World Applied Sciences Journal 24 (Information Technologies in Modern Industry, Education & Society): 133-137, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.24.itmies.80025

Processing of Spatial and Temporal Information in the Text

Alexander S. Dmitriev, Alla V. Zaboleeva-Zotova, Yulia A. Orlova and Vladimir L. Rozaliev

Volgograd State Technical University, Prospect Lenina, 28, Volgograd, 400005, Russia

Submitted: Aug 7, 2013; **Accepted:** Sep 18, 2013; **Published:** Sep 25, 2013

Abstract: The paper presents a mining methodology, which is based on processing of spatial and temporal information in natural language text. It is based on methodology capable of recognizing the relationship between events and objects in the text, using spatial and temporal relationships between these events to enhance the semantic coherence of natural language text. Heterogeneous semantic network with extended family relationships is used to implement relational-situational analysis between the words in the sentence. For correlation lining between spatial and temporal categories and events in the text the mechanism of Markov-logic networks is used, allowing to operate in the peaks of Markov-logic networks with the normal logical formulas representing the combined spatial-temporal logic. The described system can be used in tasks of semantic search.

Key words: Text mining · Natural language processing · Heterogeneous semantic network · Topo-temporal logics · Markov-logic networks

INTRODUCTION

It is well known that in recent years there are a large number of attempts (and sometimes quite successful) to create a modern intelligent systems [1-3], semantic text analysis [4, 5] and natural language speech [6]. Existing systems with varying degrees of success to cope with their tasks, but their general and chief drawback is that the semantics of the text is based primarily on the final statements and proposals [7, 8]. In this case, the meaning of the text is considered very rare. Such an approach has a strong distortion of the information described in the text, for example, if a fact is described in several successive or scattered sentences.

Naturally, to solve this problem it is necessary to use the technique of building storylines, or in other words, the chain of events. That event in the text plays a key role in a bunch of different proposals. As a rule, the events are described by verbs or verbal formations. Having defined all the events of the text, using text analysis method based on communicative grammar, one can try to build up between cause and effect relationship that will accurately describe the semantics of the text, rather than individual sentences. In addition, the use of spatial relations between objects complements event model physical model of relations between objects, which are described in the text.

Based on the aforementioned, it was decided to develop a methodology capable of recognizing the relationship between events and objects in the text, using spatial and temporal relationships between these events to enhance the semantic coherence of natural language (NL) text.

Methods for Describing the Spatial and Temporal Relations in the Text: According to the work of Vsevolodova [9], the definition of spatial relations in the text is gotten based on the so-called opposition. On the first level of decomposition of all existing locative groups in the Russian language, the most important is the relation of the localized object to lokum, which is based on finding the object at any time within the lokum. In that case, if the object at any given time was or will be within the lokum, we can say about unispatial seme, for example, "in the field, from the field, through the field". Otherwise, if the localized object is not found, or will not be found within the lokum, then we can talk about the relation of non-unispatial seme "near the field, two kilometers away from the field, to the field, by the field".

In the construction of the temporary order, a lot of grammatical categories is used. These include species-tense forms of verbs, adverbs of time, lexical and semantic information and understanding of the knowable world.

Corresponding Author: Rozaliev, Volgograd State Technical University, Prospect Lenina, 28, Volgograd, 400005, Russia.

Most of the events are described through the structure of verbs. It is known that the species-tense forms of verbs impose restrictions on the temporal order of events (past, present, future, perfect and imperfect form).

For events that are described not using the structure of verbs, i.e. with a noun, these settings are ignored.

Apart from these parameters, the events can be classified into different classes (of which the main ones are the states, processes, instantaneous event and the events in the development), methods of verbal action.

They can also be classified into modalities (i.e. the possibility, probability or obligation to perform an action) – "In this week, one can expect the rise of exchange rates of main shares".

Polarities (i.e., the positivity or negativity of the action) - "Until the firing stops, there will be no peace talk".

Temporary conjunctions are used to link two events. They often appear in complex sentences and describe the relationship between the parts. For example, before, after, during, since when as and etc.

Communicative Grammar Using Categories of Space and Time: In this work of natural language processing of text in order to identify and extract the spatial and temporal data using communicative grammar of the Russian language, developed by Zolotova. The main idea of this grammar is that the syntax and semantics are closely linked in the sense of the sentence analysis.

The main term of communicative grammar is syntaxeme. It is a word or phrase, the value of which is determined by the categorical meaning of the word and morphological forms, which in turn are realized in a specific syntactic position. The sense of a sentence is defined by a set of values contained in its syntaxemes and relations between them.

Relational-situational analysis is used to search for values of syntaxemes and the relationships between them. The main principle of its work is the search of predicate words (verbs and verbal units) based on the semantic dictionary. Then the search of syntaxemes values by means of matching predicate word and compiled by the parser dependency tree. Linguistic rules and semantic dictionary compiled by language experts.

Heterogeneous semantic network with extended family relationships, as proposed by Osipov [10], is used to implement relational-situational analysis, in the intelligent search system Exactus. The vertices of the network are the values of syntaxemes, ribs - the relationship between syntaxeme.

After setting the values of the syntaxemes, the relations are determined, which include certain pair of syntaxemes. Thus there is a circuit - a relationship in the semantic network are updated with new links.

In the described methods of analysis of natural language text, despite the fact that an analysis based on communicative grammar is one of the most advanced in the semantic analysis, there is a significant drawback. This analysis can identify the semantics within a single sentence or utterance. To review and make a semantic model of the text, this technique cannot.

On this basis, it was decided to extend the analysis of natural languages based on communicative grammar event of additional categories, which are expressed throughout the text. This will create a more complete semantic models of texts where the meaning is not disclosed for each individual proposal and, if possible, most of the analyzed text [11].

The space-temporal relationship between the predicate and syntaxeme words in sentences is used to establish the connection between the sentences. Structurally, it is an extension of heterogeneous semantic network.

For example consider the phrase: "Vasya was late to school. He cut a path through the courtyards." (Figure 1).

In the figure, the solid lines show the values of syntaxemes and the dash shows the relationship between syntaxemes. Event category is used to set the semantic relationship between the two proposals. In the narration of the text, as a rule, either sequentially or oscillates about a particular storyline. Accordingly, the text traces the chain of events that alternately succeed each other. To identify event, change the spatial and temporal data related to these events are used.

In contrast to the construction of a single utterance, semantics definition of the relationship between the different events of the proposals is a more difficult task. We use to address the problem of machine learning mechanism based on logic and Markov networks.

Combining Spatial Logics with Temporal Logics: To determine the meaning of the final sentence, heterogeneous semantic network that reveals the events in a single sentence is used. Having a semantic network, we get a large set of logical rules, among which are the rules that describe the spatial and temporal information. But this information is fairly little to do with each other, so it is advisable to use a combination of temporal logic and spatial relationship, so that the events described in the sentences would be the most accurate model in space and time.

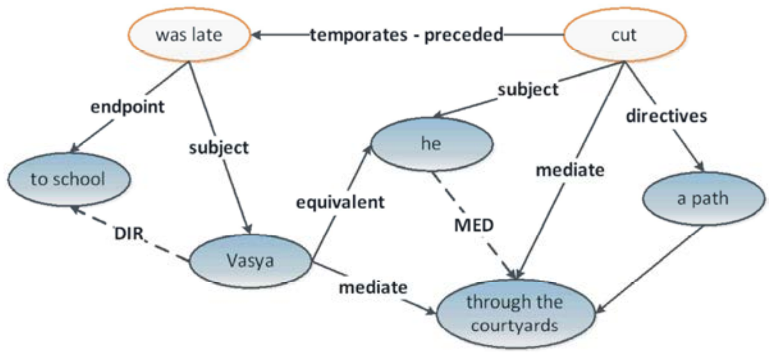

Fig. 1: Example of heterogeneous semantic network

In addition, the use of combined spatial and temporal (or otherwise topology temporal) logic reduces the consumption of computing resources. The use of logic in each individual with further attempts to link the results of their work with events requires significant time costs. Anyway, the task of building an effective and sufficiently complete topology of temporal logic requires careful study because its different combinations are fundamentally different classes of complexity.

On the basis of the spatial relationship analysis, topological logic is taken. The logic of topological spaces at the moment is one of the most successful approaches to the description of spatial relations in artificial intelligence. But so far, effective interaction model of spatial and temporal relations has not been made, as usual addition of spatial and temporal logic does not give the desired result. This is due primarily to the problems of demonstrability and maintenance of dynamic system conditions. The most comprehensive study on combining spatial and temporal logics, in studies is conducted by David Gabelaia, Roman Kontchakov, Agi Kurucz, Frank Wolter, Michael Zakharyaschev, etc [12].

We use spatial logic based on propositional logic, in which unary predicates represent spatial objects and topological relationships between them are represented by internal operators and circuits, universal and existential quantifiers in space and the usual logical operations. This logic is called a modal logic and is regarded as the logic of topological spaces. Designated as $S4_u$.

Spatial terms of this logic are expressions of the form:

$$\tau = p_i \mid \overline{\tau} \mid \tau_1 \cap \tau_2 \mid \tau_1 \cup \tau_2 \mid I\tau \mid C\tau. \tag{1}$$

where p_i - space variables, *I, C* - interior and closure operators.

In this work, to describe the reduction to a one-dimensional spatial objects, the notion of regular closed sets (or simply "Regions") is used. The RCC-8 language is used to describe the regions. For example, predicate $DC(r, s)$ means that regions r and s are not connected.

Arguments RCC-8 predicates - regional variables are interpreted as regular closed sets (regions) of topological spaces.

In this work, temporal logic is represented by the linear temporal logic LTL. Temporal logic is an approach to reasoning about time, using the temporal ligaments without specified amount of time.

The mark-up over time to LTL is any strict linear sequence (W, <), with the points of time and the ratio of precedence <. LTL formulas are constructed from propositional variables p_0, p_1 using logical operators and temporal operator *U* - "Until". Other temporal operators: *Rf* - sometimes in the future, *Ff* - always in the future *Nf* - the next moment.

Next, we consider the combination of spatial and temporal logic. The movement of objects in space-time is represented as a model of "screen shot", i.e. at each time the current position of the object is fixed. Topological-temporal model is a pair of $Mod = (P, DT)$, where $P = (U, I)$ - a topological space and *DT* is the set of spatial points *p* at each time point $n \in N$.

As a result we receive logic in which the temporal operators will be eliminated, except for *Nf*, i.e. exception of principle of variability of objects on all stretch of time, we eliminate operator *Rf* and *Ff*. Received logic LTL*RCC-8 though is less indicative, than, satisfying to 3 principles described above, logic LTL*$S4_u$, but thus the output from it is carried out repeatedly faster - PSPACE against ExpSpace. It is easy for checking up, carrying out the analysis of algorithm of the resolution from two types of logic.

For topological-temporal logic LTL*RCC-8, the following terms and formulas are entered:

$$\rho = CIp \mid CI(\rho_1 U \rho_2)$$
$$\tau = \rho \mid \overline{\tau} \mid \tau_1 \cap \tau_2 \mid \tau_1 \cup \tau_2 \mid I\rho \mid Nf\,\tau \tag{2}$$
$$\varphi = Q(\rho_1, \rho_2) \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 U \varphi_2$$

where - ρ is the regional term, φ - the formula of combined logic and Q - 8 RCC-8 predicates.

Identification of Spatial and Temporal Relations in the Text: For the description of space and temporal expressions, the following model is presented:

$$R{:}TPF \to TPR(T,P)$$
$$R{:}TPF \to TPR(e_1,e_2) \tag{3}$$

where R - functional correspondences, TPF - parameters of temporal and space relations, T = {*te, tr, ts*} (*te* - an event interval, *tr* - a reference point, *ts* - a speech interval), P = {*pe, pr*} (*pe* - an event scene of action, *pr* - a counting place).

TP (*T, P*) = {BEGIN&IN, END&ON, FUTURE & ALONG etc.}, i.e. the relation between time and an event place and also a place of counting and one of two countings of time (time of counting and speech time).

$TPR(e_1,e_2)$={*AFTER&TO*, *BEFORE&ON*, *DURING&IN*}

Identification of temporal and space relations is produced by means of a technique of machine training. For this purpose, the system on an input receives two types of a corpus. The first corpus - the labeled corpus is annotated syntactically, the second – the designation of temporal and space relations are manually placed. For correlation lining between spatial and temporal categories and events in the text the mechanism of Markov-logic networks (MLNs) is used, allowing to operate in the peaks with the normal logical formulas representing the combined spatial-temporal logic. In the further operation of algorithm the network partition on the atomic expressions used in an ordinary Markov network is produced. The Markov network carries out the task of classification by assignment of those or other events to a certain chain of events.

MLNs can answer the question of the arbitrary form "What is the probability that formula F1 is true taking into account that formula F2 is true?" If F1 and F2 are two formulas in the logician of the first order and the C is a finite dial-up of constants, including any constants which appear in F1 or F2 and L is MLNs, then:

$$P(F_1 \mid F_2, L, C) = P(F_1 \mid F_2, M_{L,C}) = \frac{P(F_1 \wedge F_2 \mid M_{L,C})}{P(F_2 \mid M_{L,C})} = \frac{\sum_{x \in X_{F_1} \cap X_{F_2}} P(X = x \mid M_{L,C})}{\sum_{x \in X_{F_2}} P(X = x \mid M_{L,C})} \tag{4}$$

where x_{Fi} is the set "worlds" (interpretations) is where the formula F_i is fulfilled.

Substituting in a network of the statement which are necessary for checking up (for example, whether an event belongs to the given chain of events) we, leaning against the trained network, we can receive with some level of reliability the answer to an event accessory to a certain chain of events.

CONCLUSION

The described system can be used in tasks of semantic search as usage of the expanded event categories which allows to derive sense not only from specific sentence, but also from the coherent texts in which the accurate sequence of events is tracked.

ACKNOWLEDGEMENT

This work is supported in part by the Russian Foundation for Basic Research 12-07-00266-a, 12-07-00270-a, 13-07-00351-a, 13-07-00459-a, 13-07-00461-a.

REFERENCES

1. Zaboleeva-Zotova, A., Y. Orlova and A. Petrovsky, 2012. Automation of software design on initial stages. In the Proceedings of the Conference Advances in Decision Technology and Intelligent Information Systems, pp: 26-30.
2. Zaboleeva-Zotova, A., Y. Orlova and V. Rozaliev, 2012. The Program Complex Automation of the Initial Stages of Software Engineering. In the Proceedings of the Workshop Operations Research and Data Mining, pp: 65-73.
3. Orlova, Y. and V. Rozaliev, 2011. Increase efficiency of software engineering by improving interaction with the user. In the Proceedings of the VI International Scientific-Practical Conference Integrated models and soft computing in artificially intelligence, pp: 907-914.
4. Zaboleeva-Zotova, A. and Y. Orlova, 2008. Computer-aided System of Semantic Text Analysis of a Technical Specification. Advanced Research in Artificial Intelligence: Supplement to International Journal IT and Knowledge, pp: 139-145.

5. Zaboleeva-Zotova, A. and Y. Orlova, 2009. Computer Support of Semantic Text Analysis of a Technical Specification on Designing Software. Intelligent Processing: Supplement to International Journal IT and Knowledge, pp: 29-35.
6. Orlova, Y. and V. Rozaliev, 2011. Analysis and evaluation of the emotional reactions of the user in the articulation of interaction with the automated system. Open education, 2: 83-85.
7. Pang, B., L. Lee and S. Vaithyanathan, 2002. Sentiment classification using machine learning techniques. In Proceeding of the conference on empirical methods in natural language processing, pp: 79-86.
8. Konig, A. and E. Brill, 2006. Reducing the human overhead in text categorization. In Proceedings of the 12th ACM SIGKDD conference on knowledge discovery and data mining, pp: 598-603.
9. Vsevolodova, M. and E. Vladimirsky, 2009. Ways of expressing spatial relations in the modern Russian language, Book House “Librokom”, Moscow, pp: 288.
10. Osipov, G., I. Tikhomirov and I. Smirnov, 2008. Relational-situational method of search and text analysis and its applications. Artificial Intelligence and Decision Making, Institute Of Systems Analysis of Russian Academy of Sciences, 2: 3-10.
11. Pham Hung, D.C. and S. Zakharov, 2008. Application of fuzzy neural network for processing temporal information in the text in Russian. In the Proceedings of the Conference, Intelligent systems. Intelligent CAD, 3: 16-22.
12. Gabelaia, D., 2005. Combining Spatial and Temporal Logics: Expressiveness vs. Complexity. Journal of artificial intelligence research, pp: 167-243.