

# FedCSIS\_2023\_paper\_2579.pdf

# Searching for Similarities among Sets of Experience (SOE) Knowledge Representation

3

Edward Szczerbicki  
Gdansk University of Technology  
ul. Narutowicza 11, 80233 Gdansk,  
Poland  
Email: esz@zie.pg.gda.pl

Cesar Sanin

Australian Institute of Higher  
Education, 545 Kent St, Sydney,  
NSW 2000, Australia  
Email: c.sanin@aih.nsw.edu.au

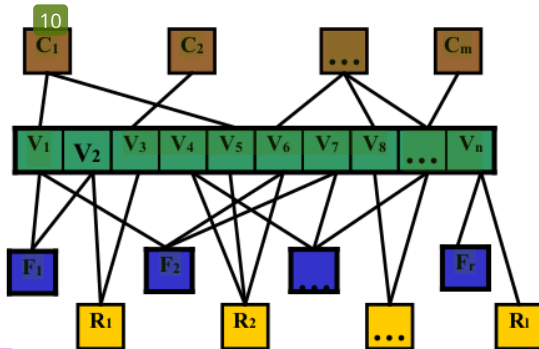
□

6

**Abstract**— In searching for patterns, it is normally not enough to consider only equality or inequality of data. Instead we need to consider how similar or different two data objects are. This is why similarity between data objects is a central concept in Data Mining (DM) and Knowledge Discovery (KD). We propose in this paper a Knowledge Representation (KR) approach in which experiential knowledge is represented by Set of Experience (SOE), and which storage, pattern discovery, and usage is managed by the concept of heterogeneous similarity metric.

## I. INTRODUCTION

SET of experience knowledge structure (SOE) has been developed to represent and store formal decision events in an explicit way [1, 2]. In this concept four basic components define decision-making events: variables, functions, constraints, and rules. They are stored in a combined dynamic configuration that embraces set of experience knowledge structure (see Fig. 1).



2

Fig 1. Set of Experience graphical interpretation: SOE is combination of four components that characterize decision making actions (variables, functions, constraints, and rules) and it includes a series of mathematical concepts (a logical component), together with a set of rules (a ruled based component) and is built upon a specific event of decision-making (a frame component).

9

Variables usually involve representing knowledge using an attribute-value language (i.e. by a vector of variables and values) [3]. Functions describe associations between dependent variable and a set of input variables [4]. A constraint, as the third component of set of experience, is a limitation of possibilities, a restriction of the feasible solutions in a decision problem, and a factor that limits the performance of a system with respect to its goals [5]. Finally, rules are suitable for representing inferences, or for associating actions with conditions under which the actions should be performed. Rules, the fourth component of set of experience, are another form of expressing relationships among variables. They are conditional relationships that operate in the universe of variables. Rules are relationships between a condition and a consequence connected by the statements IF-THEN-ELSE.

Altogether, SOE are pieces of experiential knowledge that, when collected in a structured knowledge base, can be used to search for valuable patterns augmenting future decision making.

## II. MAKING SOE TRANSPORTABLE AND UNDERSTANDABLE

Set of Experience was implemented in XML (eXtensible Markup Language) to make it transportable. This implementation provides a generic mechanism for expressing machine-readable semantics of data, as well as the possibility of presenting simple and complex structures of any kind of information or knowledge in a standardized syntax.

For illustrative purposes, an example of a head of set of experience in XML configuration can be as follows:

8

```
<?xml version="1.0" encoding="UTF-8"
standalone="no" ?>
<!-- Set of Experience Knowledge Structure -->
- <set_of_experience xmlns:
xsi="http://www.w3.org/2001/XMLSchema-
instance"
```

□ This work was not supported by any organization

```

24 noNamespaceSchemaLocation="set of
experience model.xsd">
<date>2020-11-11</date>
<hour>14:10:00</hour>
- <cate 201>
  <!-- Category encloses this set of experience into
a determined chromosome of the company -->
  <area>Human Resources</area>
  <subarea>Salary Office</subarea>
  <subject>Payment Level</subject>
  <subject>Worker's Morale</subject>
</category>

```

In practice, any semantic experience is exported into XML with the support of dedicated SOE software representation illustrated in Fig. 2. This medium maintains capture, storage, enhancement, and reuse of experience generated by formal decision events. It can be applied as a means to analyse, query, consolidate and manage semantic experience acquired by the means of SOE. Additionally, the software allows for SOE representation not only in XML, but as well in Java and OWL (Ontology Web Language) for further enhancement of transporting and sharing knowledge between various computer platforms.

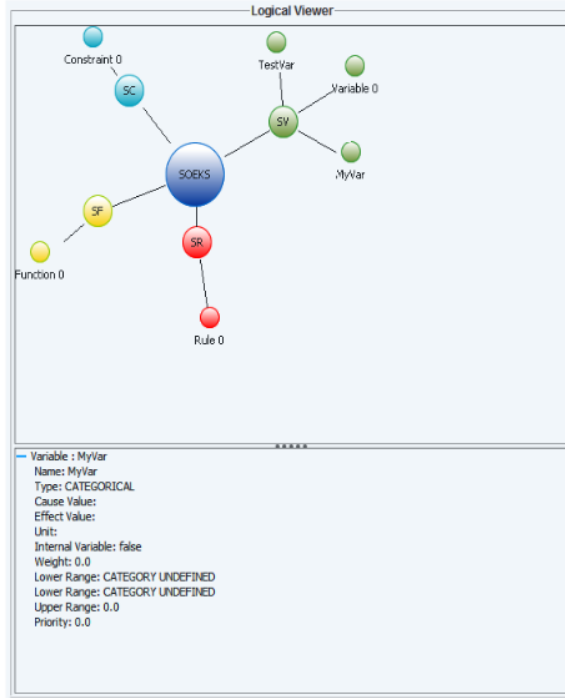


Fig.2. Software platform for SOE development and experience management

### III. SEARCHING FOR SIMILARITIES

Similarity between objects is an important concept in data mining as it can help, for example, in predictions, hypothesis testing, and rule discovery [6]. Additionally, it can offer rankings of results for choosing the best matching objects. Evaluating similarity between data objects depends essentially on the type of the data. Different similarity measures can reveal different aspects of the data, and as a result, two data objects can be classified to be similar by one measure and different by another measure. Nevertheless, it is possible to have several types of similarity measures on a single set of data, i.e. heterogeneous similarity measures. This means that it is significant to choose the one that best suits our purposes.

A typical data object is a relation that consists of several attributes. Approaches generally used for defining similarity between two data objects consider their attributes in pairs and compare them as isolated elements. Hence, similarity metrics can be calculated based on a composed addition of many different comparisons. A common approach for similarity between objects is defined in terms of a notion of distance (i.e. a geometric approach), and it is called multidimensional scaling.

The most common distance measure calculates similarity  $d$  between objects,  $s_i$  and  $s_j$  as follows:

$$d_{ij} = \left[ \sum_k w_k |s_{ik} - s_{jk}|^r \right]^{1/r} \quad (1)$$

where  $w_k$  is the weight given to the  $k^{\text{th}}$  attribute, and  $r$  is a parameter that determines which of the family of metrics is used.  $r = 2$  corresponds to the Euclidean metric,  $r = 1$  results in the City-Block, Taxi-Cab or Manhattan metric, while as  $r \rightarrow \infty$  is the Supremum or Maximum metric. More distance metrics that do not belong to this family can also be employed to compute similarity measures, for instance the Mahalanobis, Canberra, Bray-Curtis, and Bhattacharyya distances [7].

### IV. SIMILARITY FOR SET OF EXPERIENCE KNOWLEDGE REPRESENTATION

#### A. Variables

Given a pair of Sets of Experience  $E_i$  and  $E_j \in S$ , it is possible to generate a similarity metric of the variables called  $S_V \in [0, 1]$  by calculating the distance measure between each of the pairwise attributes  $k \in E_i$  and  $E_j$ :

$$S_V(E_i, E_j) = \sum_{k=1}^n w_k \left[ \frac{|E_{ik}^2 - E_{jk}^2|}{\max(|E_{ik}|, |E_{jk}|)^2} \right]^{1/2} \quad \forall k \in E_i \wedge E_j \quad (2)$$

where  $E_{ik}, E_{jk}$  are the  $k^{th}$  attribute of the sets  $E_i$  and  $E_j$ ,  $w_k$  is the weight given to the  $k^{th}$  attribute, in this case, variable; and  $n$  is the number of variables on  $E_i$ .

### B. Functions and Constraints

Similarity metrics for functions and constraints are calculated in the same way (constraints are a form of functions) simply by comparing their factors. In case of qualitative functions, they are first converted into a vector of factors. The general similarity metric  $S_{FQ} \in [0, 1]$  is calculated as follows:

$$S_{FQ}(E_i, E_j) = \sum_{k=1}^n w_k \cdot \begin{cases} 1 \Leftrightarrow F_{ik} \in E_i \wedge F_{ik} \notin E_j \\ 0 \Leftrightarrow F_{ik} \in E_i \wedge F_{ik} \in E_j \end{cases} \quad (3)$$

where  $F_k$  is the function being analyzed,  $w_k$  is the weight associated to the  $k^{th}$  function in the SOE  $E_i$ .

Rules are elements that regulate the decision event, but they are not elements of comparison due to its conditional characteristic. Rules are part of the decision event that influences the variables according to certain conditions, but because not all of them are executed, they are not part of the similarity estimation.

Finally, defining a similarity metric for the SOE is as simple as adding up all the estimations for each element: variables, functions, and constraints, and assuming that all of them are assigned with the same weight due to their equal importance on influencing the formal decision event:

$$S(E_i, E_j) = \frac{S_V(E_i, E_j) + S_F(E_i, E_j) + S_C(E_i, E_j)}{3} \quad (4)$$

With the developed similarity metrics for the SOE, which makes SOE structure comparable and classifiable, the possibilities are open for this knowledge representation concept being extendedly used within many different systems and technologies. In practice, similarity measurement is performed with the support of software platform already briefly introduced and illustrated in Fig. 2.

### V. SIMILARITY MEASUREMENT IN PRACTICE

We are developing SOE similarity application as the core of an AI engine for idream technology platform (<https://www.idream.technology>). It provides SOE technology based tools to enhance peoples' wellbeing and sustainable development. IDREAM TECHNOLOGY PTY LTD applies SOE in Drimos® engine to develop a personalized goal achieving plan using collective intelligence captured from collective experience of platform users. The implemented similarity concept allows creating

"fingerprints" distinctive for each individual user and recommends a tailored action path that helps achieving their goals and aims.

Each individual platform user generates his/her profile which is presented by a dedicated SOE. The applied algorithm captures, integrates, stores and reuses thousands of experiences occurring during the process of achieving dreams or goals. In other words, it uses collective decision-making experience and applies it to amplify the individual one. The amplification is based on finding similar experiences and then developing a systematic procedure to follow in order to achieve personal objective by a given platform user. In other words, Idream.Technology platform applies the SOE similarity notion to the distance between variables describing profiles of each participating member of Idream.Technology community. The most similar profiles are chosen by the system to assist in the creation of successful path to reach the goals set up by others. Fig.3. presents the screenshot of similarity engine in action.

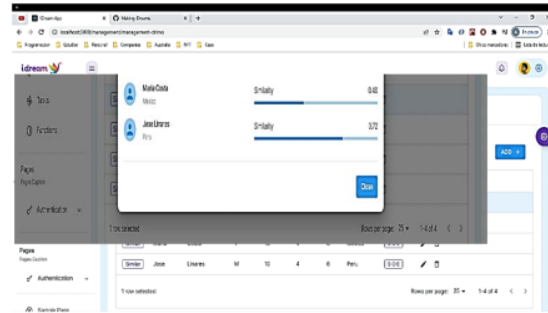


Fig. 3 Similar profile selection based on SOE similarity metric

### VI. CONCLUSION AND FUTURE OUTLOOK

Marvin Minsky [8], creator of the frame concept, referring to knowledge representation says, "each particular kind of data structure has its own virtues and deficiencies, and none by itself would seem adequate for all the different functions involved". Our comparable and classifiable SOE knowledge structure is a suitable representation for formal decision events. The presented in this paper SOE similarity measure facilitates SOE concept extension into different domains of applications.

Set of experience knowledge representation concept is at the beginning of its advance but is already making a difference to intelligence amplification. Future integration of various SOE extensions would provide an intelligent and sustainable Internet application environment that enables virtual roles (mechanisms that facilitate interoperability among users, applications, and resources) to effectively capture, publish, share and manage explicit knowledge resources. It would also provide support for on-demand

services creating vast commercial potential for our technology. Through our approach to knowledge representation and formalization embedded in the concept of SOE-based knowledge grid would incorporate epistemology and ontology to reflect human cognition characteristics and adopt the techniques and standards developed during work toward the next-generation, beyond-semantic web [9].

## REFERENCES

- [1] C. Toro, E. Sanchez, E. Carrasco, L. Mancilla-Amaya, C. Sanin, E. Szczerbicki, M. Graña, P. Bonachela, C. Parra, and G. Bueno, "Using set of experience knowledge structure to extend a rule set of clinical decision support system for Alzheimer's disease diagnosis", *Hermetics and Systems*, 43(2), 2013.
- [2] de Oliveira, C. S., C. Sanin, and E. Szczerbicki, (2019). "Towards Knowledge Formalization and Sharing in a Cognitive Vision Platform for Hazard Control (CVP-HC)". *Proceedings Asian Conference on Intelligent Information and Database Systems* (pp. 53-61). Springer, 2019.
- [3] Richardson, M., Domingos, P.: Markov logic networks. *Mach Learn* 62, 107–136 <https://doi.org/10.1007/s10994-006-5833-1>
- [4] Malhotra, Yogesh, AI Machine Learning & Deep Learning Risk Management & Control, Beyond Deep Learning and Generative Adversarial Networks, Princeton Fintech & Quant Conference, Princeton University, April 21, 2018., <https://ssrn.com/abstract=3167035> or <http://dx.doi.org/10.2139/ssrn.177035> (2018).
- [5] Mabin, V.J., & Balderstone S.J.: The World of the Theory of Constraints: A Review of the International Literature, CRC Press. <https://doi.org/10.4324/9780429273056> (2020).
- [6] Sowmya, R., and K. R. Suneetha: Data mining with big data. In *2017 11th International Conference on Intelligent Systems and Control (ISysC)*, pp. 246-250. IEEE, (2017).
- [7] Fabrikant, S.I. and B.P. Battenfield: Formalizing semantic spaces for information access, *Annals of the Association of American Graphers*, 91(2), pp. 263-280 (2001).
- [8] Minsky M, Singh P, Sloman A. (2004) The St Thomas common sense symposium: Designing architectures for human-level intelligence *AI Magazine*. 25: 113-124. doi: 10.1609/Aimag.V25I2.1764
- [9] T. de Souza Alves, de Oliveira C.S., C. Sanin, and E. Szczerbicki, (2018), "From Knowledge-based Vision Systems to Cognitive Vision Systems: A Review", *Proceedings of Knowledge-Based Intelligent Information and Engineering Systems 22nd International Conference KES 2018*, in *Advances in Knowledge-Based and Intelligent Information and Engineering Systems*: R. J. Howlett, L. C. Jain (Eds.), Belgrade, Sep 2018, Elsevier Procedia Computer Science, 2018, DOI: 10.1016/j.procs.2018.08.077

# 72%

SIMILARITY INDEX

### PRIMARY SOURCES

- 1

"Knowledge Management and Engineering with Decisional DNA", Springer Science and Business Media LLC, 2020  
Crossref

246 words — 11%
- 2

mostwiedzy.pl  
Internet

196 words — 9%
- 3

annals-csis.org  
Internet

175 words — 8%
- 4

Edward Szczerbicki, Cesar Sanin. "Experience-based Intelligence Augmentation with Decisional DNA: Upcoming Direction", Procedia Computer Science, 2022  
Crossref

155 words — 7%
- 5

Cesar Sanin. "DEVELOPING HETEROGENEOUS SIMILARITY METRICS FOR KNOWLEDGE ADMINISTRATION", Cybernetics & Systems, 9/1/2006  
Crossref

142 words — 7%
- 6

nozdr.ru  
Internet

106 words — 5%
- 7

Cesar Sanin, Edward Szczerbicki. "EXTENDING SET OF EXPERIENCE KNOWLEDGE STRUCTURE INTO A TRANSPORTABLE LANGUAGE eXTENSIBLE MARKUP LANGUAGE", Cybernetics and Systems, 2006  
Crossref

103 words — 5%

|    |                                                                                                                                                                                                 |               |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| 8  | <a href="http://nova.newcastle.edu.au">nova.newcastle.edu.au</a><br>Internet                                                                                                                    | 62 words — 3% |
| 9  | <a href="http://ehu.eus">ehu.eus</a><br>Internet                                                                                                                                                | 48 words — 2% |
| 10 | Cesar Sanin, Edward Szczerbicki. "Using Set of Experience in the Process of Transforming Information into Knowledge", International Journal of Enterprise Information Systems, 2006<br>Crossref | 45 words — 2% |
| 11 | <a href="http://papers.ssrn.com">papers.ssrn.com</a><br>Internet                                                                                                                                | 39 words — 2% |
| 12 | <a href="http://depot.ceon.pl">depot.ceon.pl</a><br>Internet                                                                                                                                    | 36 words — 2% |
| 13 | <a href="http://neurotree.org">neurotree.org</a><br>Internet                                                                                                                                    | 29 words — 1% |
| 14 | <a href="http://link.springer.com">link.springer.com</a><br>Internet                                                                                                                            | 25 words — 1% |
| 15 | <a href="http://cartesia.org">cartesia.org</a><br>Internet                                                                                                                                      | 24 words — 1% |
| 16 | <a href="http://journals.biologists.com">journals.biologists.com</a><br>Internet                                                                                                                | 21 words — 1% |
| 17 | <a href="http://www.mdpi.com">www.mdpi.com</a><br>Internet                                                                                                                                      | 21 words — 1% |
| 18 | <a href="http://www.tandfonline.com">www.tandfonline.com</a><br>Internet                                                                                                                        | 16 words — 1% |

|    |                                                                                                                                                                                     |                 |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| 19 | Lecture Notes in Computer Science, 2006.<br>Crossref                                                                                                                                | 13 words — 1%   |
| 20 | www.slideserve.com<br>Internet                                                                                                                                                      | 13 words — 1%   |
| 21 | vdoc.pub<br>Internet                                                                                                                                                                | 12 words — 1%   |
| 22 | epdf.tips<br>Internet                                                                                                                                                               | 10 words — < 1% |
| 23 | Hayashi, T.. "Coribbon Hopf (Face) Algebras<br>Generated by Lattice Models", Journal of Algebra,<br>20001115<br>Crossref                                                            | 8 words — < 1%  |
| 24 | Cesar Sanin, Edward Szczerbicki. "Chapter 135<br>Using XML for Implementing Set of Experience<br>Knowledge Structure", Springer Science and Business Media<br>LLC, 2005<br>Crossref | 7 words — < 1%  |

EXCLUDE QUOTES OFF  
EXCLUDE BIBLIOGRAPHY OFF

EXCLUDE SOURCES OFF  
EXCLUDE MATCHES OFF