Discovering Concept-Level Event Associations from a Text Stream

Tao Ge[1,2], Lei Cui[3], Heng Ji[4], Baobao Chang[1,2], and Zhifang Sui[1,2](✉)

[1] Key Laboratory of Computational Linguistics, Ministry of Education, School of EECS, Peking University, Beijing, China
{getao,chbb,szf}@pku.edu.cn
[2] Collaborative Innovation Center for Language Ability, Xuzhou, China
[3] Microsoft Research, Beijing, China
lecu@microsoft.com
[4] Rensselaer Polytechnic Institute, Troy, NY, USA
jih@rpi.edu

Abstract. We study an open text mining problem – discovering concept-level event associations from a text stream. We investigate the importance and challenge of this task and propose a novel solution by using event sequential patterns. The proposed approach can discover important event associations implicitly expressed. The discovered event associations are general and useful as knowledge for applications such as event prediction.

1 Introduction

People often seek event associations because such knowledge enables them to predict the future, take certain precautions, or make wise decisions under a specific circumstance. For example, if one knows landslides often occur after earthquakes, the risk of damages can be reduced.

Due to the importance of event associations, this paper studies concept-level event association discovery. In contrast to the previous work studying specific and context-dependent events (e.g., *Jim hit John yesterday*), concept-level events (e.g., *earthquake*) are context-independent and thus their associations (e.g., ⟨*earthquake-landslide*⟩) are general and useful as knowledge, which has attracted so much attention that some semantic networks and knowledge bases (e.g., ConceptNet[1]) have started to incorporate concept-level event association knowledge due to its potential ability in knowledge inference and decision making. Formally, given two concept-level events e_i and e_j, we define e_i and e_j are associated if e_j tends to to be triggered, caused or affected by e_i, and e_j is not a part of e_i.

Despite extensive studies on event relations [1–5,13,16–19,22] in NLP field, the task of concept-level event association discovery has not been much explored. Most work [9–12,20,21,23] related to concept-level event association discovery

[1] http://conceptnet5.media.mit.edu/.

C.-Y. Lin et al. (Eds.): NLPCC-ICCPOL 2016, LNAI 10102, pp. 413–424, 2016.
DOI: 10.1007/978-3-319-50496-4_34

mainly focused on causality extraction based on text clues (e.g., causal verbs and connectives), which is usually insufficient because event associations are not limited to causality explicitly expressed. For many associations that are implicitly expressed, it is difficult for text-based approaches to discover. Although the implicit event associations might be discovered by the methods based on word co-occurrence (e.g., Point-wise Mutual Information (PMI)), these methods do not work well for our goal for two reasons. First, computing PMI of arbitrary event pairs is time-consuming and will introduce many trivial and uninformative event pairs like (*say, sit*). Second, event pairs with high PMI may not be truly associated since events in some pairs are minor events (e.g., *donation* and *evacuate* in Fig. 1) triggered by a major event (e.g., *earthquake*) and they are not associated though they always co-occur.

To solve this problem, we study this task from a novel viewpoint – exploiting Burst Sequential Patterns (BSPs[2]) of events in a text stream to discover event associations. Intuitively, if a word describing an event always bursts after or co-bursts with another event word throughout a text stream, these two events are probably associated (e.g., the word *donation* usually bursts after *earthquake*). By analyzing such BSPs in a text stream, it is possible to discover event associations even if they are implicitly expressed. For this goal, we propose to use *Burst Information Networks* (*BINets*) [6–8] as a representation of a text stream, which can overcome the limitations of traditional PMI-based methods.

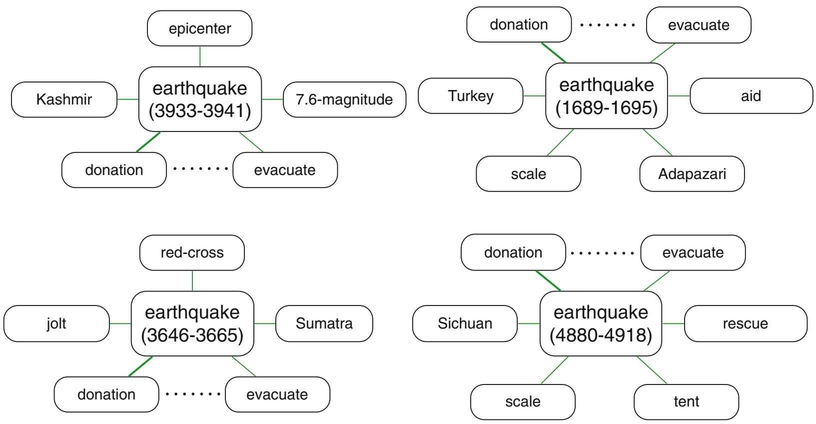

Fig. 1. A BINet example. The numbers in the round brackets denote the burst period of the node. Due to space limitation, we only show *earthquake*'s burst period (days after Jan 1, 1995). Dash lines denote false associations.

In a BINet (Fig. 1), a node is a burst word (including entities and events) with the time span of one of its burst periods, and an edge between two nodes indicates how strongly they are related. Since only burst words are in a BINet,

[2] We treat co-burst as a special case of BSPs.

trivial events are naturally excluded. In a BINet, event BSPs (e.g., *donation* and *earthquake*) can be clearly observed. Moreover, nodes in a community in a BINet are not only topically but also temporally coherent; thus, we can say a community describes an event's topic. Based on a community's structure, it is easy to distinguish major events and minor events for removing false association pairs like (*donation*, *evacuate*) in Fig. 1.

Experiments show the BINet-based approach can not only discover concept-level event associations with comparable precision to text clue based approaches but also discover many important event associations that are not explicitly expressed, and that the BINet and the text clue based approach can nicely complement each other, yielding significant improvement of performance.

2 Burst Information Networks

2.1 Burst Detection

To build a Burst Information Network mentioned in the above section, we first need to detect bursts of words. In general, a word's burst might indicate important events or trending topics. For example, as shown in Fig. 1, the word *earthquake* has a burst from the 4880th to the 4918th days because of a strong earthquake occuring in China on May 12, 2008. For a timestamped document collection $\mathbf{C} = \{D_1, D_2, ..., D_t, ..., D_T\}$, we define a word w's burst sequence $\boldsymbol{s} = (s_1, s_2, ..., s_t, ..., s_T)$ in which s_t is either 1 or 0 to indicate whether the word w bursts or not at time t. Based on the idea of [14,24], this burst sequence can be simply found by searching for the optimal sequence $\boldsymbol{s}^*$ to minimize the cost function defined as follows:

$$Cost(\boldsymbol{s}, \boldsymbol{p}, q^{(0)}, q^{(1)}) = \sum_{t=1}^{T} |\log p_t - \log q^{(s_t)}| + \sum_{t=1}^{T-1} \beta * \mathbf{1}(s_t \neq s_{t+1}) \quad (1)$$

where $\boldsymbol{p} = (p_1, ..., p_t, ..., p_T)$ in which p_t is the probability of the word at t, $q^{(0)}$ is the base probability of the word and it is often defined as the probability of the word on the whole data/corpus, $q^{(1)}$ is the probability of the word in the burst state and it is often defined as $q^{(1)} = \alpha q^{(0)}$ $(\alpha > 1)$.

The first term of Eq. (1) measures the difference between p_t and $q^{(s_t)}$. If a word bursts (i.e., p_t is high), $|logp_t - logq^{(1)}|$ will be smaller than $|logp_t - logq^{(0)}|$ and thus in the optimal sequence $\boldsymbol{s}^*$, s_t tends to be 1; otherwise, s_t tends to be 0. The last term of Eq. (1) is for penalizing transition of burst states through the time to avoid too frequent transition of burst states for smoothing and β is the parameter controlling this part's weight.

Specifically, if a word w is in a burst state at every time t during a period, we call this period as a burst period of w, and w has a burst during this period. In Fig. 1, *earthquake* has 3 burst periods (i.e., (3646–3665), (3933–3941), and (4880–4918)).

Formally, we define $\mathcal{P}_i(w)$ as the ith burst period of the word w. It is a consecutive time sequence (i.e., time interval) during which w bursts at every time epoch t:

$$\mathcal{P}_i(w) = [t_i^s(w), t_i^e(w)]$$
$$\forall t \in \mathcal{P}_i(w) \quad s_t(w) = 1$$

where $t_i^s(w)$ and $t_i^e(w)$ denote the starting and ending time of the ith burst period of w, and $s_t(w)$ denotes the burst state of w at time t.

2.2 BINet Construction

A "*Burst Information Network (BINet)*" represents associations between key facts in a text stream, which has been proven to be effective in multiple knowledge mining tasks [6–8]. The basic component of a BINet is burst elements which are nodes of the information network:

A **Burst Element** is a burst of a word. It can be represented by a tuple: $\langle w, \mathcal{P}_i(w) \rangle$ where w denotes the word and $\mathcal{P}_i(w)$ denotes one burst period of w.

A BINet is defined as $G = \langle V, E \rangle$. Each node $v \in V$ is a burst element and each edge $e \in E$ denotes the association between burst elements. Intuitively, if two burst elements frequently co-occur, then they should be highly weighted. We define $\omega_{i,j}$ as the global weight of an edge between v_i and v_j, which is equal to the number of documents where v_i and v_j co-occur, and $\pi_{i,j}$ as the local weight, which equals to the number of documents in which v_i and v_j form a bigram (i.e., v_i and v_j are adjacent in context). Since a node in BINet contains both semantic and temporal information, nodes in a community are topically and temporally coherent.

3 Event Association Discovery

We first extract events (Sect. 3.1), identify major events for removing false associations (Sect. 3.2) from the BINet, and then rank event associations (Sect. 3.3).

3.1 Event Extraction

Since there is no available open-domain event extraction systems despite some event extractors for limited types of events (e.g., 33 event types in ACE evaluation), we use words in the following list as event trigger words to identify nodes describing events in the BINet, and call the nodes whose word is in the following list **event nodes**:

– Nouns and verbs in frames with *time* attribute in FrameNet[3].
– Trigger word list in ACE evaluation.
– Natural hazards in Wikipedia.

[3] https://framenet.icsi.berkeley.edu/fndrupal/.

Since one node is a unigram[4] (with its burst period), an event node sometimes may not describe an event well (e.g., "test" is too general to describe an event). Hence, for an event node v_i, we try to find its adjacent node to form a bigram to represent the event (e.g., for a node whose word is "test", we may use its adjacent node "nuclear" to represent the event as "nuclear test"). Specifically, we first find the set of nodes locally strongly related to v_i:

$$C(v_i) = \{v_j | \pi_{i,j} > \pi_t\}$$

where π_t is a threshold. Then, we find v_i's most globally related node v_k from $C(v_i)$:

$$v_k = arg\max_{v_j \in C(v_i)} \omega_{i,j}$$

Table 1. Designed POS pattern for event bigram phrase extraction. The bold means it is the head word of the bigram which should be an event node.

Bigram	POS pattern
v_i,v_k	JJ,**NN** | NN,**NN** | NN,**VB** | **VB**,NN

If part-of-speech (POS) tags of v_i and v_k match the patterns in Table 1, then v_i and v_k form a valid event bigram. If we cannot find a v_k making v_i and v_k form an event bigram, v_i is considered as an independent event unigram. Note that if a bigram contains a named entity, we use the type of the entity to replace the entity string for generalization. For example, *Tohoku earthquake* will be replaced with *LOCATION earthquake*.

3.2 Major Event Identification

To identify major events, we first need to detect topics in the text stream and then identify the major event of every topic. As mentioned before, nodes in a community in a BINet describe an event's topic. Therefore, we model topic detection as a community discovery problem.

We first compute PageRank value of nodes in a BINet and rank them by their PageRank values. Note that the weights for PageRank computation are the global weights (ω) of the BINet. Then, we repeatedly choose the node that has the highest PageRank value but does not belong to any community, with its closely related nodes to form a new community $\mathcal{E}$. The algorithm is summarized in Algorithm 1 where $\mathcal{L}$ is the ranking list of nodes by their PageRank values, $V' \subset V$ is the set of nodes that does not belong to any communities, $\hat{\omega}_{v,u}$ is the normalized weight of the edge between v and u, and σ is the threshold for selecting closely related nodes. The algorithm discovers communities greedily and thus is fast.

After topics in a text stream are detected, we identify major events for each topic. Intuitively, a major event must be most frequently mentioned and it should

[4] Here, a named entity is considered as a unigram even if it is composed of multiple words such as *Hong Kong*.

be strongly related to other nodes in its community; thus, its PageRank value should be at the top in the community. Hence, we select the event phrase (unigram or bigram) whose PageRank value is the highest among all event phrases in a community as the major event, as shown in Table 2. Note that a bigram's PageRank value is the average of its words.

Table 2. An example of communities (topics) discovered by our approach. Major events (the bold words) usually have the top PageRank value.

Topic	Key phrases
1	**Iraq war**, Iraqi, US-led, Baghdad
2	**Attack**, terrorist, New York, Washington, Afghanistan
3	**Earthquake**, quake, Wenchuan, Sichuan, quake-hit
4	**Hong Kong return**, motherland, handover, hk
5	Deng Xiaoping, Deng, **death**, condolence, mourn

Algorithm 1. Topic detection

Input: $\mathcal{L}$, $G = \langle V, E \rangle$;
Output: A list of communities: $\mathcal{C} = [\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_k]$
$V' \leftarrow V$
while $\|\mathcal{L}\| > 0$ **do**
  $v \leftarrow \mathcal{L}[0]$ (the first element in $\mathcal{L}$)
  $\mathcal{E} \leftarrow \{v\} \cup \{u | u \in V' \wedge \hat{\omega}_{v,u} > \sigma\}$
  $\mathcal{C}.add(\mathcal{E})$; $\mathcal{L} \leftarrow \mathcal{L} - \mathcal{E}$; $V' \leftarrow V' - \mathcal{E}$
end while

3.3 Event Association Pair Ranking

We select all event pairs in which two events are adjacent in the BINet as candidates and remove (minor event, minor event) pairs which account for most false association cases.

Moreover, we exclude the pairs in which the semantic similarity[5] of two events is higher than a threshold τ because they usually refer to the same event (e.g., *quake* and *earthquake*).

For the remaining pairs, we rank event association pairs (e_1, e_2) using the following metric inspired by Pointwise Mutual Information (PMI):

$$M(e_1, e_2) = \frac{n_{e_1,e_2}}{n_{e_1} \times n_{e_2}} (\log n_{e_1,e_2} + \alpha) \tag{2}$$

where e is an event uni- or bi-gram in Sect. 3.1, n_e is the count of e, n_{e_1,e_2} is the count of cases where e_1 is adjacent to e_2 in a BINet (e.g., $n_{earthquake}$ =

[5] Cosine similarity computed based on word embeddings trained on English Gigaword corpus.

$n_{earthquake,donation} = 4$ for the BINet in Fig. 1), and the factor $(\log n_{e_1,e_2} + \alpha)$ is for promoting event pairs with high support where α is a smoothing parameter for avoiding (2) being 0 if $n_{e_1,e_2} = 1$.

4 Experiments and Evaluations

4.1 Data

We evaluate our approach on 1995–2010 Xinhua news in English Gigaword[6] which contains 1,482,560 news articles.

We used Stanford CoreNLP toolkit to perform POS tagging, lemmatization, named entity recognition, and apply our Burst Information Network (BINet) construction algorithm on this dataset. We remove edges whose global weights are less than a threshold ω_t for reducing noise. The resulting BINet includes 414,944 nodes and 3,699,537 edges.

4.2 End-to-end Evaluation

We discover event associations in an end-to-end fashion. Hyper-parameters ($\sigma = 0.0005$, $\tau = 0.7$, $\pi_t = 5$, $\omega_t = 5$, $\alpha = 0.01$) are tuned on a development set. Totally, we mined 6,084 event association pairs.

We compare the following approaches:

PMI-E: Ranking association pairs by PMI computed over all event words based on their co-occurrence in documents.

PMI-S: PMI of event words are computed based on co-occurrence in sentences.

BINet-E: BINet-based approach without removing false association pairs.

BINet-E+: BINet-based approach where false association are removed.

Text-E: This model extracts causality of event trigger words based on the most commonly used unambiguous causal verbs and connectives, as [20] did, and ranks by frequency. The details of the implementation of this baseline is introduced in the Appendix Section.

Combine: we re-rank the results of BINet-E+ by combining the results of Text-E:

$$\hat{M}(e_1, e_2) = M(e_1, e_2) + \log n_t(e_1, e_2)$$

where $n_t(e_1, e_2)$ is the count of cases where causality of e_1 and e_2 is explicitly expressed by causal verbs and connectives.

In baseline methods, event words include the event bigrams in Sect. 3.1 for fair comparison. We do not compare to [11,12] because their supervised approaches require annotated data that is not publicly available, and do not make a comparison to [23] due to their limited focus on deverbal nouns. [9,10,21] are not compared either because their focus is not mining event associations.

Event association discovery is an open text mining problem and there is no closed gold standard for this task though some knowledge resources (e.g., ConceptNet) can be used as references but they are far from complete. Alternatively,

[6] https://catalog.ldc.upenn.edu/LDC2011T07.

Table 3. Precision of top 500 discovered event association pairs in end-to-end evaluation.

Model	Precision@500
Pmi-E	1.6%
Pmi-S	4.4%
BINet-E	17.2%
BINet-E+	35.6%
Text-E	36.2%
Combine	**43.0%**

we manually evaluate the quality of discovered event associations and use Precision of top K (500) pairs to measure the performance. We do not evaluate recall since it is impractical to find all event associations. We pooled the top K pairs outputted by each system evaluated in this paper for annotation. The annotation[7] is done by 2 annotators who are asked to tell if words/phrases in a pair are associated events by considering whether a word/phrase pair satisfy the event association definition and the association is informative and self-interpretable.

The annotations have fairly good agreement (84.4% overlapping). The difference in the annotators' background knowledge accounts for most annotation disagreement cases. During evaluation, we consider an event association pair as correct if both of the annotators annotate it as correct.

As shown in Table 3, Pmi-E and Pmi-S yield poor performance because many event pairs are either about trivial events or are not associated. Introducing BINets improves PMI-based methods because large numbers of trivial events are excluded. When we remove false association pairs based on the network structure, the performance (BINet-E+) gets significant boost (18.4% gain) and achieves comparable performance to Text-E. When we re-rank the results of BINet-E+ with text clue information, the performance is markedly improved (7.4% and 6.8% gain over BINet-E+ and Text-E respectively), demonstrating these two approaches can well complement each other.

Moreover, we show the performance of models with various Ks in Fig. 2. Text-clue based approach can accurately mine event associations if K is small while its performance drops drastically with K increasing because the number of explicitly expressed event associations is limited. In contrast, the BINet-based approach is more stable, which outperforms the text-based model when K is large. As the results in Table 3, the combination of these approaches improves both of them.

We analyze error cases of the discovered event pairs. The event extraction mistakes are the main source of errors because event extraction is a challenging task, especially for open domains, which affects event association discovery

[7] The annotators mainly used ConceptNet and Wikipedia as references to help with the annotation.

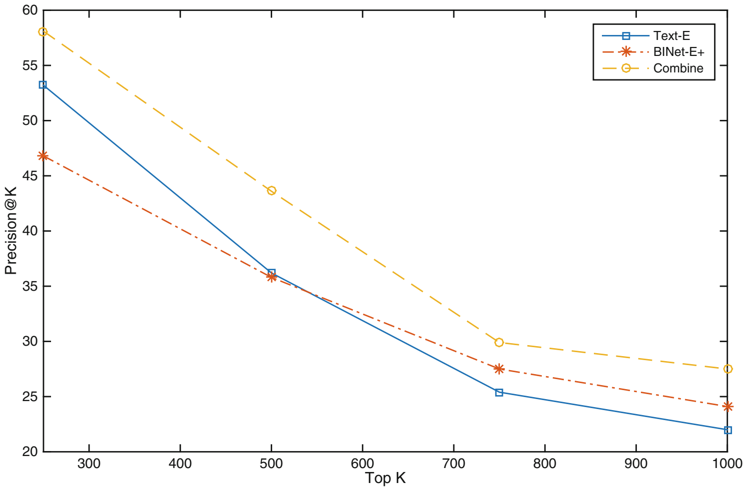

Fig. 2. Precision curves of various models.

results. Another type of errors is that some events are over-generalized because unigram and bigram event representations sometimes are insufficient to describe a complicated event. For example, in the association pair (*financial crisis, impact*), the event *impact* is too general to be informative. In addition, events in some pairs do not satisfy the definition of association (e.g., one event is a part of the other event in a pair like (*match,goal*)).

Table 4. Examples of discovered event association pairs. Of these 32 event association pairs, only 14 (bold) are explicitly expressed by textual clues.

Earthquake	Flood	Financial crisis	Protest
Donation	Divert floodwater	**Shrink**	Election
Landslide	Mine accident	**Financial reform**	Declaration
Humanitarian aid	**Dike breach**	Stimulate economic	**Violence**
Mourn	Remain trapped	**Loan**	Nuclear test
Search	**Evacuation**	Rate cut	War
Death	Rehabilitation	**Slump**	**Conflict**
Evacuation	Flood control	**Plunge**	Invasion
Medical treatment	**Damage**	**Unemployment**	**Arrest**

As a qualitative evaluation, we present examples of event associations discovered by BINET-E+ in Table 4. The event associations are general and useful as knowledge. Moreover, we analyze these 32 event pairs and find only 14 (43.75%) of them are explicitly expressed by the textual clues used in TEXT-E, showing the limitation of the textual clue based approach and the importance of studying BSP-based approaches for this task.

4.3 Future Event Prediction with Association Knowledge

Moreover, we evaluate if the discovered event association knowledge could help us predict future events. We collect 36,129,066 news articles from February to December 2015 on the web. For each event association pair discovered by our approach, we verify if an event in this pair happened after the other. Specifically, if the events in an association pair occurred (burst) one after another within 7 days during this period, we consider this pair helps event prediction.

Table 5 lists the number of association pairs useful for event prediction. Among the top 5,000 event association pairs discovered by our approach, approximately 20% of them help predict events during the period. Specially, we also test those 32 event association pairs in Table 4 which are considered correct. 15 (46.9%) of them help event prediction.

It is notable that the news articles in the corpus are mainly from American and European news agencies and many of them are about events in USA and European countries while news articles in the corpus we used for discovering event associations are from Chinese news agency and they tend to report Chinese local events. Even so, the discovered event associations are still successfully used for prediction, showing that the event association knowledge is general and location-independent.

Table 5. The number of association pairs helpful for future event prediction

Top	500	1000	2000	5000
Predicted	111	197	382	848

5 Related Work

Most work [9–12, 20, 21, 23] related to concept-level event association discovery mainly study extracting causality based on text clues (e.g., causal verbs and connectives). Among them, [9, 10] studied mining textual patterns that describe causal relations, [23] derived event associations by focusing on deverbal nouns within a discourse, [20] proposed to extract event causality and use it to predict future events based on explicit discourse connectives, [21] focused on estimation of the probability that an event occurring after the other given a query event pair, [11, 12] used supervised models to extract event causality, and [15] utilized hierarchical topic structure to capture event associations. In contrast, our approach is unsupervised, efficient, and does not rely on explicit discourse connectives but can nicely complement the textual-based approach. The discovered association knowledge is general and related to important events and thus useful for applications like event prediction and event-centric knowledge base construction.

Another research branch related to this paper is event relation extraction [3–5, 22]. Different from our task that discovers concept-level event associations that can be used as general knowledge, these studies focus on extracting relations between events in a local context (a sentence or a document).

6 Conclusion

We study an open text mining problem – concept-level event association discovery based on burst sequential pattern mining by using a novel graph-based text stream representation, which makes it possible to discover massive implicit event associations and presents chances for event knowledge discovery and event prediction from big data.

Acknowledgements. We appreciate the helpful comments of the reviewers. This work is supported by the National Key Basic Research Program of China (No. 2014CB340504), the Research Fund for the Doctoral Program of Higher Education (20130001110027) and the National Natural Science Foundation of China (No. 61375074, 61273318). The contact author is Zhifang Sui.

Appendix

We introduce how we implement the TEXT-E approach mentioned in Sect. 4. As [20] did, we use the most commonly used unambiguous causal verbs and connectives in Table 6 to extract causality as event associations. We did not use *as* and *after* because *as* is ambiguous, and *after* cannot guarantee that events connected by it are associated according to definition in our paper.

Table 6. Patterns for extracting causality. Note that for *because*, *because of* and *due to*, both of a and b should have a direct path to the causal markers.

Causal marker	Dependency pattern	Instance(a,b)
Because	advcl(a,b)	They **killed** him because he **divulged** the secret
Because of	prep_because_of(a,b)	The election is **postponed** because of the **outbreak** of plague
Due to	prep_due_to(a,b)	The province has **suffered** heavy losses of arable land due to water **erosion** for the past several years
Cause	nsubj(cause,a); dobj(cause,b)	The **earthquake** caused severe **damages** in Japan
	vmod(a, cause); agent(cause, b)	The move is aimed at increasing investment in key sectors and reducing the **burden** caused by inefficient public **enterprises** on the economy
Affect	nsubj(affect,a); dobj(affect,b)	Same with cause
	vmod(a, affect); agent(affect, b)	Same with cause
Lead to	nsubj(lead, a); prep_to(lead, b)	In fact, such **activities** not only harm reforms, national economic development and social stability but lead to high production and construction **costs** for the local economies

References

1. Abe, S., Inui, K., Matsumoto, Y.: Two-phased event relation acquisition: coupling the relation-oriented and argument-oriented approaches. In: COLING (2008)
2. Bethard, S., Martin, J.H.: Learning semantic links from a corpus of parallel temporal and causal relations. In: ACL (2008)
3. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative schemas and their participants. In: ACL (2009)
4. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative event chains. In: ACL (2008)
5. Do, Q.X., Chan, Y.S., Roth, D.: Minimally supervised event causality identification. In: EMNLP (2011)
6. Ge, T., Cui, L., Chang, B., Li, S., Zhou, M., Sui, Z.: News stream summarization using burst information networks. In: EMNLP (2016)
7. Ge, T., Cui, L., Chang, B., Sui, Z., Zhou, M.: Event detection with burst information network. In: COLING (2016)
8. Ge, T., Dou, Q., Pan, X., Ji, H., Cui, L., Chang, B., Sui, Z., Zhou, M.: Aligning coordinated text streams through burst information network construction and decipherment. arXiv preprint. arXiv:1609.08237 (2016)
9. Girju, R.: Automatic detection of causal relations for question answering. In: Workshop on Multilingual Summarization and Question Answering (2003)
10. Girju, R., Moldovan, D.I., et al.: Text mining for causal relations. In: FLAIRS Conference (2002)
11. Hashimoto, C., Torisawa, K., Kloetzer, J., Oh, J.H.: Generating event causality hypotheses through semantic relations. In: AAAI (2015)
12. Hashimoto, C., Torisawa, K., Kloetzer, J., Sano, M., Varga, I., Oh, J.H., Kidawara, Y.: Toward future scenario generation: extracting event causality exploiting semantic relation, context, and association features. In: ACL (2014)
13. Hashimoto, C., Torisawa, K., Kuroda, K., De Saeger, S., Murata, M., Kazama, J.: Large-scale verb entailment acquisition from the web. In: EMNLP (2009)
14. Kleinberg, J.: Bursty and hierarchical structure in streams. Data Min. Knowl. Disc. **7**(4), 373–397 (2003)
15. Li, R., Wang, T., Wang, X.: Tracking events using time-dependent hierarchical dirichlet tree model. In: SDM (2015)
16. Mirza, P., Tonelli, S.: An analysis of causality between events and its relation to temporal information. In: COLING (2014)
17. Mulkar-Mehta, R., Welty, C., Hoobs, J.R., Hovy, E.: Using granularity concepts for discovering causal relations. In: FLAIRS (2011)
18. Oh, J.H., Torisawa, K., Hashimoto, C., Sano, M., De Saeger, S., Ohtake, K.: Why-question answering using intra- and inter-sentential causal relations. In: ACL (2013)
19. Pantel, P., Bhagat, R., Coppola, B., Chklovski, T., Hovy, E.H.: ISP: learning inferential selectional preferences. In: HLT-NAACL (2007)
20. Radinsky, K., Davidovich, S., Markovitch, S.: Learning causality for news events prediction. In: WWW (2012)
21. Radinsky, K., Horvitz, E.: Mining the web to predict future events. In: WSDM (2013)
22. Riaz, M., Girju, R.: Another look at causality: discovering scenario-specific contingency relationships with no supervision. In: ICSC (2010)
23. Tanaka, S., Okazaki, N., Ishizuka, M.: Acquiring and generalizing causal inference rules from deverbal noun constructions. In: COLING (2012)
24. Zhao, W.X., Chen, R., Fan, K., Yan, H., Li, X.: A novel burst-based text representation model for scalable event detection. In: ACL (2012)