Application of Associations to Assess Similarity in Situations Prior to Armed Conflict

Ahto Kuuseok

1 Introduction

Very often decision-makers while planning for actions in critical situations that are happening, for example, in process of business process management or in military mission control are relaying on experience.

Therefore, when it comes to a new situation, it is worth looking at whether similar situations have occurred in the past, how were decisions made and how the situations evolved.

This chapter focuses on situation assessment and action control in military conflict situation assessment on the eve of military invasions. Unfortunately, the material available is scarce, problematic, and often fictional. Moreover, the volume of material available is too small to employ suitable methods for analyzing numerous examples. Therefore, tools must be found and implemented that allow at least something to be evaluated in a clearly justified way, and at least some explanation of how and on what basis it was decided. In previous works tools have been tried to use applying a descriptive similarity coefficient for situations taken in pairs. Unfortunately, it became clear that similarity estimates that emerged when comparing pairs of descriptions available for study turned out to be somehow oddly variable, with surprisingly little value.

I also got some inspiration from news article, which described situation of prediction. With the help of Artificial Intelligence, computer systems were able to predict a takeover of large Biomedical Corporation 5 days before that fact officially announced [11]. Although the methodology is not naturally the same or even close, it is tempting to go further with the idea and try to combine solutions of artificial

A. Kuuseok (✉)
Estonian Police and Border Guard Board, Estonian Business School, Tallinn, Estonia
e-mail: ahto.kuuseok@eesti.ee

H. R. Arabnia et al. (eds.), *Advances in Artificial Intelligence and Applied Cognitive Computing*, Transactions on Computational Science and Computational Intelligence, https://doi.org/10.1007/978-3-030-70296-0_46

intelligent (e.g., neural networks) and method of evaluation of descriptive similarity, to create some kind of predictions and compare them with our results. Right now, so fare it is purely “manual paper-based process” and it takes a lot of work to get results. However, it is reasonable to turn into a dialog system, where human and machine act in collaboration.

One of the inherent features of our approach is the fact that we need to evaluate similarity based on a very modest amount of information (human-accessible). Therefore, methods suitable for the analysis of voluminous data were initially excluded.

The chapter is structured as follows. Section 2 describes eves of military conflicts, which we observed, and some results we came to. Section 3 refers to highlights previous works and authors that I have relied on in this work. Section 4 explains and reveals the methodology I have used in this work. Section 5 presents some conclusions and brief description of the results and problems to be addressed for future work.

2 Military Domain

Some time ago, at the beginning of this research topic, I started with observations of numerical evaluations of situations and developments. We elaborated similarities from structural and descriptive aspects. The basis of the structural similarity is homomorphism of algebraic systems. Under observation was mainly of descriptive similarities – claims in descriptions of situations, expressible with formulas of calculations of the predictions.

The situations we observed were prior military conflicts, since Second World War II up to these days. The conflicts we chose to investigate were somehow related to the region of interest or related to the subject of the country under study (Estonia). In addition, we concentrated to the cases where one state (or group of states) were aggressor and attacked another, just a matter to make procedure of picking situations more clearly [4].

In later work, I did formalize the procedure of mining relevant examples of armed conflicts [3].

However, the results of pairwise evaluated situations were surprisingly scattering, and with surprisingly low numerical similarity indices [5]. Some of the observed pairs we expected to have very high indexes turned out low, and there were examples to the contrary. Some examples. “Gulf War” (02.08.1990–28.02.1001) was a conflict between coalition forces and Iraq; and Suez Crisis (10.1956–03.1957) armed conflict between Coalition and Egypt. Looking at the history of the conflict, the eve of the outbreak of the conflict, and even the underlying causes, the similarity should be quite high. In fact, the similarity index calculated by the above method was only 0.111. In addition, contrary to expectations, in some cases the calculated index turned out to be higher than expected. For example, the story of the Vietnam War (between coalition leaded by the US and coalition leaded by Soviet Union

1964–1972) and the outbreak of the Russo-Georgian War (between Russia and Georgia – August 7–15, 2008) and the eve of it should not have had much in common. However, the calculated similarity index came out 0.363.

In addition, it emerged that it was not possible to separate the intersection part in the descriptions of the situations available for examination. This made it necessary to find a more appropriate method of dealing with sets of claims from descriptions that may not all have an intersection part.

3 Related Works

Situations and developments dealt with as "*sets of statements*" that characterize them. These are finite number of finite sets that could estimate by the descriptive similarity coefficient of the descriptions of the situations. Defined by P. Lorents as following: $Sim_{LT}(A,B) = E(equ_T(A,B)):[E(A) + E(B) - E(equ_T(A,B))]$, where

- **E(H)** is the number of elements of a set H, **T** this is the way of equalization,
- $\mathbf{equ_T(A, B)}$ is such a set, where $x \in A$ or $y \in B$ do belong to i*n case, if* x and y, **are *equated*** [4]. Remark: basis (and method) of the equalization will be selected and fixed by the implementer of the current method (expert, analyst, etc.). For example, based on logical equivalence, if there are suitable tools to "rewrite" the respective statements into formulas of the predicate calculus [2, 9]. On the other hand, for example, based on the ability to define statements with the same meaning by fixing the corresponding definitions in the appropriate table [4, 6, 7].

It is important to note that the same and identified are essentially different concepts. The same things may turn out to be the identified; the things that identified may not be the same.

The approach described above has already been implemented in areas other than security. For example, to investigate the circumstances of IT project failure, to assess the actual similarity of certain Federal and Estonian legislation [4]. It has also been used for comparative study of the public transport situation in small towns [6, 7].

4 Proposed Methodology

Treating situation descriptions as a set of relevant statements. Definition of similarity of statements. For example, in one case based on logical equivalence of the predicate calculus formulas which express those statements [12, 13]. In the second case, for example, based on the common meaning of the statements. Association of certain things, including texts with meanings, is essentially the formation of relevant knowledge [14].

If there are two comparable descriptions, then we can use the descriptive similarity coefficient to assess the similarity of the situations (equ.) [4]. However, if there are more than two comparable descriptions, then the association similarity coefficient (cos) should be used (see Def. 4).

To use multiple sets of similarity coefficients, it is necessary to define associations of sets (situation descriptions). As well as interset contacts, connectors, associators, contact networks, etc. (see Def. 1, 2, 3).

The results of pair-wise observation of situation descriptions immediately preceding armed conflicts were not suitable to provide practically usable assessments of the similarity of situation descriptions. This method explained more closely in previous research work, see citation [5].

Consequently, a number of new (mathematical) concepts and appropriate procedures had to be defined so that even in the absence of a common element, it would be possible to quantify the similarity of some final sets, including sets of statements derived from situation descriptions. Such terms were, for example, associations, associators, connectors, contact network, etc. of finite sets, as well as overhead networks of association sets, and finally the coefficient of similarity of association sets and the procedure for calculating it [8].

5 Definitions and Procedures

It is important to note it again that same and identified are essentially different concepts. Therefore, the "intersection-like set" is not automatically the same what could be considered as set–theoretical intersection.

The proposed method of assessing similarity between situations happening in armed conflicts is based on set–theoretical mathematical apparatus introduced in [4, 5, 8].

In the following, let us state that for sets, we speak of finite sets. We use the symbol E(H) to denote the final number of finite set H elements. A set whose elements are other sets we call a class.

One of the most important classes in this work is the sets of the descriptions of situations. To identify sets that belong to a class K, we use a fixed method of identification, such as representation of pairs of objects that been identified by an appropriate table, see reference [5].

Let it be a sign of this equalization **ε**. During the identification process, pairs of elements formed. We know from set theory that a set of pairs of elements of any set forms some kind of binary relationship between the elements of these sets [10].

The relation created by the identification method (proposed by P. Lorents) **ε** we denote by a symbol $\equiv_{\boldsymbol{\varepsilon}}$. The relations under consideration could possess different algebraic properties, such as symmetry, different forms of transitivity (e.g., weak transitivity [15]) etc. Inter alia, those properties often used as constraints for building reasoning mechanisms over situations.

Definition 1 Let A and B be two sets in class K. We call ***contact*** between sets A and B a pair {α,β}, where α∈A, β∈B, and $\alpha \equiv_{\varepsilon} \beta$. Elements α and β will be called

connectors. All connectors from the set A are constitute a ***binder of A,*** and denoted as **binA**. We call set **H** ***isolated*** in class **K**, unless **H** is directly associated with any other class of **K**.

Definition 2 The two sets A and B of class K are called associate sets by the method of equalization ε in short associate sets. If (I) these sets are directly associated, that is, there is at least one contact between these sets (II) there is a set C, which also comes from class K, where sets A and C are associated and sets C and B are associated [8].

Definition 3 As the ***associator*** of the class K association G we call the binH'∪binH"∪ . . . ∪binH"', that is, the set whose elements are all the connectors that come from the sets H ', H' ', ..., H' "forming the given association. We designate this associator with the symbol assG. In the ***network G*** of association G, we refer to the number of contacts that occur between sets in that association. We designate this network with the symbol netG.

The **associator** of some sets is what could be considered as an expression of the intersection, even if, if there is no "pure intersection" by set theory.

Definition 4 The ***similarity coefficient of the sets of the association G***, which we denote by cosG, is called the ratio E(netG):E(∪G), where E(netG) is the number of all contacts in the network netG, and E(∪G) is the number of all elements derived from union ∪G. We denote the similarity coefficient of association G by the symbol **cosG**. While doing so, cosG = E(netG):E(∪G).

Below we will deal with sets of class **K** elements that contain descriptions of 16 situations immediately preceding the military conflict. Elements of these descriptions, in turn, are relevant statements, which are identified/unidentified by using an appropriate table that contains pairs of identifiable statements. These 16 conflicts selected by appropriate procedure [3]. This procedure is based on a certain direct or indirect connection with the security of the Republic of Estonia, the strict formulation of which (i.e., the procedure) follows certain clearly formulated criteria.

Based on the fixed form of identification in the form of a table and the relationship of the identification equalized with it, we formed associations in class K. In addition, exactly one isolated set revealed. We then found out the number of contacts that make up the overhead network of contacts, as well as the number of all the elements in each set of associations. It is necessary to assess the overall similarity of situations that have arisen prior to the military conflicts addressed.

Based on clearly and rigorously formulated terms and the formulas needed to implement them, inequalities and equations, it is possible to explain what could be similar in descriptions of many different situations. Defining, handling, and applying in some sort of common or somehow similar set to several sets proved to be problematic. Especially when it turned out those descriptions of situations from a larger class of situations did not contain identifiable statements that could be found in all relevant (i.e., situations in a given class) descriptions. What was particularly irritating was the fact that, although all of these situations were followed exclusively by armed conflict, comparisons between these (i.e., a small numerical value of the similarity coefficient). However, again, it was clear that some claims or their equivalents were present, sometimes in groups, in descriptions of many and many different situations. It took quite a long, laborious journey to notice such strange "common-like" collections, and to define them precisely. In doing so, one had to create and integrate, on the one hand, carefully collected, structured, and analyzed empirical material, and, on the other, mathematical constructions that often did not want to fit together. The result, however, was eventually to find the relevant concepts, tools, and obtain numerical estimates. In doing so, the results, including the concepts of association, associators of this association, and the coefficient of similarity of the catenary and finally the association, helped. Their implementation helped to produce some noticeable and partly surprising results. Including pleasant surprises. For example:

In previous work, Kuuseok [3] has developed procedure of mining proper and suitable cases as examples. Because there is an unimaginable amount of data on the topic of interest to us in publicly available materials, it needs to be prefiltered. The purpose of the filtering is to exclude from consideration such cases where the historical, military–political, and other such aspects would not fit in the case of the country under investigation, in this case Estonia. Geopolitical, military–political, historical, and other aspects have been considered. The analysis of the original material revealed the following dimensions and boundaries:

- Dimensions of Time: II MS to Contemporary
- The geopolitical dimension: On the one hand, delimited by the geopolitical characteristics of its immediate neighborhood and allied countries
- The military–political dimension is limited to potentially hostile neutral, friendly and allied countries

It is important that at least one condition is met, but there may be a number of conditions, all of them. The application of the appropriate procedure ultimately left the sieve under scrutiny, leaving 15 military conflicts in which these conditions were met. We can take the example of the Russo-Georgian War in 2008, where for Estonia have met several conditions – a close neighbor (Russia) and military–political cooperation (Georgia).

Their implementation helped to produce some noticeable and partly surprising results. Including pleasant surprises. For example:

- It turned out that if we allowed only two sets (D′ and D′′) of G to be considered as association G, the cosG value of the coefficient would be equal to the corresponding descriptive (discussed and used in previous works) similarity coefficient to Sim (D′, D′′).
- The sets of numbers in the figure provide a description of the situations. In sets (Fig. 1), numbers represents the statements. The numbers in the colored squares represent the connectors. The arcs between the connectors indicate the contacts. If different sets have squares of the same color, then this means that the corresponding statements represented by the numbers in the squares

equalized to each other. All statements from a description of a situation, with corresponding numbers in colored squares, form the binder of the set of the statements describing that situation. All colored squares in the figure represent the associator of the class under observation. However, all arcs represent the corresponding contact network.

- Of the observed armed conflict descriptions, one of them remained an isolated set. There was 145 statements from the 15 cases observed. During the identification process, 28 identifiable statements constituting contacts emerged, while the isolated set contained seven statements, which not identified in the case descriptions of any of the other claims. Unexpectedly, allegations emerged that were not technically much related to the military invasion of one country by another. An example that has repeatedly been stated is: "high international condemnation but no real action."
- And 83 contacts emerged between the 28 statements identified in the body of allegations that formed the association. Interestingly, when using the terminology of graph theory [1], the degree of vertices of associated statements as vertices of a graph varied up to seven (!) times, ranging from 3 to 20.
- Unexpectedly, the assertion "prompted international condemnation, but nothing realistic" came up frequently. Certainly different experts would have found these descriptions of correlated relationships, but it was not primary in this study.
- If the pairwise comparison turned out to have an average of 0.28 similarity ratings for claims derived from descriptions of military conflict eve, then the value of the similarity coefficient used in this work is considerably higher: 0.601 (see Fig. 1). This confirmed the intuitively perceived situation: When in a given set of situations each of which grew into a real armed conflict, there had to be, and as it turned out to be, more similarity than it seemed when comparing individual situations to only two.

6 Conclusions and Future Works

In this work, we studied possibilities to assess numerical similarity of several descriptions of situations immediately preceding the outbreak of a military conflict. The descriptions of the situations considered as sets of relevant statements. The central problem turned out to be the absent of an intersection (in set theoretical meaning) of the sets mentioned above. A mentioned intersection of descriptions would perhaps have been a clearer way of highlighting the more general characteristics of the outbreak of armed conflict. It was necessary to compensate for this lack of intersection.

We did it with a number of new concepts and proper definitions, and procedures for relevant assessments. For example, contacts between sets, associations of the sets, contact networks, coefficients of similarities of the sets from associations.

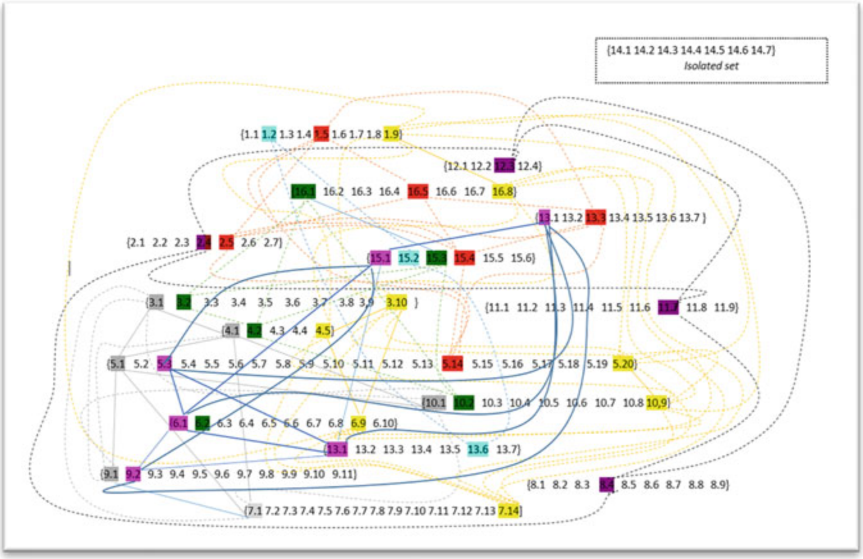

Fig. 1 Isolated Set and Similarities

However, there are already more problems that need to be addressed. For example:

- How does the change in the composition of the association (e.g., the addition of new pools or the removal of existing pools) be reflected in the change in the value of the coefficient of similarity of the association sets?
- How to interpret the changes just mentioned?
- What is the relation between the values of the similarity coefficient of the association sets to other conceivable numerical values, which, to one degree or another, arise from the treatment of associations (such as the mean of each descriptive similarity coefficient obtained by pairwise comparisons of sets)?
- What could be the interpretations of the aforementioned relationships?
- How interpret the cooperation dialog system, where human and machine act in collaboration.

One key issue here is the creation of a system based on the decisions of a particular expert and learning how to assert claims from different sources as a specific expert would do. What is important is that the system would allow equalize claims from different sources as expert would do. One approach would be to transform step-by-step the statements into suitable predicate calculus formulas. Basically in the same way as it was implemented in the Matsak 2010 prototype of the DST system [13]. Basically in the same way as it was implemented in the DST system prototype created by Matsak 2010 [13]. However, it must be acknowledged that neither DST nor other systems of this kind can in principle be so-called fully automatic. Unfortunately, this is due to algorithm–theoretical considerations. As mentioned above, it should be a self-learning system whose "mentor" is a particular expert, or team of experts, using that system. In a sense, this is the most genuine example of human–computer interaction where neither human nor computer is fully autonomous. People because data volumes and processing speeds are not feasible. Computers, however, because equalization of statements is not algorithmic.

References

1. F. Harary, Graph Theory (Published 2018 by CRC Press Taylor & Francis Group. First published 1969 by Westview Press, 2018), New York, USA
2. S.C. Kleene, Mathematical Logic (Dover Publication INC; Originally published in 1967 by John Wiley & Sons INC, 2002), Massachusetts, USA
3. A. Kuuseok, Procedure of mining relevant examples of armed conflicts to define plausibility based on numerical assessment of similarity of situations and developments. (IHSI International 2020, International Human Systems 2020) Published. Modena, Italy 19–21 February 2020, Springer (2020)
4. P. Lorents, E. Matsak, A. Kuuseok, D. Harik, Assessing the similarity of situations and developments by using metrics. 15 pages. Intelligent Decision Technologies, (KES-IDT-17), Springer, Vilamoura, Algarve, Portugal, 21–23 June 2017
5. P. Lorents, A. Kuuseok, E. Lorents, Applying systems' similarities to assess the plausibility of armed conflicts. 10 pages. Smart information and communication technologies (Smart-ICT-19), Springer, Saidia, Morocco, 26–29 September 2019.
6. P. Lorents, M. Averkina, *Application the Method of Similarity for the Decision- Making Process. Smart ICT* (Springer, Saidia, 2019)
7. P. Lorents, M. Averkina, Some mathematical and practical aspects of decision making based on similarity. (168–179), 21-st HCI international conferences, HCII 2019, Orlando, FL, USA. Springer, July 26–31, 2019 Proceedings
8. P. Lorents, Comparison of similarity of situations in associations. ICAI 2020 (2020)
9. E. Matsak, Dialogue system for extracting logic constructions in natural language texts. IC-AI 2005: 791–797 (2005)
10. M. Potter, S. Theory, its Philosophy (Oxford University Press, Oxford. UK (2004)
11. R. Aliya, R. Wigglesworth, When Silicon Valley Came to Wall Street. Financial Times, October 28, 2017. https://www.ft.com/content/ba5dc7ca-b3ef-11e7-aa26-bb002965bce8
12. E. Matsak, Dialogue system for extracting logic constructions in natural language texts. IC-AI 2005: 791–797 (2005)
13. E. Matsak, *Discovering Logical Constructs from Estonian Children Language* (Lambert Academic Publishing, Germany, 2010)
14. Lorents P, Knowledge and information. Proceedings of the international conference on artificial intelligence, IC- AI 2010. Las Vegas. CSREA Press (2010)
15. L. Esakia, Эсакиа Л. Слабая транзитивность - реституция. // Логические исследования / Logical Investigations. 2001. T. 8. C. 244–255 (2001)