Decision-making and computational linguistic tools application for overall estimation of the level of social tension

Maksym Shchoholiev¹,*, Oleh Andriichuk², Vitaly Tsyganok² and Violeta Tretynyk¹
¹National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine
²Institute for Information Recording of National Academy of Sciences of Ukraine, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

*E-mail: shchoholiev.maksym@gmail.com

Abstract. The study of social tension, its causes and laws of variation, as well as establishing its level is crucial during crises, epidemics and wars. The purpose of this research is to develop the system for determining the level of tension in society on the basis of data from social networks using decision-making and computational linguistic tools. The impact of various factors on the growth of social tensions is determined with the help of knowledge bases built by means of decision support systems. To analyse the emotionality of comment on news publications the TF-IDF and Word2vec methods are used. Within the framework of the method of goal dynamic estimating of alternatives, the level of social tension is determined as the degree of achievement of the main goal, the ratings of news publications and the ratings of news events that contributed to the increase in social tension are calculated. The study presents the efficient method of determining the level of social tension and defining the main factors that have the greatest impact on the growth of social tension. Further studies are needed to improve the estimates of factors and the overall estimate of social tension.

1. Introduction
Social tension (ST) is one of the main characteristics that is taken into account in determining social stability. On the one hand ST is an indicator of social well-being and on the other, an indicator of information attacks.

There are different methodological approaches to determining the level of tension (in some sources mentioned as a level of deprivation, stress level) in society. Sociologists divide all methods into two groups: survey methods of sociological research and analysis of statistical material [1].

Survey methods of sociological research [2-3] provide many opportunities for different types of analysis, but research will be limited to a specific social group. The analysis of statistical material [4-8] provides an opportunity to study the dynamics of tension, and also allows to determine tension within large regions and at the national level, but such studies will be retrospective.

There is also a group of methods in which the main role is played by machine learning algorithms and sentiment analysis tools [9-11]. The use of these methods allows to encompass many social groups and determine the level of tension in real time.
For example, to determine the level of tension, the authors of the article [11] used the publications of users in social networks. For determining the attitude of the authors of publications to a particular news the support vector machine (SVM) classification algorithm was used. Besides, the TF-IDF (Term Frequency - Inverse Document Frequency) [12] method was applied for feature extraction.

ST can be forcefully formed in target groups through information operations [13-14]. Research of information operations is an urgent issue in the conditions of modern rapid development of information technologies and their comprehensive and deep penetration into all spheres of human life. Consequently, an information environment forms, which in a certain way affects society, social groups and individuals. By means of information operations, it is possible to form a certain opinion and attitude towards a target issue, topic, object in the target audience, which can even provoke some ST [15].

The purpose of this article is to describe the decision-making and computational linguistic tools, as well as the system for determining the level of ST, created by using these tools.

Rapid identification of information operations gives a great advantage to government agencies in combating them. The system presented in the article is being developed for state law enforcement and defence agencies in Ukraine as part of the NATO CyRADARS project.

Taking into account the specifics of ST, some properties innate to such areas should be noted. The measurement of ST depends on the scale of the object for which it is measured (country, region, city, district), the specificity of the mentality of the local population, the uniqueness of local socio-economic and environmental factors. Dynamism is determined by the irreversibility of the development processes of complex systems, and the ever-increasing speed of modern social and information processes. The consequence of dynamism is also the incompleteness of the description, because ST can be created by the processes that have just begun to occur, but have not yet managed to fully manifest themselves. This is especially true for operational monitoring of ST. The use of statistical data in the measurement of ST is certainly associated with vagueness, errors, ambiguities, inconsistency and inaccuracy in the collection of information. The literature describes a number of methods for measuring ST, but it is impossible to explicitly formalize ST as a function or a system of equations and, as a result, it is impossible to build its analytical model. It is difficult to fully describe the characteristics of ST quantitatively, and therefore it is irrelevant to talk about the existence of reference values for these characteristics. ST affects people who have free will. Predicting human behaviour as a component of a social system is often impossible.

Thus, ST falls under the category of ill-structured subject areas [16-18], and it is advisable to model such subject areas using expert decision support systems (DSS) [19-20].

2. Approach to the use of DSS to measure social tension

Consider the approach to the use of decision support tools for measuring ST. The proposed approach in some way overlaps with the ones, proposed for the recognition of information operations [21]. In the framework of the suggested approach, DSS tools are used to build knowledge bases, in which, through consistent decomposition, factors of ST are revealed, relevant news events and news publications are indicated. Here the results of content monitoring (particularly of social networks) and the knowledge of experts are used. An analysis of the emotionality of comments on news publications using the TF-IDF and Word2vec [22] methods determines the level of emotional tension associated with certain news events and, as a result, their contribution to the increase in ST. Subsequently, using the method of hierarchical goal dynamic estimating of alternatives [23], the level of ST is determined as the degree of achievement of the main goal, the ratings of news publications and the ratings of news events that contributed to the increase in ST are calculated.

The contents of the proposed methodology of the use of DSS tools for measuring ST are as follows:

1) A preliminary study of the object for which the ST will be measured is carried out, a group of experts is selected.
2) An expert decomposition of factors (criteria) of ST is conducted. The criteria of ST are introduced to the DSS knowledge base in the form of corresponding goals.
3) The goals corresponding to news events are entered into the DSS knowledge base and are tied to the corresponding factors of ST.
4) The projects corresponding to news publications are entered into the DSS knowledge base, and their influence on the relevant news events is established.
5) Projects can be combined into complex projects corresponding to each news event.
6) The partial coefficient of influence (PCI) of the project (news publication) is determined as the normalized value of the product of the number of times a news publication is viewed by its level of emotional tension.
7) The PCI of the goals, that correspond to news events, are determined as the normalized value of the sum of products of the number of views of related news publications by their level of emotional tension.
8) The PCI of goals, that correspond to ST criteria, are determined either by expert means [24] or by finding normalized values of the sum of products of the number of views tied to the corresponding news events of news publications by their level of emotional tension.
9) The degree of achievement of the main goal is calculated as an indicator of ST level. If necessary, the level of ST is also calculated for specific factors.
10) The ratings of the efficiency of projects and complex projects are calculated as indicators of the contribution of each of the news publications to the creation of ST.

Preceding calculations are used in the formation of DSS recommendations.

3. Application of sentiment analysis methods to determine the level of tension in society
To calculate the rating of news publications and news events, an estimate of ST is used, which is determined on the basis of analysis of comments from social networks, and on the basis of popularity of pages of news resources from which these comments are obtained.

The scheme for calculating the estimate of ST for a particular news event or news resource is presented in figure 1.

![Diagram](image)

Figure 1. The scheme of determining the level of tension in society.

Estimates of the emotionality of comments are determined on the basis of analysis of their keywords. To extract keywords, among which according to the conducted experiments the emotional vocabulary occurs most often, the TF-IDF method is used. From the extracted words a dictionary is formed, in which for each word its context vector is given. The latter is formed using the Word2vec method with the CBOW (Continuous Bag of Words) learning algorithm. The group of keyword vectors that characterizes the comment under study is input to the classifier, which determines the estimate of the emotionality of this comment. An artificial neural network is used as a classifier.

Estimates of comments are determined on the interval [0; 2], where 0 shows the highest positive emotionality, 2 - the highest negative emotionality, 0 – neutrality of a comment.

The level of nesting of comments (comments on the comments) is also taken into account. The higher the level of nesting, the less impact this comment has on the overall estimate of tension.

On the basis of obtained estimates for each publication, it is possible to build the scheme of distribution of tension. Schematically, this distribution can be represented as one of eight histograms if all the estimates are combined in three equal groups (figure 2). An indicator of the fact that publication provokes the increase in tension is the big number of comments with the estimates 1,4-2. Therefore,
further attention will be focused mainly on those publications that have a distribution of emotionality of the type b, d, f, or h.

![Figure 2](image_url)

**Figure 2.** Simplified schemes of comments distribution by emotionality.

The average number of comments with estimates 1.4-2 is calculated as a weighted arithmetic mean:

$$x_1 = \frac{\sum_{i=1}^{n_1} N_i w_i}{n_1},$$

(1)

$N_i \in \mathbb{N}$ – number of comments with a certain estimate of emotionality, $n_1 = 7$ – number of possible estimates of emotionality in the interval 1.4-2, $w_i = \frac{1}{10} + 0.6; i \in [1;7]$ – weights of estimates (shown in table 1).

**Table 1.** Coefficient distribution for the comments with emotion estimates 1.4-2.

| Estimates of comments emotionality | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|-----------------------------------|-----|-----|-----|-----|-----|-----|-----|
| Weighting coefficients            | 0.7 | 0.8 | 0.9 | 1   | 1.1 | 1.2 | 1.3 |

The overall average number of comments is calculated as the usual arithmetic mean:

$$x_2 = \frac{\sum_{i=1}^{n_2} N_i}{n_2},$$

(2)

$N_i \in \mathbb{N}$ – number of comments with a certain estimate of emotionality, $n_2 = 21$ – overall number of possible estimates of emotionality.

The coefficient showing the ratio between the total average number of comments and the average number of comments with negative emotional connotation is calculated by the formula:

$$k = \frac{x_2}{x_1}.$$  

(3)

The main focus will be on those publications for which $k \leq 1$, that is on those where the average number of comments with high negative emotionality exceeds the total average number of comments. In addition, the values of $n_{neg}$ and $n_{pos}$ are calculated which show the number of publications under study for which $k \leq 1$ and $k > 1$ respectively.

To take into account the popularity of new resources the value of $k^*$ for each $k$ is calculated by the formula:

$$k^* = 1 - a(1 - k),$$

(4)
a – number of subscribers, normalized by all news resources; \( a \in [0;1] \). The graph of values \( k^* \) for different \( a \) and \( k \) on the interval \([0;1]\) is shown in figure 3.

\[ X = \sqrt[n]{\prod_{j=1}^{n} k_j^*}, \] (5)

\( X \in [0,1] \). Depending on the problem formula (5) can be used for calculating negative tension either for a certain news resource or for a certain news event.

4. The practical example of calculating social tension by means of DSS

Consider the application of the above methodology for measuring ST on the example of the State of Ukraine.

| #  | Goal formulation                                                                 |
|----|---------------------------------------------------------------------------------|
| 0  | Social Tension in Ukraine                                                       |
| 1  | Criminogenic situation                                                          |
| 2  | Inability to get a decent job                                                    |
| 3  | Health issues                                                                   |
| 4  | Issues in the field of housing and utilities                                   |
| 5  | Threat of military conflict                                                     |
| 6  | Environmental problems                                                          |
| 7  | The Security Service of Ukraine exposed Major General Shaitanov, who worked for the Federal Security Service of the Russian Federation |
| 8  | Fires in the Chornobyl Exclusion Zone                                           |
| 9  | Kyiv has returned to the ranking of cities with the most polluted air          |
| 10 | Tariffs, subscription fee for mobile networks                                   |
| 11 | Imposition of quarantine                                                        |
| 12 | Closure of public transportation due to quarantine                              |
| 13 | Shortage of masks in pharmacies                                                 |
| 14 | Health care reform                                                              |
| 15 | Forest fires in the Zhytomyr region                                             |
| 16 | Product prices are rising                                                        |
| 17 | The growth of the dollar                                                        |
| 18 | Interview of Hordon with Girkin                                                 |
| 19 | Exchange of prisoners. 20 hostages were released from the occupied territories  |
| 20 | MH17 case                                                                       |
| 21 | House collapsed in Odessa                                                       |
| 22 | A bridge fell into the river in Dnipropetrovsk region                            |
| 23 | Zhytomyr. Shooting by the lake.                                                 |
| 24 | Robbery in Kharkiv. Unknown people blew up an ATM and took money.               |
| 25-68 | News publications                                                                |

Figure 3. The graph of values \( k^* \) for different \( a \) and \( k \) on the interval \([0,1]\).
Figure 4 shows an example of a knowledge base of ST in Ukraine built by means of DSS for the period 01.01.2020 - 23.05.2020.

Table 2 contains the list of formulations of all goals and projects of the knowledge base. As a result of calculating the level of social tension by means of DSS as the degree of achievement of the main goal for the period 01.01.2020 - 23.05.2020, a value of 0.42278 was obtained.

A fragment (top 5) of the result of calculating the rating of news publications as the efficiency of projects using DSS is shown in table 3.

Table 3. Fragment (top 5) of rating of news publications.

| Efficiency   | News publication related to the event                  |
|--------------|-------------------------------------------------------|
| 0.051279277  | Imposition of quarantine                              |
| 0.034086009  | Closure of public transportation due to quarantine    |
| 0.031551813  | Interview of Hordon with Girkin                      |
| 0.020846635  | Zhytomyr. Shooting by the lake.                       |
| 0.019569476  | Shortage of masks in pharmacies                       |

A fragment (top 5) of the result of calculating the rating of news events as the efficiency of the complex projects using DSS is shown in table 4. Each complex project consists of simple projects that correspond to some news publications.

Table 4. Fragment (top 5) of rating of news events.

| Efficiency   | News event                                              |
|--------------|---------------------------------------------------------|
| 0.143001     | Imposition of quarantine                                |
| 0.10456      | Closure of public transportation due to quarantine      |
| 0.096787     | Interview of Hordon with Girkin                        |
| 0.069838     | Fires in the Chernobyl Exclusion Zone                   |
| 0.057445     | Shortage of masks in pharmacies                         |

5. The concept of creating a system for determining social tension

Figure 5 shows the Use Case diagram of the concept of a system for determining ST, which is an implementation of the methods proposed above and system integration of the following software tools: system for distributed collection and processing of expert information (SDCPEI), expert estimation system (EES), and DSS.

Actors:
An Expert is a specialist who is invited or hired for a monetary reward to provide a qualified opinion, assessment or judgment on certain issues.

A Knowledge Engineer is a specialist who forms a knowledge base and calculates recommendations based on it using the DSS toolkit.

SDCPEI is a system for remote group work of experts in a global network.

EES is a software complex which flexibly adapts to the level of competence of experts and allows to fully acquire knowledge from them without distortion.

DSS is a system that calculates and issues recommendations based on the knowledge (both objective and expert) introduced in its knowledge base.

Use Cases:

Building the Knowledge Base. This process, like all the use cases described below, is initiated by the knowledge engineer. As part of this process, the DSS tools are used to construct a problem-oriented domain model based on knowledge obtained from experts and the results of content monitoring of the global network.

Group Expert Decomposition. Experts participate in a group decomposition of ST into factors. According to the object of ST, the knowledge engineer forms a group of experts competent in the issues to be considered. The process of group decomposition occurs during a phased dialogue with experts from the group by successive division of ST into components. At each stage of decomposition, an expert from the group is invited to formulate many factors of ST (choose factors among existing ones or introduce new ones). Further, in a similar way, decomposition of factors of ST into components can be done. The decomposition process is completed when specific news events act as constituents, on the basis of which relevant requests can be made for content monitoring of news publications.

Group Expert Estimation. The process is initiated when the group decomposition is over. During this process, for each group of factors of ST, the nature of their impacts and the magnitude of their degrees of influence are determined.

Entering Objective Information to the Knowledge Base. During this process, information that can be obtained by measuring certain values can be entered into the knowledge base.

Calculation of the Overall Estimate of the Level of Tension in Society. During this process, on the basis of the constructed knowledge base, the degree of achievement of the goals of the hierarchy is calculated. The degree of achievement of the main goal corresponds to the overall estimate of the level of tension in society. The degrees of achievement of the subgoals correspond to ST by each of the factors in particular. The time dynamic is also taken into account.

Calculation of Rating of the News Publications. In the course of this process, on the basis of the constructed knowledge base, the relative efficiency of news publications is calculated, that is, the “relative contribution” of each of these publications to the achievement of the main goal which is the creation of ST. An integral estimate of the efficiency of news publications is also calculated. The time dynamic is taken into account.

Calculation of Rating of the News Events. The course of this process is similar to the use case “Calculation of Rating of the News Publications”. A complex project is taken as a news event, where corresponding news publications are taken as simple projects included in it.

Searching for Recent News Events and Publications. During this process, on the basis of the constructed knowledge base, content monitoring and expert knowledge, a search for new news events and publications is carried out, and the knowledge base is supplemented.

Sentiment Analysis of Comments of the News Publications. During this process, the analysis of the emotionality of comments on news publications is carried out using the methods TF-IDF and Word2vec, the level of tension associated with certain news events is determined, and, as a result, their contribution to the factors of ST is determined as well.

6. Conclusions
Today, information operations conducted both by certain political forces within the country and by some countries against others are becoming commonplace in our lives. The main indicator of an information attack is the level of ST. In addition, the level of tension can determine the level of social welfare.

In the paper was described the system for determining the level of ST using tools of computational linguistics and the DSS knowledge base. A methodology was proposed for using the DSS toolkit to build a knowledge base in which the main factors of social tension were determined by successive decomposition. In this case, the results of content monitoring and expert assessments were used. During the study of the emotionality of comments, sentiment analysis methods were applied, in particular the TF-IDF and Word2vec methods. The calculation of social tension in Ukraine was carried out in the period 01.01.2020 - 23.05.2020, as a result of which were identified news publications and news events that showed the greatest efficiency in increasing the level of tension. The concept of the system was presented in the form of a diagram depicting the relations of its use cases and actors. Their main functions were also described.

Further research will focus on improving the accuracy of determining the influence of different factors on the level of tension, as well as improving the accuracy of determining the estimate of the overall level of ST.

Acknowledgments
This study is funded by the NATO SPS Project CyRADARS (Cyber Rapid Analysis for Defense Awareness of Real-time Situation), Project SPS G5286.

References
[1] U. V. Kaira (2017). Methodology for the study of social tension [Metodika issledovaniya sotsialnoy napryazhennosti], Bulletin of Tula State University, Humanitarian sciences, Vol. 3, pp. 107-118.
[2] G. Artemov, A. Aleinikov, D. Abgadzhava, A. Pinkevich, A. Abalian (2017). Social Tension: the Possibility of Conflict Diagnosis (on the Example of St. Petersburg), Economics and Sociology, Vol. 10, No. 1, pp. 192-208. DOI: 10.14254/2071-789X.2017/10-1/14
[3] G. Baranova, V. Frolov, A. Kondrashyn (2011). Features of social tension in the regions of Russia [Osobennosti sotsialnoy napryazhennosti v regionah Rossii], Social researches, № 6, pp. 48-55.
[4] V. A. Marenko, O. Luchko (2014). Cognitive Modelling Application for Social Tension Study, Mathematical Structures and Modeling, N. 4(32), pp. 116-127.
[5] O. Andrenko and S. Mordovtsev (2015). Integrated assessment of the social tension in the regions [Intehralna otsinka sotsialnoi naprakennosti rehioniv], Socio-Economic Problems and the State [online], 13 (2), pp. 161-168.
[6] O. Andrenko (2013). Social tensions as a type of threat to the socio-economic security of the region [Sotsialna napratenist yak vyd zahrozy sotsialno-ekonomicnhii bezpantsi rehionu], Business-inform, № 9, pp. 119-123.
[7] P. Akinina and S. Ryazantseva (Ed.). 2002. Diagnostics of social tension in society: a regional aspect [Diagnostika sotsialnyy napryazhennosti v obschestve: regionalnyiy aspekt], Servisskhola, Stavropol, 240 p.
[8] M. Galiullin (2015). Using social networks to measure public concern about life safety issues [Ispolzovanie sotsialnykh setey dlya otsenki urovnya obespookennosti naseleniya voprosam bezopasnosti zhiznedeyatelnosti], Vol. 12, 2015, No. 2 (44), pp. 70-73.
[9] P. Burnap, O.F. Rana, N.J. Avis, M.L. Williams, W. Housley, A.M. Edwards, J. Morgan, L. Sloan (2015). Detecting tension in online communities with computational Twitter analysis, Technological Forecasting and Social Change, Vol. 95, pp. 96–108.
[10] G. Osipov, O. Vybornova, I. Smirnov (2011). Social tension detection and intention recognition using natural language semantic analysis, European Intelligence and Security Informatics
Conference (EISIC 2011), art. no. 6061249, pp. 277–281.

[11] D. Donchenko, N. Ovchar, N. Sadovnikova, D. Parygin, O. Shabalina, D. Ather (2017). Analysis of Comments of Users of Social Networks to Assess the Level of Social Tension, 6th International Young Scientists Conference in HPC and Simulation (YSC 2017), Kotka, Finland, pp. 359-367.

[12] D. C. Manning, P. Raghavan, H. Schütze (2009). An Introduction to Information Retrieval, 544 p. [https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf]

[13] Information operations roadmap. 2003. Washington, D.C., Department of Defense, 74 p.

[14] Military Information Support Operations. Joint Publication 3-13.2. 07 January 2010. Incorporating Change 1, 20 December 2011. 10 p. [http://fas.org/irp/doddir/dod/jp3-13-2.pdf]

[15] M. Shchoholiev and V. Tretnyuk (2019). The System of Operative Determination of the Level of Tension in Society Based on Data from Social Networks, Information & Security: An International Journal, Sofia, Bulgaria, vol. 43, no. 3, pp. 375-382. DOI: 10.11610/isij.4328.

[16] A. N. Averkin, O. P. Kuznetsov, A. A. Kulinich, N. V. Titova (2006). Decision-making support in weakly structured subject domains: Analysis of situations and evaluation of alternatives, Journal of Computer and Systems Sciences International, vol. 45, no. 3, pp. 469–479.

[17] V. Tsyganok, S. Kadenko, O. Andriychuk, P. Roik (2017). Usage of multicriteria decision-making support arsenal for strategic planning in environmental protection sphere, Journal of Multi-Criteria Decision Analysis, vol. 24, issue 5-6, pp. 227-238. DOI: 10.1002/mcda.1616.

[18] S. V. Kadenko (2016). Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area, CEUR Workshop Proceedings, Selected Papers of the XVI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2016), Kyiv, Ukraine, pp. 8-14. [http://ceur-ws.org/Vol-1813/paper2.pdf]

[19] T. L. Saaty (2000). Fundamentals of decision making and priority theory with the analytic hierarchy process, RWS Publications, Pittsburgh, USA, 477 p.

[20] D. T. Lee (1988). Expert Decision-support Systems for Decision-making, Journal of Information Technology (3:2), pp. 85-94.

[21] A. Dodonov, D. Lande, V. Tsyganok, O. Andriichuk, S. Kadenko, A. Graivoronskaya (2019). Information Operations Recognition. From Nonlinear Analysis to Decision-Making, Lambert Academic Publishing, 275 p.

[22] T. Mikolov, Q. V. Le, I. Sutskever (2013). Exploiting Similarities among Languages for Machine Translation, arXiv: Computation and Language, 10 p. [https://arxiv.org/pdf/1309.4168.pdf]

[23] V. G. Totsenko (2001). One Approach to the Decision Making Support in R&D Planning. Part 2. The Method of Goal Dynamic Estimating of Alternatives, Journal of Automation and Information Sciences, vol. 33, issue 4, pp. 82-90.

[24] V. G. Totsenko, V. V. Tsyganok (1999). Method of paired comparisons using feedback with expert. Journal of Automation and Information Sciences, 13 (7-9), pp. 86-96.