Automatic Detection of Discourse Indicating Emerging Risk

Critical Approaches to Discourse Analysis across Disciplines
http://cadaad.net/journal
Vol 4 (2): 171 – 179
ISSN: 1752-3079

ÁGNES SÁNDOR

Xerox Research Centre Europe

Agnes.Sandor@xrce.xerox.com

Abstract

Detecting emerging risk is a major concern in the 'risk society' we live in. Risk can be detected among other sources from discourse describing events. Automatic language processing tools can help monitoring large amounts of electronic text, and recent advances in syntactic and semantic processing allow fine-grained analyses that produce normalized event descriptions, which can be used in risk detection. We have implemented normalized event extraction in the Xerox Incremental Parser.

We propose filtering all the normalized event descriptions in order to get events that indicate emerging risk based on two theories of detecting weak signals of emerging risk: one based on scenario models, implemented in the tool EventSpotter, and the other on detecting events that show characteristic features of weak signals.

In this article we describe the three modules (normalized event extraction and two ways of filtering) and propose them for industrial application as well as for social scientists involved in the analysis of discourse on risk.

1. Introduction[1]

The term 'risk society' was first coined by Ulrich Beck as far back as in 1986 - the year of the Chernobyl disaster. This term has been used ever since reflecting that the concept of risk is widely present in every facet of contemporary society: we are continuously exposed to political, environmental, financial, economic, health-related etc. risks (see e.g. the wide range of 'risk' topics on in this volume). Risk is possible danger in the future. The earlier we become aware of it the more effectively we can take measures to prevent it. Thus detecting emerging risk has become a major concern and challenge in our society.

Risk detection is formalized in strategic early warning systems (SEWS):

> to assist organizations in dealing with discontinuities or strategic "surprises". It is based on methods for detecting "weak signals" (Igor Ansoff, 1975), which can be perceived as important discontinuities in an organizational environment [...]. The underlying assumption of SEWS is that discontinuities do not emerge without warning. (Wikipedia)

Although SEWS were originally conceived for managing business organizations, they are also applied for detecting risks in any facet of the 'risk society'. SEWS are thus being developed and used in homeland security (Delavallade et al. 2007), health-programs (Smith et al. 2007), economy (Goldstein et al. 1999), etc.

The first task for building an early warning system is the detection of the weak signals. Weak signals can be detected either directly by monitoring the environment or indirectly by monitoring descriptions of events in the environment. The work that we are presenting in this paper follows the latter approach: we monitor event descriptions in written discourse with the aim of automatically detecting weak signals, of emerging risk, i.e. descriptions of events that potentially lead to risk. We use EventSpotter, an application of XIP (Xerox Incremental Parser), a natural language parser for syntactic and semantic analysis.

The general domain-independent strategy for detecting descriptions of events as weak signals is the following: As a first step EventSpotter provides a normalized form of all event descriptions. A normalized event description consists in a table whose columns are the name of the event, its various coordinates, when they are mentioned in the text, the source of the event description as well as its factuality according to the source. Once all the events are normalized, we further process them in order to point out events that are potentially weak signals of emerging risk. In this process we follow two different approaches: detecting risk by matching scenarios of early warning events designed by scenario experts or by spotting individual early warning signals that are recognized by some characteristic features (Hiltunen 2006). We have applied EventSpotter in the detection of weak signals of nuclear proliferation. This application follows the scenario-based approach, and is integrated into the early warning system ADAC for monitoring news articles. Our module for recognizing weak signals by some characteristic features has not been integrated and tested in a real-life application yet.

In the consecutive sections of the paper we describe the principles and techniques that we use for normalized event detection, for extracting events that fit scenarios and finally those that are considered as individual early warning signals.

2. Normalized Event Extraction

As we have said above, the aim of normalized event detection is to assign an event name and some coordinates whenever they are mentioned in the text for every event described, as well as a source and the factuality of the event according to the source. The coordinates are AGENT, TARGET[2], PLACE and TIME. The factuality is FACTUAL or NOT FACTUAL.

In terms of dependency parsing this consists in mapping grammatical categories and functions into the normalized categories. Thus event names – which we call event cores – are verbs and event nouns, agents are subjects, modifiers of event nouns and various other grammatical functions. Thus sentence (1) is normalized as Table 1:

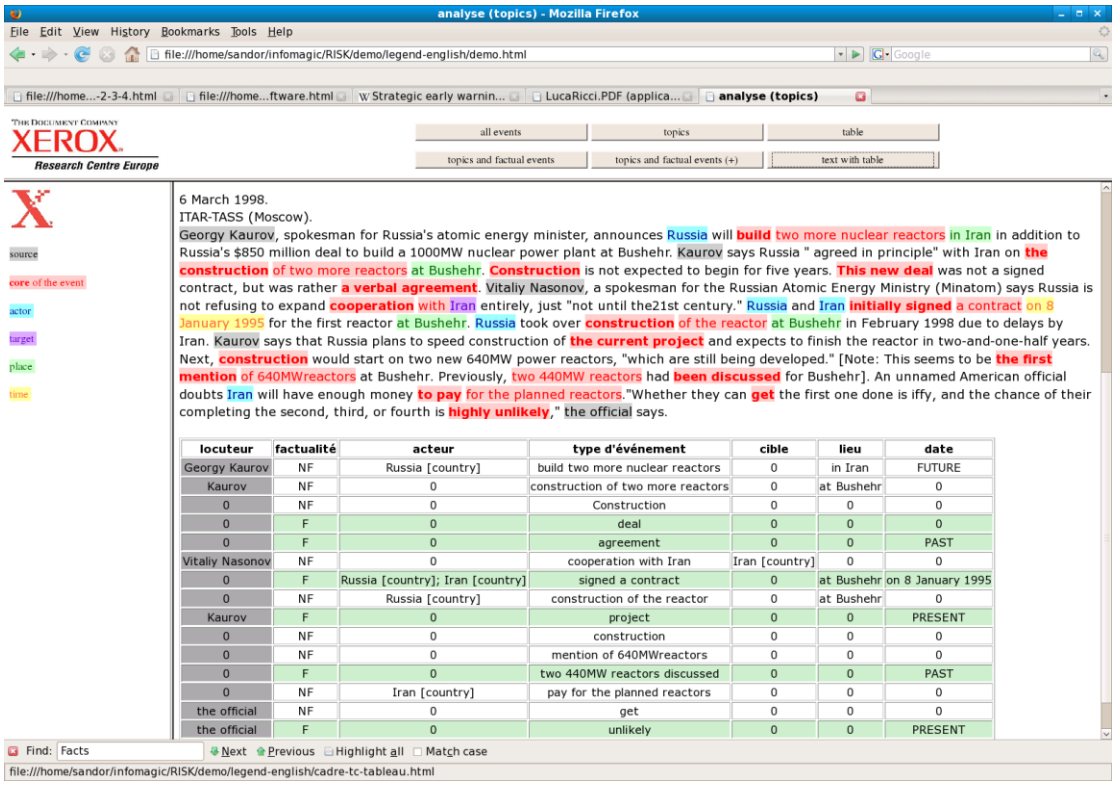
(1)
After a meeting with Iranian Minister of Mines and Metals Eshaq Jahangiri, Russian First Deputy Prime Minister Nikolai Aksyonenko tells journalists that Gholam Reza Aqazadeh, the chairman of the Atomic Energy Organization of Iran, would visit Moscow in August or September. (Itar-Tass, 28 July 1999)

Table 1. Normalized table of sentence (1)

source	factuality	actor	core	target	place	time
Itar-Tass	F	Nikolai Aksyonenko	meeting	Eshaq Jahangiri		
Itar-Tass	F	Nikolai Aksyonenko	tell	journalists		
Nikolai Aksyonenko	NF	Gholam Reza Aqazadeh	visit	Moscow	Moscow	in August or September

We have created a user interface for consulting events described in news articles (see Table 2).

Table 2. User interface for consulting event descriptions



Besides providing the basis for subsequent processing normalization makes it possible to query the table for precise information about various facets of events. The user can query for participants in particular events, like negotiations, for different sources quoting the same event, for factual event descriptions (i.e. events that actually took place according to the source) related to a particular location.

We carried our a small scale evaluation on 50 news articles containing 149 events (Capet et al. 2007). The results are reproduced in Tables 3a. and 3b.:

Table 3a. Evaluation results

	Source	*Factuality*
Correct	*131*	*122*
Errors	*8*	*17*

Table 3b. Evaluation results

	Core	*Actor*	*Target*	*Place*	*Time*
Precision	*99*	*80*	*78*	*85*	*63*
Recall	*95*	*71*	*41*	*57*	*63*

3. Matching scenarios

We use the term scenario as describing a set of event types whose co-occurrence may lead to crisis. Scenario analysis is a method in futures studies and refers to a 'a hypothetical sequence of events constructed for the purpose of focusing attention on causal processes and decision points' (Kahn and Wiener 1967). However, we have used scenarios in our system for describing templates of event types whose co-occurrence led to crisis in the past. In this context the event types contained in scenarios are weak signals for future risk based on models of past crises. The scenario that we used in the early warning system of nuclear proliferation was constructed by scenario experts (Capet et al. 2008).

EventSpotter filters the normalized event descriptions by retaining only those events that match the pre-established list of weak signal event types. This processing is carried out using the concept-matching framework (Sándor 2007). For technical details refer to (Capet et al. 2008). Some event types are listed in Table 4:

Table 4. List of event types

Event types as weak signals	
1	to work on secret nuclear programs
2	to work on secret nuclear equipment
3	to sell military equipment
4	to negotiate a secret nuclear related agreement
5	to get involved in a cooperation in the nuclear domain
…	

The output of EventSpotter is illustrated by sentences (2) through (4), which are assigned to the event type 'TO GET INVOLVED IN A COOPERATION IN THE NUCLEAR DOMAIN':

(2)
A delegation from Syria arrives in Iran to begin negotiations on a possible Iranian-Syrian nuclear pact.

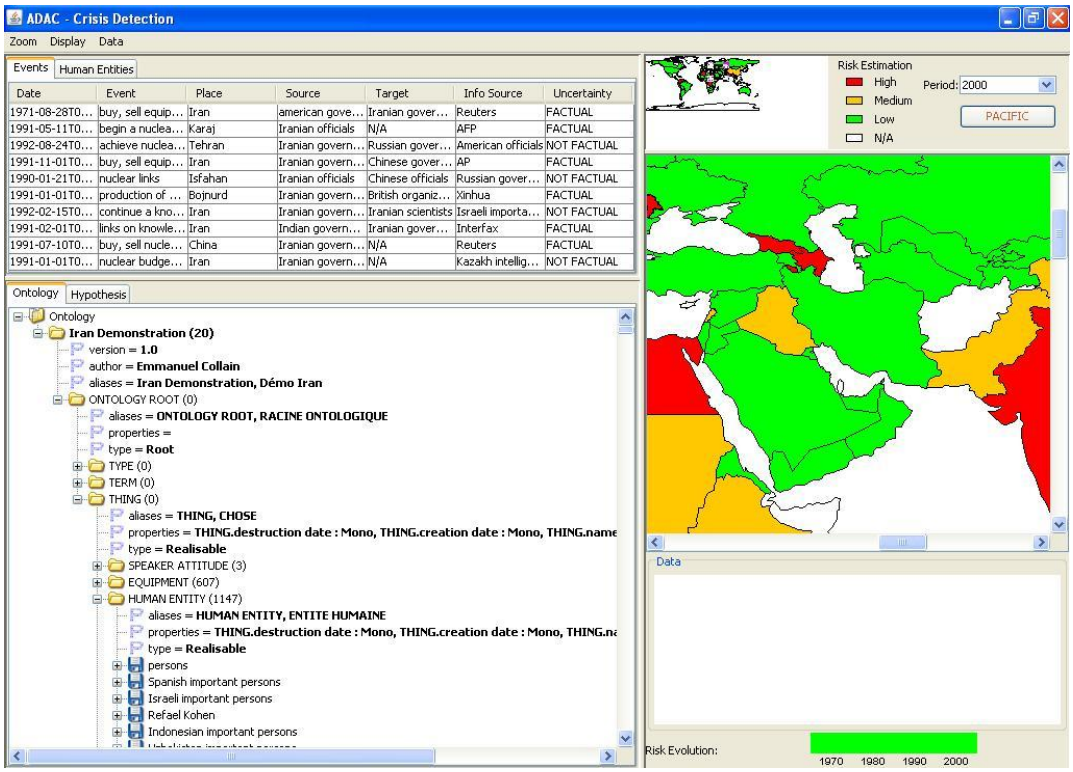
(3)
The Middle East Newsline reports that Iran is preparing to receive a light water nuclear reactor from Russia.

(4)
Former chief nuclear negotiator for Iran Hassan Rowhani says Tehran is ready to negotiate a mutual start for the Natanz nuclear facility.

EventSpotter is integrated into the early warning system ADAC. ADAC has a recognition engine that computes the probability of risk based on 'the degree of match between event sequences and experts' scenarios, taking into account some, spatial, temporal and operational constraints' (Capet et al. 2008). Table 5 shows the user interface of ADAC. The table of the automatically extracted normalized event types extracted from incoming news articles is in the upper left zone. If the user validates an event extracted, it will be automatically inserted into the scenario template of the lower left zone. The right hand part of the interface is designed for the presentation of the results, i.e. the probability of nuclear proliferation leading to crisis situations. This part of the interface is not implemented at this point.

EventSpotter as implemented in ADAC detects event types that are parts of scenario templates of past events, which is a theoretical limitation for detecting risk arising from new situations. This approach could be completed by taking into account possible future scenarios as those carried out in futures studies. The methodology used in EventSpotter could be used with any types of scenarios.

Table 5. User interface of ADAC



4. *Individual events as weak signals*

In this section we present an alternative approach to filtering events, in which the normalized event descriptions detected are spotted as weak signals due to some general characteristic features. Like in the application of the scenario-based approach our goal is not to establish a list of characteristic features but to provide a tool for testing experts' hypotheses and propositions of such features. We obtain a list of characteristic features from definitions of weak signals as well as from a list of synonyms of the term 'weak signal':

> … weak signals mean today's information that can foretell the changes in the future. This information might sound funny or strange and it can cause confusion, because it offers a totally new way of thinking/idea/innovation. (Hiltunen 2008)

> Examples of such terms are "distress" […], "emerging issues" […], "surprise" […], "dramatic development" […], "dramatic short-term events" […], "shocks" […], "unique events" […], "sudden events" […], "disruptive events" […], "unprecedented developments". (Sleegers 2003)

Based on these characteristic features of weak signals, we have constructed an 'event tagger' that labels events according to the following features:

- SHORT_TIME (cf. 'dramatic short-term events', 'sudden events')
- NEW_KNOWLEDGE (cf. 'new way of thinking/idea/innovation', 'unprecedented developments')

- TENDENCY (cf. 'emerging issues')
- EMOTION (cf. 'might sound funny or strange', 'surprise')
- WARNING (cf. 'can cause confusion','distress')

The event tagger takes advantage of the author's discourse for labeling the event descriptions (Sándor 2006). This means that an event will be labeled e.g. NEW_KNOWLEDGE if the author conveys this comment, but will not be labeled as such if she does not, even if the event described strikes the reader as new knowledge. When applied on news articles, event labeling uses the journalist's insight for the recognition and expression of events that 'make news' (cf. Uskali 2005). The following sentence for example is tagged as WARNING by the system due to the comment 'which have caused concern':

(5)
President Ahmadinejad just announced a plan to expand the number of centrifuges by 6,000, in addition to the 3,000 already operating, which have caused concern in the international community.

Event tagging could be used in conjunction with classical queries in order to obtain weak signals concerning particular topics in early warning systems. The following excerpt from a news article about the dangers of fluoridation of water is tagged as follows (the query and the expression conveying the characterization of weak signals is underlined):

(6)
SHORT_TIME_Fluoride recently began flowing through the tap water into millions of Southern California households. WARNING_But the pipeline of information to warn the public about fluoride exposure is apparently clogged up with something. ... TENDENCY_Watch for fluoride toothpaste, mouthwash, dental treatments, soft drinks, juice, commercially raised fruits and vegetables (grown with fluorine-containing pesticides, herbicides and fertilizer), processed and canned food, wine, beer, coffee and tea, not to mention increasing fluorine pollution in the environment, to name a few.

For event labeling we use XIP and the same processing framework, concept-matching, as for scenario processing. Being based on syntactic analysis this framework yields hits that exactly match the query subject matter being a weak signal. The following sentences both contain the query fluoride and the characteristic keywords of weak signals, however they are syntactically unrelated. Thus these sentences are not spotted as weak signals related to fluoride.

(7)
This Health Canada is warning that counterfeit toothpaste products, falsely labelled as Colgate Fluoride Toothpaste Herbal and Colgate Fluoride Toothpaste Maximum Cavity Protection, have been found to contain high levels of harmful bacteria.

(8)
Cavities Not Increasing, But Decreasing, When Fluoridation Stops

This system could be completed with other characteristic features of weak signals and applied in monitoring news articles covering particular topics.

5. Conclusion and Future Work

We have presented a natural language processing system for detecting normalized event descriptions augmented by two independent systems that apply experts' strategies for detecting events that can be considered as weak signals of emerging risk. The two subsequent systems are based on scenario templates and characteristic features of weak signals respectively. All the three systems are based on syntactic analysis augmented by advanced semantic analysis. EventSpotter, the system detecting scenario templates has been integrated into the early warning system ADAC, whereas the other system has not been tested in a real-life application yet. Expert users might propose different scenarios or different characteristic features from those presented here. The implementation of systems with different user needs necessitates some lexical work without interfering with the basic infrastructure of the systems. Future work consists in the development of a user interface that allows subject-matter experts to construct and evaluate new systems.

We propose these tools for both industrial applications and for social scientists.

Notes

[1] The integration of EventSpotter into the early warning system ADAC has been carried out in the framework of the Infom@gic project supported by the French National Research Agency (ANR).

[2] TARGET is a general term for all the participants other than the AGENT.

References

Ansoff, H.I.(1975). Managing strategic surprise by response to weak signals. *California Management Review* 18 (2): 21-33.

Capet, Ph., Delevallade, Th., Jacquelinet, J., Martineau, C., Nakamura, T., Rebotier, A., Sandor, A. and Voyatzi, S. (2007). Intelligent awareness: Event extraction, information evaluation and risk assessment. In Zygmunt Vetulani (ed.), *Proceedings of the 3rd Language and Technology Conference*. Poznan: Widawnictwo Poznanskie. pp. 539-543.

Capet, Ph., Delevallade, Th., Nakamura, T.,, Sandor, A., Tarsitano., C. and Voyatzi, S. (2008). A risk assessment system with automatic extraction of event types. In IFIP, *Intelligent Information Processing IV*. Boston: Springer. pp. 220-229.

Delavallade, T., Mouillet, L., Bouchon-Meunier, B. and Collain, E. (2007). Monitoring event flows and modelling scenarios for crisis prediction: Application to ethinc conflict forecasting. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 15: 83-110.

Hiltunen, E. (2006). Was it a wild card or just our blindness to gradual change? *Elina Journal of Futures Studies* 11 (2): 61 - 74.

Hiltunen, E. (2008). Good sources of weak signals: A global study of where futurists look for weak signals. *Journal of Futures Studies* 12 (4): 21-44.

Kahn, H. and Wiener, A.J. (1967). *The Year 2000*. New York: MacMillan.

Goldstein, M., Kaminsky, G. and Reinhart, C. (1999). *Assessing Financial Vulnerability: An Early Warning System for Emerging Markets*. Washington DC: Institute for International Economics.

Sándor, Á. (2007). Using the author's comments for knowledge discovery. Semaine de la Connaissance: Atelier Texte et Connaissance. Nantes. June 29.

Sándor, Á. (2007). Modeling metadiscourse conveying the author s rhetorical strategy in biomedical research abstracts. *Revue Française de Linguistique Appliquée, No. 2*. pp. 97-109.

Sleegers, A.M. (2003). *The Future Shocks*. Masters thesis. International Business Studies, University of Maastricht. http://www.icis.unimaas.nl/publ/downs/The_Future_Shocks.pdf

Smith, G., Hippisley-Cox, J., Harcourt, S., Heaps, M., Painter, M., Porter, A. and Pringle, M. (2007). Developing a national primary care-based early warning system for health protection: A surveillance tool for the future? Analysis of routinely collected data. *Journal of Public Health* 29 (1): 75-82.

Uskali, T. (2005). Paying attention to weak signals. *Innovation Journalism* 2 (11). http://www.innovationjournalism.org/archive/INJO-2-11.pdf