Ontology-Driven Data Mining Platform for Fuzzy Classification of Mental Maps

Konstantin RYABININ [1], Konstantin BELOUSOV and Roman CHUMAKOV

Perm State University, Perm, Russia

Abstract. Mental maps are valuable material for Digital Humanities research since they represent a summary of humans spatial experience reflected in their minds. Contributing to this research, we developed a high-level Web-based software platform that allows to collect drawings of mental maps and to perform corresponding data mining and fuzzy classification. The novelty of the proposed platform is the ontology-based integration of mental maps drawing engine and data mining engine, wherein all the essential steps of data mining, including data acquisition, transformation, fuzzy classification, and visual analytics are driven by ontologies. The platform consists of a high-level graphical editor to draw maps and a data flow diagram editor to define the data mining pipeline. The operators available to construct this pipeline are described by ontology, which ensures the platform's extensibility on the knowledge base level. Thereby, the platform created can be used not only for Digital Humanities research but also for testing and evaluation of new data mining and fuzzy classification methods. Currently, we have evaluated weighted fuzzy pattern matching for mental maps fuzzy classification and confirmed the reasonable performance of this method.

Keywords. Mental Map, Data Mining, Fuzzy Classification, Ontology Engineering, Data Flow Diagrams

1. Introduction

Mental maps are a representation of spatial experience in the human mind [1]. They have been a research object in different social sciences since the second half of the 20th century, serving a tool able to reflect space perception, cognitive processes, and environmental behavior of individuals [1,2]. Information technologies like data mining (DM) and visual analytics bring new tools to leverage this research, raising the mental maps processing to the next level of automation [3].

Mental maps of particular sites (cities, regions, countries, etc.) can be reconstructed by so-called "sketch maps". While some authors treat "sketch maps" and "mental maps" as synonyms [4], in this work we denote "sketch map" a particular drawing made by an informant, and "mental map" a meaningful model behind this drawing. Therefore, the DM is performed on mental maps, taking sketch maps as raw input.

It turned out that informants when developing a mental map of their country are highly influenced by their regional identity, reflecting the complex state of the corre-

---

[1]Corresponding Author: Konstantin Ryabinin, Perm State University, Bukireva Str. 15, 614990 Perm, Russia; E-mail: kostya.ryabinin@gmail.com.

sponding region from the residents point of view [5]. This is why regional mental maps are valuable material for Digital Humanities research.

One of the most important tasks of mental maps analysis is classification. It allows discovering the informants attitude to the surrounding spatial environment, which, in turn, helps to justify the public moods in a particular region or social group [2,5], as well as to measure individual differences of informants [6]. Mental maps by their nature contain a lot of variations and uncertainties, so classical methods (including classical set theory) are inappropriate for their comparison/classification. In this regard, fuzzy set theory can be used to tackle the mental maps classification problem.

When it comes to sketch maps, traditionally, informants draw them "by hand on a blank sheet of paper, expressing a subjective reality of space using individual and social memory" [4]. However, this approach hinders a straightforward digital analysis of such maps, because the problem of pattern recognition should involve not only map analysis itself, but also non-trivial preprocessing stage based on computer vision. Therefore, special software sketch map editors are required to digitize the drawing process thus improving the speed and efficiency of mental maps DM.

The goal of our work is to develop a software platform comprising both an informant-friendly sketch map editor to collect the appropriate material and analyst-friendly extensible tools to perform fuzzy classification for the purpose of mental map DM to support Digital Humanities research.

The distinctive feature of the approach proposed is the use of ontology engineering methods and means to unify all the essential steps of DM, including data acquisition, appropriate transformation, fuzzy classification, and subsequent visual analytics within a single Web-based ontology-driven platform. The easiness of extensibility ensured by ontologies allows using this platform not only for DM itself but also for testing and evaluation of new DM and fuzzy classification methods.

2. Related Work

Mental maps are extensively used in social sciences as a material to elicit and analyze socio-spatial representations [4], as well as to study the human skills of understanding the structure and functions of a surrounding space [7]. For example, C. Didelon-Loiseau et al. utilized mental maps to identify the regional perception of the whole world by respondents from different countries [1]. In contrast to this, S. Gül used mental maps as a basis to study the urban perception of students at the city level [8]. K. Hátlová et al. found out that mental maps are most of all influenced by biological, psychological, sociocultural, environmental, and educational factors [6].

One of the algorithms suitable for mental maps classification is weighted fuzzy pattern matching [9]. It is relatively old yet pretty reliable, so we decided to use this algorithm for the first test of our approach. In the future, we plan to consider algorithms based on machine learning in the way suggested by Y. Bodyanskiy et al. (which takes numerical data as input) [10] and by H.-X. Li et al. (which works with linguistic rules) [11].

Regarding the software, F. Aram et al. introduced the system called AMMA that enables quantitative evaluation of mental maps complexity and accuracy based on landmarks and paths [3]. Although this system implements a lot of evaluation methods, it is mainly restricted by urban-scale operation and does not provide tools for drawing men-

tal maps. Instead, it allows comparing scanned hand-drawn mental maps with accurate geographical maps.

The well-known popular software like KNIME, Weka, and RapidMiner [12] incorporate state-of-the-art DM methods, which makes them quite universal standalone desktop tools for analysts. However, when it comes to seamless integration with Web applications or evaluation/testing of new DM techniques, there arise certain limitations. In this regard, high flexibility of customization is required. A promising way to achieve flexibility is an ontology-driven software development approach [13].

According to the elaborate survey of D. Dou et al., ontologies enable so-called semantic DM that bridges "semantic gaps between the data, applications, data mining algorithms, and data mining results", as well as provide "a formal way for representing the data mining flow" [14]. Y. Li et al. proposed an ontology-based framework for self-service knowledge discovery, utilizing ontology to ensure user-centric DM model management capabilities, including selection and reuse of appropriate models [15]. P. Yan et al. proposed an ontological "model to uncover semantic paths between concepts [...] across documents incorporating Wikipedia knowledge" as a background [16]. This model gains Map-Reduce approach to ensure parallelism and outperforms its analogs in terms of semantic information retrieval and relation discovery [17]. K. Joshi et al. proposed using ontology in a fuzzy classification process to improve semantic information retrieval [18]. K. Okoye suggested to enrich fuzzy models with ontology-based semantic annotations, which allow "to determine [...] the presence of different patterns within the discovered models" [19].

Relying on this positive experience, we have adapted our previously created ontology-driven visual analytics platform SciVi [20] to the tasks of fuzzy classification within the DM process to ensure its configurability and integration capabilities.

3. Proposed Solution

3.1. Background

While working on our previous research, we have created two independent software systems: SciVi ontology-driven client-server platform focused on DM and visual analytics (https://scivi.tools/) [20], and Creative Maps Studio focused on creating and editing sketch maps.

The functioning of the SciVi platform is governed by ontologies, which describe available data extraction, transformation, load, and mining mechanisms, whereby forming a SciVi knowledge base (KB). Correspondingly, there are two SciVi user categories: knowledge engineers and analysts. A knowledge engineer acts as a KB administrator, who adapts SciVi to solve the new DM tasks by extending underlying ontologies with necessary functions description or modifying the existing functions descriptions to fine-tune related data processing mechanisms. The ontologies are modified with ONTOLIS visual editor [21] and stored in the lightweight JSON-based format called ONT. An analyst acts as a SciVi end-user, who builds actual DM pipelines using data flow diagrams (DFDs) and performs the data analysis. For this, SciVi provides a high-level graphical DFD editor. DFDs proved their efficiency as a visual language for defining DM pipelines in a variety of real-world case studies in different popular DM software, including KN-

IME, Weka, and RapidMiner. What makes SciVi platform distinctive is that available DFD operators are described by underlying ontologies, so to introduce new operators knowledge engineer just needs to extend the SciVi KB without modifying its source code. Thereby, knowledge engineer can prepare the appropriate operators palette for analyst, without engaging SciVi developers. Technically, each operator is a micro-plugin for the SciVi platform managed by a corresponding fragment of ontology. An ontology-driven approach to managing plugins facilitates the incorporating both self-written and third-party data processing libraries into SciVi platform and documenting them.

Creative Maps Studio (https://creativemaps.studio/) is a Web-based editor designed for sketch map drawing. It is implemented using React as a frontend library, FastAPI as a backend framework, and PostgreSQL as a database management system. Unlike traditional maps drawn on paper, digital maps possess full power of automated data processing, including DM, fuzzy set theory, and visual analytics. Moreover, the digital form makes the data collection process much easier, enlarging the geographical coverage of the experiments beyond the borders of particular cities up to nationwide scale.

3.2. Integrating Creative Maps Studio with SciVi

To enable in-place mental map DM inside the Creative Maps Studio, we decided to integrate the SciVi platform into it on the basis of loose coupling using ontology-driven REST API.

We have developed a special account type in the Creative Maps Studio, enabling analyst user category along with the regular end-user (informant). In this special account, analyst can see all available sketch maps created by informants and select the arrays of maps for DM. Selected maps can be exported to SciVi. To optimize export process in terms of ergonomics and network traffic, we extended the SciVi server and client with corresponding REST API that enables to request available operators and to order instancing of specified operators with specified settings in the DFD.

Communication between Creative Maps Studio and SciVi platform is as follows. When the analyst logs in to his/her account, Creative Maps Studio requests available data source connectors from the SciVi server, specifying the name of corresponding SciVi DM operators ontology. The SciVi server responds with an ONT-formatted ontology O describing available data extraction operators. This ontology defines the available settings and URI of each operator.

The export menu in the Creative Maps Studio is automatically filled up with available operators, enabling analyst to select maps of interest and give a command to load them to SciVi using particular data extraction operator. Upon this command, the Creative Maps Studio opens up the window with the SciVi client (providing the SciVi DFD editor) and instructs it via WebSocket to instantiate the chosen operator with appropriate settings in DFD editor. WebSocket API call syntax is automatically inferred by Creative Maps Studio from the ontology O. This action can be repeated multiple times when analyst needs to export different maps or map arrays for upcoming DM. After the export of needed maps, analyst specifies the rest of the DM pipeline within the SciVi client DFD editor by instantiating required operators and linking them to define a data flow.

Upon the explicit command of analyst, SciVi executes the created pipeline, wherein the execution of particular operators is being distributed between the SciVi server and

client according to the meta-information stored in the operators ontology. Data flow between server and client is seamless thanks to a direct WebSocket connection. Execution of data extraction operators leads to the specific GET requests to the Creative Maps Studio server that responds with JSON-formatted maps or map arrays. Maps are preprocessed by bringing all the numerical values of objects parameters to the single discrete scale suitable for subsequent classification.

Being ontology-driven, the communication protocol of Creative Maps Studio and SciVi is very flexible: should SciVi platform be extended with new data extraction mechanisms suitable for mental maps, these mechanisms will be automatically available in Creative Maps Studio.

3.3. Fuzzy Classification of Mental Maps in SciVi

In our research, we applied a weighted fuzzy pattern matching algorithm [9] to measure similarity between mental maps created by random informants and mental maps of people living in particular regions. The DM pipeline declared as a DFD within the SciVi platform environment is shown in Figure 1. Each DFD node represents a single operator described in the SciVi ontology.
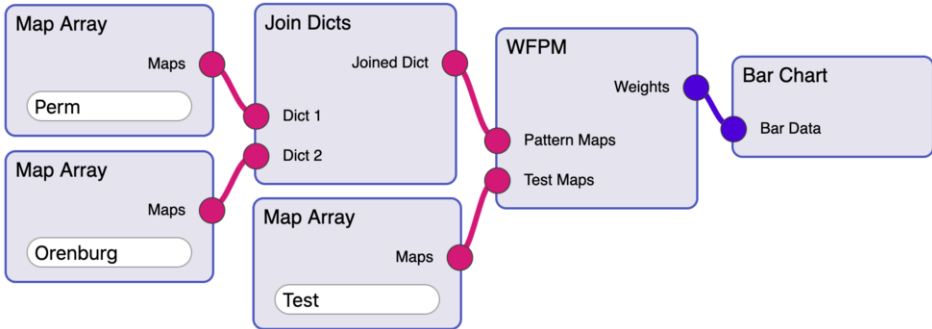

Figure 1. DM pipeline defined in the SciVi platform using DFD graphical notation.

"Map Array" nodes depict the data extraction operators, which are generated automatically according to analyst's choice as described in Section 3.2. The extracted data are encoded as a key-value dictionary, where the key is the name of map array ("Perm", "Orenburg", and "Test", respectively; these names are assigned by analyst when performing export operation), and the value is the actual array of corresponding maps. The "Join Dicts" node represents the concatenation of dictionaries.

"WFPM" node stands for "Weighted Fuzzy Pattern Matching" operation, which relies on the algorithm presented in [9]. Formally speaking, each map is a set of objects, and each object is a set of properties like coordinates, size, color, etc. Currently, we take into consideration only numerical properties of settlements.

Let us denote the fuzzy classification of maps as F. Then

$$F(P,T,W) = \bigcup_{j=1}^{|T|} \bigcup_{i=1}^{|P|} \max_{k=1,|T_j|} \left(\max_{l=1,|W|} \left(\min \left(\mu_{k,l}(P_i, T_{j,k,l}), W_l \right) \right) \right), \tag{1}$$

where P is a pattern set (array of maps of known regions; this array is depicted in Figure 1 as "Pattern Maps" slot), T is a test set (array of maps, for which to find fuzzy belonging to the regions from the pattern set; this array is depicted in Figure 1 as "Test Maps" slot), P_i is an individual pattern representing the i-th region, T_j is a j-th test map, $T_{j,k,l}$ is an l-th property of k-th settlement of j-th test map, W_l is a weight denoting the importance of the l-th property, $\mu_{k,l}$ is a function of fuzzy belonging of a j-th test map to the i-th pattern calculated for k-th settlement according to its l-th property. Denoting $T_{j,k,l}$ as x,

$$\mu_{k,l}(P_i,x) = D_{i,k,l}(g(x)) + \frac{D_{i,k,l}(h(x)) - D_{i,k,l}(g(x))}{h(x) - g(x)}(x - g(x)), \tag{2}$$

where $D_{i,k,l}$ is a frequency distribution of $\{P_{i,m,k,l} \mid m = \overline{1,|P_i|}\}$ set (m denotes particular map in the i-th pattern), $g(x)$ clamps x to the nearest element of $P_{i,k,l}$ range, and $h(x)$ picks the element next to the one given by $g(x)$. The μ function is extended by 0 wherever (2) is undefined.

The ontology fragment describing the "WFPM" operator is shown in Figure 2. It declares that "WFPM" is a "Fuzzy Classification" method with two inputs ("Pattern Maps" and "Test Maps") of the "MapArrayDict" type and one output ("Weights") of the "Grid" type. There are also multiple settings defining the weights of different map objects properties used in fuzzy pattern matching (in Figure 2 only "X Weight", "Y Weight", and "Size Weight" are shown, however in fact there are more weights; the entire ontology of WFPM operator is available in the SciVi OpenSource repository under https://github.com/scivi-tools/scivi.web). An important part of the description is "WFPM Worker" concept representing an actual client-side implementation of the operator that contains a link to the JavaScript source code. The source code is automatically appended to the SciVi Web client when the corresponding operator is requested. This is how the SciVi plugin management works. This way of adding plugins appears to be very fast, easy, and flexible, as we have proved earlier [20].
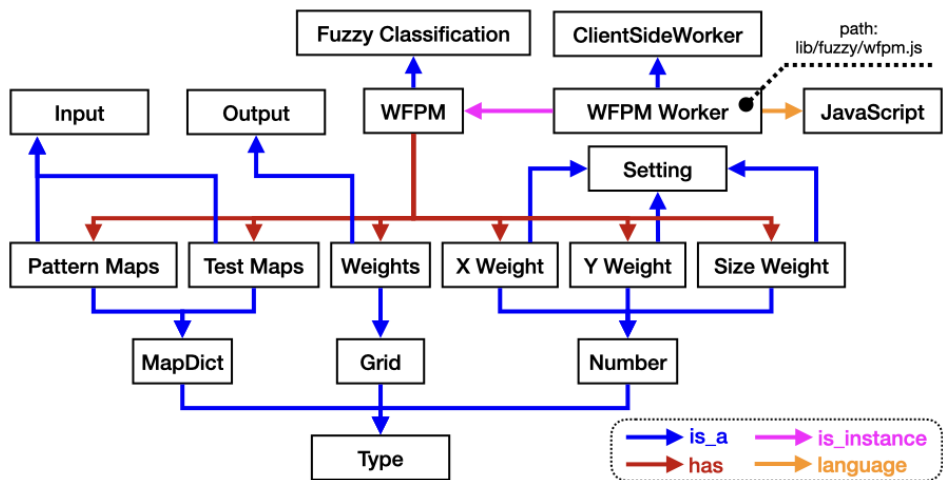

Figure 2. SciVi ontology fragment describing the fuzzy classification operator.

The last step of the DM pipeline is data visualization as a bar chart ("Bar Chart" node), which makes the classification more illustrative. The visualization result is

shown in Figure 3. The interactive visualization is available online at `https://scivi.semograph.com/mmaps?preset=fuzzyMentalMaps.json`. The quantitative analysis and significance test of results obtained is presented in Table 1. WFPM and Reference columns indicate the obtained and ground-truth values in a format (*Perm*; *Orenburg*). The ground-truth values are known to us since we used maps of the informants with known residence to test the WFPM algorithm. Residual is calculated based on normalized Euclidean distance from WFPM result to the Reference value. The average residual is 0.33, so we consider WFPM as an acceptable fuzzy classification algorithm.
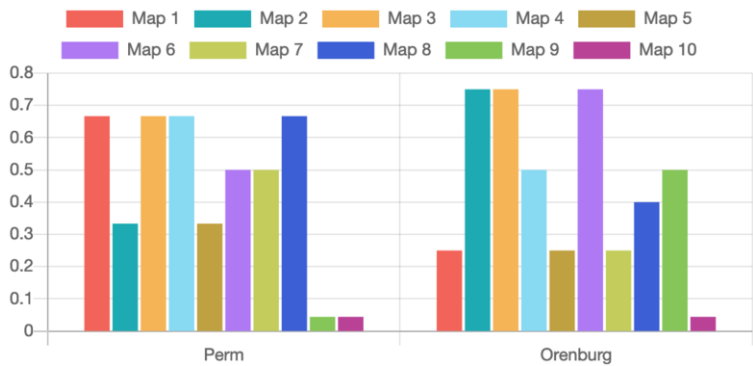

Figure 3. Visualization of mental maps fuzzy classification using a bar chart.

Map	WFPM	Reference	Residual
1	(0.67; 0.25)	(1; 0)	0.29
2	(0.33; 0.75)	(0; 1)	0.29
3	(0.67; 0.75)	(0; 1)	0.51
4	(0.67; 0.5)	(1; 0)	0.42
5	(0.33; 0.25)	(0; 0)	0.29
6	(0.5; 0.75)	(0; 1)	0.4
7	(0.5; 0.25)	(1; 0)	0.4
8	(0.67; 0.4)	(1; 0)	0.37
9	(0; 0.5)	(0; 1)	0.35
10	(0; 0)	(0; 0)	0

Table 1. WFPM quantitative analysis and significance test.

4. Conclusion

We intensively use mental maps drawn by informants to study regional identity within Digital Humanities research. This research requires flexible and extensible DM tools leveraged by fuzzy maths. To suit this need, we integrated Creative Maps Studio map drawing software with SciVi visual analytics system using ontology engineering methods and means. Thereby, the entire research, from the data collecting up to the visualization of the fuzzy classification results, can be conducted within a single software platform. We tested this platform performing weighted fuzzy pattern matching of the mental maps to determine the fuzzy regional identity of particular informants. An ontology-driven approach ensures easy extending of this platform with new DM algorithms, which will help us to evaluate other fuzzy classification methods in the future.

Acknowledgments

References

[1] Didelon-Loiseau C, de Ruffray S, Lambert N. Mental Maps of Global Regions: Identifying and Characterizing "Hard" and "Soft" Regions. Journal of Cultural Geography. 2018;35(2):210–229.

[2] Henrikson AK. The Geographical "Mental Maps" of American Foreign Policy Makers. International Political Science Review / Revue internationale de science politique. 1980;1(4):495–530.

[3] Aram F, Solgi E, García EH, Mohammadzadeh D, Mosavi A, Shamshirband S. Design and Validation of a Computational Program for Analysing Mental Maps: Aram Mental Map Analyzer. Sustainability. 2019;11(14).

[4] Dernat S, Johany F, Lardon S. Identifying Choremes in Mental Maps to Better Understand Socio-Spatial Representations. Cybergeo: European Journal of Geography. 2016.

[5] Zelyanskaya N, Belousov K, Ichkineeva D. Naive Geography and Geopolitical Semiotics: The Semiotic Analysis of Geomental Maps of Russians. Semiotica. 2017;2017(215):235–253.

[6] Hátlová K, Hanus M. A Systematic Review into Factors Influencing Sketch Map Quality. ISPRS International Journal of Geo-Information. 2020;9(4).

[7] Castellar SMV, Juliasz PCS. Mental Map and Spatial Thinking. In: Proceedings of the ICA. vol. 1; 2018. .

[8] Gül S. A Case Study on the Urban Perception of Disabled Students in Samsun (Turkey). Review of International Geographical Education Online. 2020;10:226–248.

[9] Dubois D, Prade H, Testemale C. Weighted Fuzzy Pattern Matching. Fuzzy Sets and Systems. 1988;28(3):313–331.

[10] Bodyanskiy Y, Dolotov A, Peleshko D, Rashkevych Y, Vynokurova O. Associative Probabilistic Neuro-Fuzzy System for Data Classification Under Short Training Set Conditions. In: Contemporary Complex Systems and Their Dependability; 2019. p. 56–63.

[11] Li HX, Wang Y, Zhang G. Probabilistic Fuzzy Classification for Stochastic Data. IEEE Transactions on Fuzzy Systems. 2017;25(6):1391–1402.

[12] Naik A, Samant L. Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science. 2016;85:662–668.

[13] Pan JZ, Staab S, Assmann U, Ebert J, Zhao Y. Ontology-Driven Software Development. Springer; 2013.

[14] Dou D, Wang H, Liu H. Semantic Data Mining: A Survey of Ontology-Based Approaches. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015); 2015. p. 244–251.

[15] Li Y, Thomas MA, Osei-Bryson KM. Ontology-Based Data Mining Model Management for Self-Service Knowledge Discovery. Information Systems Frontiers. 2017;19:925–943.

[16] Yan P, Jin W. Mining Semantic Relationships between Concepts across Documents Incorporating Wikipedia Knowledge. In: Advances in Data Mining. Applications and Theoretical Aspects. vol. 7987; 2013. p. 70–84.

[17] Yan P, Jin W. Building Semantic Kernels for Cross-Document Knowledge Discovery Using Wikipedia. Knowledge and Information Systems. 2017;51:287–310.

[18] Joshi K, Verma A, Kandpal A, Garg S, Chauhan R, Goudar RH. Ontology Based Fuzzy Classification of Web Documents for Semantic Information Retrieval. In: 2013 Sixth International Conference on Contemporary Computing (IC3); 2013. p. 1–5.

[19] Okoye K. Technique for Annotation of Fuzzy Models: A Semantic Fuzzy Mining Approach. Frontiers in Artificial Intelligence and Applications. 2019;320:65–75.

[20] Ryabinin KV, Belousov KI, Chuprina SI. Novel Circular Graph Capabilities for Comprehensive Visual Analytics of Interconnected Data in Digital Humanities. Scientific Visualization. 2020;12(4):56–70.

[21] Chuprina S, Nasraoui O. Using Ontology-based Adaptable Scientific Visualization and Cognitive Graphics Tools to Transform Traditional Information Systems into Intelligent Systems. Scientific Visualization. 2016;8(1):23–44.