УДК 004.912 : 004.048

O. V. Bisikalo
O. O. Maksymov
A. T. Maksymova
R. V. Slobodian

DETECTION OF THE TERRORIST THREATS IN THE TEXT MESSAGES

Vinnytsia national technical university

Анотація

Запропоновано метод визначення терористичних загроз у текстових повідомленнях, який базуються на моделі асоціативного образного мислення людини. Забезпечено можливість використання параметрів складних зв'язків між словами у реченнях текстового документу для індексування текстової інформації, а також зменшення розмірності онлайнової класифікації лінійних векторів. Розроблено програмний прототип системи моніторингу загроз, що базується на лінгвістичному пакеті DKPro Core.

Ключові слова: текстові повідомлення, терористичні загрози, складні відношення, лінійний класифікатор, машинне навчання, модель, образне асоціативне мислення.

Abstract

The method of determination of terrorist threats in text messages, which based on the model of associative imaginative thinking of a person, is proposed. It is possible to use the parameters of complex links between words in the text document sentences for indexing textual information, as well as to reduce the dimensionality of the online classification of linear vectors. The software prototype of the threat monitoring system, based on the DKPro Core linguistic package, has been develop.

Keywords: text messages, terrorist threats, complex relationships, linear classifier, machine learning, model, associative imaginative thinking.

Problem

Hidden content of terrorist threats in the text messages is usually deliberately veiled by using a neutral vocabulary; also, malicious people often resort to synonymous and metaphorical constructions. Therefore, recognition of terrorist messages by searching for keywords from a "terrorist dictionary" does not bring the desired results. Using machine-learning techniques based upon already known cases of "lexical support" for terrorist threats also does not work if the attackers have agreed on their own original slang ("Gypsy language").

Approach to solution

The proposed approach based upon the model of human figurative associative thinking. Essential natural processes of intelligence formation and development were reduced to the model of the associative network of images' life cycle. There was determined a new concept of linguistic image that represented as the subset of the set of words that are verbal signs of the image to recognize. The approach provides the definition of the components of linguistic images, as well as the strength of the links between them based upon the known methods of linguistic and statistical processing of text information. It proposes to represent the terrorist messages in the form of graphs / subgraphs of linguistic images that are invariant to verbal signs of those images. So really, we are looking not for words, but for stable networks / subnets of the links between them.

Brief description of the proposed method

The new method aimed at implementing the proposed approach through the combination of the known methods of information retrieval and machine learning. Formally, we have a set of documents $D = \{d_1, \dots, d_{|D|}\}$, information about each of them can be obtained only from the document itself without involving external sources. Also known is $C = \{c_1, \dots, c_{|C|}\}$ – a set of categories or abstract labels, each of them may

mark one of the multiple documents from D. Let it be $\Phi: D \times C \rightarrow \{0,1\}$ – unknown target function that is determined from pair $\langle d_i, c_j \rangle$ or the document d_i belongs of category c_j (True) or no (False). The problem is to construct (classification) the function Φ^k, as close as possible to Φ. The formally ranked classifier will be considered the definition of the function $CSV_i: D \rightarrow [0,1]$, that for each document d_j returns the categorization status value d_j to c_i. You can build an exact classifier this way: or immediately construct a function $CSV_i: D \rightarrow [T, F]$ or compute a similarly ranked function $CSV_i: D \rightarrow [0,1]$, and then define the threshold τ_i such that $CSV_i \geq \tau_i$ considers as T, and $CSV_i < \tau_i$ considers as F.

To solve the problem of constructing the function of the classifier of threats Φ^k, it is proposed to apply the possibilities of modern linguistic packages to determine the complex relationships between the sentences of the text document and to put the obtained numerical parameters in the basis of the procedure for indexing text information. Unlike the known methods, the proposed method, instead of words (terms), uses the coordinates of the vector of a separate text document as links between words. New method ensures numerical indicators of the completeness and accuracy of the threatening documents' identification.

Expected Results

The proposed method implements on the platform of the Apache UIMA framework and the DKPro Core linguistic package with Java / Maven / Eclipse software support. The received software prototype of the threat monitoring system builds a network of linguistic images (meaningful forms of word) based on storing complex syntactic relationships between the lemmas. Formal analysis of the obtained graph of lexical ontology allows us to determine the degree of its proximity to the classified space of similar graphs that correspond to known types of threats.

After performing the main task of the prototype, the formation of a relevant knowledge base to recognizing text-based terrorist messages based upon the new approach, it leads to development of professional software to support the threat monitoring system online.

REFERENCES

1. Бісікало О.В. Метод визначення ключових слів англомовного тексту на основі DKPro Core / О.В. Бісікало, О.В. Яхимович // Технологический аудит и резервы производства: Информационные технологии. – 2015. – Том 1, № 2(21). – С. 26–30.

2. Бісікало О.В. Визначення змістовних ознак тексту на основі аналізу зв'язків між лексичними одиницями / О.В. Бісікало, А.І. Лісовенко, О.В. Яхимович, С.С. Траченко // Вісник НТУ «ХПІ». Серія: Механіко-технологічні системи та комплекси. – Х: НТУ «ХПІ», – 2015. – № 21 (1130). – С.83–89. – Бібліогр.: 10 назв. – ISSN 2411-2798.

Максимов Олексій Олексійович — студент групи 3АКІТ-17м, кафедра комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця.

Максимова Анастасія Тарасівна — студентка групи 3АКІТ-17м, кафедра автоматики та інформаційно вимірювальної техніки, Вінницький національний технічний університет, м. Вінниця.

Слободян Роман Віталійович — студент групи 3АКІТ-17м, кафедра автоматики та інформаційно вимірювальної техніки, Вінницький національний технічний університет, м. Вінниця.

Науковий керівник: ***Бісікало Олег Володимирович*** — доктор технічних наук, професор, декан факультету Комп'ютерних систем і автоматики, Вінницький національний технічний університет, м. Вінниця.

Maksymov Oleksii O. — student of group 3ACIT-17m, Department of Computer Control System, Vinnytsia National Technical University, Vinnytsia.

Maksymova Anastasiia T. — student of group 3ACIT-17m, Department of Automation and Information-Measuring Devices, Vinnytsia National Technical University, Vinnytsia.

Slobodian Roman V. — student of group 3ACIT-17m, Department of Automation and Information-Measuring Devices, Vinnytsia National Technical University, Vinnytsia.

Supervisor: ***Bisikalo Oleg V.*** — Dr. Sc. (Eng.), Professor, Head of the Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia.