactic pattern (DAP) of the form:

"*semantic-class* such as *seed* and *"

where *semantic-class* is the class we want to learn, *seed* is the input example with which the algorithm initiates the harvesting procedure and * corresponds to the position on which the new terms are extracted.

The algorithm has a bootstrapping mechanism that takes the terms learned on the * position and (if not yet explored) places them on the position of the *seed* for a new round of learning. In this manner new lexico-synatctic patterns are automatically created and used to extract terms. The algorithm is implemented as a breadth-first search which can run for a certain number of iterations or until no new terms are discovered (complete exhaustion). The output of the algorithm is a set of terms related to the *semantic-class*. We denote this set as E.

To assign confidence to each extracted term, Kozareva et al.[1] represent the harvested terms as a directed graph where the nodes correspond to the extracted terms and the edges show for each term v which term u generated it. To rank the terms they use

$$outDegree(u) = \frac{\sum_{\forall(u,v)\in E}(u,v)}{|V|-1} \quad (1)$$

where u is the current term for ranking, $\sum_{\forall(u,v)\in E}(u,v)$ is the sum of all outgoing links from term u and $|V|$ is the number of all unique terms extracted during the bootstrapping process. This measure captures how often a term is seen in the *seed* position of the DAP pattern. Intuitively, the more often the term is seen, the more likely it is to be of the semantic class.

B. Learning Patterns between Events and Temporal Units

Next, given an event $e_i \in E$ our objective is to learn all temporal aspects and their distributions. For example, a *war* can last for *months*, *years* or *centuries*, but it is less likely to last for *seconds* or *minutes*.

For the purpose we use ten temporal units (tu) some of which were previously used by Pan and Jurafsky (*seconds*, *minutes*, *hours*, *days*, *weeks*, *months*, *years*, *decades*) and two additional ones (*centuries* and *eras*). Unlike [10] who relied on only three manually defined patterns to learn the temporal aspect duration, we employ an automated pattern harvesting procedure which allows us to capture all temporal aspects and distributions associated with the event.

[1]For more details on how to build the graph and rerank the terms please refer to the paper of [14].

To harvest the patterns characterizing the temporal aspects, we use the automatically harvested events from the previous step and the ten different temporal units. We build Web queries of the form:

$$\text{“}e_i * tu_j\text{”}$$
$$\text{“}e_i \text{ that } * tu_j\text{”}$$

where e_i is the event of interest, tu_j is a temporal unit and * corresponds to the position on which the terms located between e_i and tu_j will be extracted.

We submit the queries as exact phrases. We keep all snippets for the returned results and filter out those that do not contain the event and the temporal unit terms. For example, for the event *concert* and the temporal unit *hours* the submitted queries are "*concert * hours*" and "*concert that * hours*". One of the retrieved sentences is "*We only want to participate in a **concert** that takes 2 **hours***." from which we extract the pattern "*takes*". The pattern acquisition process is done for the rest of the retrieved snippets and queries. The output of the algorithm is quadruples of the form event, temporal unit, pattern and frequency of the extraction.

In addition to the Web patterns which are composed of consecutive words between the event and temporal terms, we also use TextRunner [16] which generates more loose patterns. TextRunner is a large database containing background knowledge about pre-existing relations of the form NOUN-VERB-NOUN. This information was obtained by parsing all sentences in large document collection and collecting the subject, object and predicate of the sentences. For our task, we query TextRunner with the event term as arg_1, the temporal unit as arg_2 and collect all predicates (verb phrases) matching the query. For example, the retrieved patterns from TextRunner for *"storm"* and *"hours"* include: *"lasted"*, *"lasted about"*, *"raged for"*, *"continued for"*, *"hit during"* among others. The output of this step is the same as the Web pattern extraction one.

IV. Data Collection

A. Event Data

For our experiments we used the bootstrapping procedure described above. We instantiated it with the semantic class *events* and the seed term *earthquakes*. Each pattern was submitted to Yahoo!Boss as a Web query. The top 1000 snippets related to the query were retrieved and part-of-speech tagged with TreeTagger [17]. The algorithm ran until exhaustion for 14 iterations and extracted 16347 unique event terms, 12135 of which were seen more than once in the extraction pattern. We reranked the harvested information with $outDegree$ and selected the top 1000 results for manual annotation. We found that 98% of the terms were related to events, the majority of which correspond to social events like *birthdays*, *anniversaries*, *concerts*, disaster events like *hurricanes*, *tornadoes*, health events like *heart attacks*, *strokes* among others. Table II shows 40 examples of the automatically harvested states and events.

concerts	birthday party
conferences	christmas
weddings	football game
floods	dinners
stroke	banquets
festivals	basketball game
seminars	wedding reception
heart attack	war
parties	picnics
workshops	death
meetings	fire
exhibitions	conventions
birthdays	graduation
fairs	lectures
hurricanes	golf tournament
droughts	music festival
parades	marathons
storms	wine tasting
anniversaries	cyclone
dances	triathlons

TABLE II
Examples of Learned States and Events.

B. Pattern Extraction Data

To harvest the patterns from the Web, we used the top 1000 event terms from the previous step. For each event and the different temporal units, we built the Web queries described in the previous section and submitted them to Yahoo!Boss. For each query we retrieved a maximum of 1000 sentences. Similarly we used the events and the temporal units to form the TextRunner queries and retrieved the predicates. In total the algorithm learned 1187683 patterns, 220315 of which were unique.

Of course, the different events have their individual distribution of patterns. In Table I we show the top 10 learned patterns (if present) for the events *concert*, *dance*, *flood* and *storm*.

V. Evaluation and Discussion

Overall, the pattern harvesting step generates a rich set of patterns associated with the events and the temporal units. We have randomly selected six events and shown in Figure 1 their pattern-specific distribution of counts over the various temporal units. On the x-axis we show the number of patterns learned for each temporal unit. As can be seen, the most frequent temporal values for *concert*, *stroke*, and *concert* is *years*, which we know to be their individual build-up / preparation times and not their durations. It is somewhat surprising that the patterns harvested from open text for all time units reflect far more attention to the Before-After aspect (BA) of events than the Duration aspect (D), contrary to most researchers' intuitions and prior work.

These graphs show the importance of differentiating the aspects when one is harvesting times for individual events and states.

In order to determine whether and how one can differentiate the various aspects, and to verify our results, we have performed a small amount of manual annotation. To each learned

	SECONDS	MINUTES	HOURS	DAYS	WEEKS	MONTHS	YEARS	CENTURIES
CONCERT	sells out in	that starts in	started	was	is	are held	was	has disappeared in
		is	was	is	was	was	are held throughout	that spans
		was cancelled at	lasted	was on	are held	take	are held	takes to
		started	lasted about	will be held	is in	are held during	saw	is composed of over
		will last	will last	take	take	are at	seen	were performed in
		lasted	lasts	is held	recorded	took	seen in	are performed in
		was	is	took place	took	was recoded	are scheduled	
		sold in	was	is in	take	was over	is	
		enjoyed	shows	will take	has been sold out for	are held in	is held	
		began	was over	will end	was recoded	was scheduled	live in	
DANCE	start	lasts	is created after	that occupies	did	move	was	was
	can create in	took	ran	was halted	is in	must end in	was born over	dates
	is	coming	lasts for	took	was	happens	has evolved over	is
		starts in	spend	continued for	pop up		aged from	
		was	is in	was in	been		seen in	
		takes	that resulted in	is in	is presented in		was on	
		has lasted	spent	were			performed	
		had lasted	can take been				is on	
		took	starts in	had			is for	
		is	continues for	takes			has spent	
FLOOD		lasted	hit within	was	was		raged though	
		can take	needed	continued	affected		occurred	
				lasted	that lasted		changed	
				happened			was	
				lasted for			that will occur	
				began			caused	
				was increasing			received for	
				started on			destroyed	
				swept during			will happen	
				remains for			damaged	
STORM		hit	was	had	prepares for	lasted for	hit	may continue
		was	lasted for	can last for	hit	erupted	that hit	
		developed within	raged for	takes	was	lasting	destroyed	
		lasted for	was moving at	lasted	passed through	happened	were	
		lasted	spent	had been for	comes	was	lash	
		turned at	is	was for	are expected	took	have killed	
		struck	had lasted	can last	moved	starts	recovering	
		broke	delayed for	occurs in	was forecast	occurred	threatens	
		arrived	continued for	was forecast	to hit	had in	that occurred	
		began	went through	took	that dumped	will begin at	is not expected	

TABLE I

Learned Patterns for the Events Concert, Dance, Flood and Storm

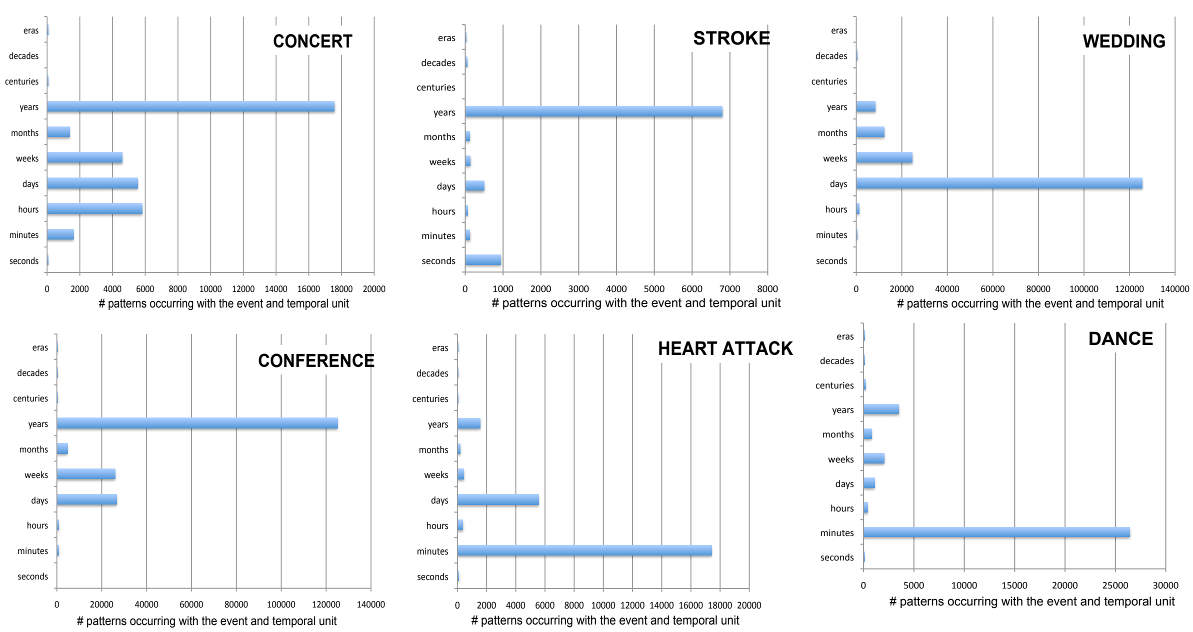

Fig. 1. Distributions of Temporal Values for each Temporal Unit.

	DURATION	TIME-AT	INTERVAL BEFORE/AFTER	INCLUSION DURING	SPEED	AGE
	lasted	was	has been sold out	seen in		
C	was	was recorded	was postponed for	been to in		
O	is	started	had been scheduled	attended in		
N	lasted for	ended	had been waiting for	was played		
C	lasts	was broadcast	will be	will mark		
E	will last	is in	was after	was held on		
R	lasted over	saw	sells out in	take throughout		
T	ran for	began	is in	scheduled		
	sold in	was shot	was recorded	produced over		
	will last about	took	was cancelled	has changed over		
	lasted	was	took	spent		was
	lasts	is in	that resulted in	seen in		is
D	is	in	takes	have in		dates
A	was	been	started in	has evolved over		was born over
N	moves for	was	start	has changed over		to forget
C	lasts for	performed	spent	was in		to celebrate
E	can last for	is presented at	in created after	performed		that was developed over
	would practice	is for	been	has spent		originated over
	spent	was halted	move	that occupies		be
	ran	is on	was in	continued for		performed
	lasted	was	hit within	that occurs		
	lasted for occurred	was	will happen	occurred during		
	lasts	happened	can be predicted	that struck		
F	was	that occurred		is expected during		
L	lasts for	occurs		hits		
O	continued	having		have occurred in		
O	was on	have hit		devastates		
D	was increasing	began		destroyed		
	that lasts for	was		comes in		
	continued for			cause		
	lasted	hit	has been brewing for	had in	was moving at	
	lasted for	came	develops during	hit during	is moving at	
	continued for	was	are expected for	occurs during		
S	took	started	had been forecast	to close in		
T	raged for	will hit	followed	happened during		
O	lasts	that pummeled	would form in	began in		
R	can last	hits	was predicted for	happened		
M	would last for	comes	threatens	has continued		
	spent	arrived	stalled for	that hit		
	passed	occurred	may not come for	that blows		

TABLE III
Temporal Aspects Associated with the Events Concert, Dance, Flood and Storm

pattern and each time-scale it harvested, we manually assigned one of the aspects defined in Section II, namely *D, T, BA, I, S*, and *A*, or when unsure used the class *Other*. The results were very informative.

As shown in Table III, some patterns reliably return only a single aspect, and can be used as the primary guides for their aspect during harvesting. For example, the patterns *"X lasted Y"* and *"X lasted for Y"* occur for all events, but only within the Duration column. They therefore reliably deliver the Duration aspect, and (we surmise) for other events and states as well. Similarly, *"seen in"*, *"has occurred in"*, *"comes in"* and other patterns including the preposition *"in"* pretty reliably indicate the Inclusion aspect (I). The Time-At aspect (T) seems to be preferred by patterns including the preposition *"at"*, and the Speed aspect (S) and Age aspect (A) have characteristic patterns not shared by the others.

It may therefore be sufficient, when harvesting the temporal characteristics for a new event or state, simply to start with four or five of the most reliable patterns for each aspect. If they all yield the same time scale then one can safely conclude that this is the time scale appropriate for that aspect. If however they differ, further harvesting (first of more patterns, and then of more temporal values) is in order. Additional manual annotation is required to test the accuracy of this hypothesis.

VI. Conclusion

This paper describes six temporal aspects very frequently associated with events and states, and provides a minimal-effort multi-stage automated procedure for harvesting this sort of information from the web. It illustrates the typical characteristics of a few dozen event types, together with the patterns that best harvest this information. This is the first time that these aspects are identified with patterns that can be used to learn their information.

What remains to be done is an actual compilation of the temporal aspects of the most common 5000 or so verbs, for general distribution. We hope that readers are inspired to create and make available such lists.

Acknowledgments

We acknowledge the support of DARPA contract number FA8750-09-C-3705 and NSF grant IIS-0429360. We would like to thank Alan Ritter, Anthony Fader and Robert Bart for allowing us to use the TextRunner system. We are grateful to Jerry Hobbs for his useful comments.

References

[1] Z. Vendler, *Linguistics in Philosophy*, 1967.

[2] J. F. Allen, "Maintaining knowledge about temporal intervals," *Commun. ACM*, vol. 26, pp. 832–843, 1983.
[3] D. Dowty, *Word meaning and montague grammar*, 1979.
[4] W. V. Quine, "Events and reification," in *Actions and Events: Perspectives on the philosophy of Donald Davidson, Oxford: Blackwell. Reprinted in R. Casati and A.C. Varzi, eds., Events*, 1996, pp. 107–116.
[5] A. Lascarides and N. Asher, "Temporal interpretation, discourse relations and commonsense entailment," *Linguistics and Philosophy*, vol. 16, no. 5, pp. 437–493, 1993.
[6] H. Kamp and U. Reyle, "A calculus for first order discourse representation structures," *Journal of Logic, Language and Information*, vol. 5, pp. 297–348, 1996.
[7] J. Pustejovsky, J. Castao, R. Ingria, R. Saur, R. Gaizauskas, A. Setzer, and G. Katz, "Timeml: Robust specification of event and temporal expressions in text," in *in Fifth International Workshop on Computational Semantics (IWCS-5*, 2003.
[8] J. Pustejovsky, P. Hanks, R. Saur, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo, "The timebank corpus," pp. 647–656, 2003.
[9] F. Pan, R. Mulkar, J. R. Hobbs, G. W. Bush, and V. Putin, "Learning event durations from event descriptions," in *in Proceedings of the 44th Conference of the Association for Computational Linguistics (COLING-ACL*, 2006, pp. 393–400.
[10] A. Gusev, N. Chambers, D. R. Khilnani, P. Khaitan, S. Bethard, and D. Jurafsky, "Using query patterns to learn the duration of events," in *In International Conference on Computational Semantics*, 2011.
[11] M. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *Proceedings of the 14th conference on Computational linguistics*, 1992, pp. 539–545.
[12] M. Pasca, "Acquisition of categorized named entities for web search," in *Proc. of the thirteenth ACM international conference on Information and knowledge management*, 2004, pp. 137–145.
[13] O. Etzioni, M. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates, "Unsupervised named-entity extraction from the web: an experimental study," *Artificial Intelligence*, vol. 165, no. 1, pp. 91–134, June 2005.
[14] Z. Kozareva, E. Riloff, and E. Hovy, "Semantic class learning from the web with hyponym pattern linkage graphs," in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics ACL-08: HLT*, 2008, pp. 1048–1056.
[15] E. Hovy, Z. Kozareva, and E. Riloff, "Toward completeness in concept extraction and classification," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2009, pp. 948–957.
[16] A. Yates, M. Cafarella, M. Banko, O. Etzioni, M. Broadhead, and S. Soderland, "Textrunner: open information extraction on the web," in *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, ser. NAACL-Demonstrations '07, 2007, pp. 25–26.
[17] H. Schmid, "Probabilistic part-of-speech tagging using decision trees," in *Proceedings of the International Conference on New Methods in Language Processing*, 1994, pp. 44–49.