

Measuring Multimedia Content Proximity via Artificial Intelligence Methods

Yury Megel

*dept. of Cybernetics
Kharkiv Petro Vasylenko National
Technical University of Agriculture
Kharkiv, Ukraine
megelye@gmail.com*

Olena Mikhnova

*dept. of Cybernetics
Kharkiv Petro Vasylenko National
Technical University of Agriculture
Kharkiv, Ukraine
mikhnova@yahoo.co.uk*

Svitlana Kovalenko

*dept. of Cybernetics
Kharkiv Petro Vasylenko National
Technical University of Agriculture
Kharkiv, Ukraine
svetkovalenko@gmail.com*

Sergey Kovalenko

*dept. of System Analysis and
Information-Analytical Technologies
National Technical University "KhPI"
Kharkiv, Ukraine
adbc@ukr.net*

Antonina Rybalka

*dept. of Physics
Kharkiv National University of Radio
Electronics
Kharkiv, Ukraine
antonina.rybalka@nure.ua*

Igor Chalyi

*dept. of Cybernetics
Kharkiv Petro Vasylenko National
Technical University of Agriculture
Kharkiv, Ukraine
ivchaly@gmail.com*

Ivaylo Blagov

*dept. of Precise Techniques
and Measuring Instruments
Technical Universitet of Sofia
Sofia, Bulgaria
ivayioblagov@abv.bg*

Abstract—The paper describes existing and novel approaches to word sequence similarity measuring which uses image processing procedures to reveal proximity relations between words and their meaning. This is implemented due to the fact the each word sequence can be associated with a particular pictorial domain, and thus any textual information can be potentially presented via an image series. Theoretical results of rule based approach, genetic algorithms and neural networks application are analyzed in terms of solving the problem under study. The proposed information processing technology may be useful for automatic translation applications, known as CAT-tools by now, search engine optimization, etc. The proposed content analysis techniques are designed to boost translation efficiency in particular. Romano-germanic and east-slavonic language pairs have been used as an example of practical application. It aids translation experts in finding the best solution for a specific domain that he (or she) is usually unfamiliar with. Thus, eliminating human errors and decreasing time needed for contextual search and deep examining of the topic.

Keywords—*computer-assisted translation, evolutionary approach, multimedia processing, neural network, rule based algorithm*

I. INTRODUCTION

At the current stage of development of bilingual text processing techniques there are a number of automatic translation systems, but professional translators use only automated translation means in their everyday work. These are so called CAT-tools. Here, CAT is an abbreviation for Computer Assisted Translation. Such software means are designed to help experts in translation without his or her total replacement or exclusion. The latter cannot be currently implemented in full with quality preservation. Despite significant perfection of automatic translation systems during the last decade, which is testified by the appearance of mobile applications with relatively high quality of synchronous translation solutions, written translation in a

specific knowledge domain still remains the authority of a translator expert (see fig. 1 below as a strong evidence). In fig. 1 one may see a machine translation in the right column of two text fragments taken from a technical standard and a vintage item description.

For those who know Russian, it can be clearly seen that the Russian text looks absurd in both cases, though in general one may understand what it is about. However, special terms translated into the target language look as much absurd if performed by a specialist with poor qualification characteristics. At the same time, this expert may also have very severe time limits to do his or her work at a high quality level. It should be also mentioned that in order to find the correct translational solution, an expert should be familiar with the key concepts his translation is about which may appear too complicated for a linguist in case of a technical, medical or pharmaceutical domains. The current publication aims at suggesting a novel approach on automating search for the correct translation solution. The translation quality increase in application to the field-specific terms is proposed to be implemented via context related methods of intelligent image processing.

Among a variety of CAT-tools available in the translation market today, SDL Trados Studio gains the leading position by all means in the field of text translation and software localization (namely, SDL Passolo). TM (Translation Memory) and sometimes TB (Term Base) constitute any automatic translation system in its majority, and Trados is not an exception. These databases are accumulated with translation pairs of source and target segments while sole translator's work at a project or a group of translators work. In other words, TB and TM are constructed based on the translator's knowledge who further uses them. That is why the quality of computer-assisted translation differs greatly from the quality of fully automated translation where terms and grammatical links are translated solely by the machine, without any expert involvement [1].

Acknowledgements: The authors would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support.

In this work we propose the enhancement of current term translation means by introducing the analysis procedure and image matching. Such an approach can be implemented for both completely and partially automated systems. This will enlarge the percent of automation of the translation process and eliminate errors in specialized term translation by the

experts. Artificial intelligent methods, described and developed up till now and used for image matching, can be observed in [2-8]. To compare graphical data, common similarity metrics are often used along with cluster analysis, histogram difference, statistical analysis, optical flow analysis, etc.

<p>1930's Lion front door knocker. Antique brass this has a fantastic patina that only age can bring. This is a reclaimed example off an old front door. Instant character for your front door.</p>	<p>1930-е годы: дверной молоточек льва. Античная латунь имеет фантастическую патину, которую может принести только возраст. Это исправленный пример со старой парадной двери. Мгновенный персонаж для вашей входной двери.</p>
<p>Con Cast Pipe is the first precast and concrete pipe manufacturer in Canada to exclusively offer the PERFECT PIPE system. This innovation will revolutionize the landscape of infrastructure systems as we know it. With PERFECT PIPE, enjoy increased peace-of-mind when committing to a long-lasting pipe product backed by the dependability of the Con Cast Pipe brand.</p>	<p>Con Cast Pipe - первый производитель сборных и бетонных труб в Канаде, который эксклюзивно предложил систему PERFECT PIPE. Это нововведение революционизирует ландшафт инфраструктурных систем, какими мы его знаем. С ИДЕАЛЬНОЙ ТРУБОЙ вы можете быть спокойны, когда выбираете долговечный продукт для труб, подкрепленный надежностью бренда Con Cast Pipe.</p>

Fig. 1. Automatic English-Russian translation of 2 text fragments of different topics

II. EFFICIENCY OF RULE BASED, GA AND NEURONET APPROACHES FOR THE PROBLEM UNDER STUDY

The rule based approaches that use fuzzy logic models are considered to be the most applicable mathematical tool for text analysis despite apparent drawbacks that are further revealed in this study. Designing human approximate reasoning models allows us to make decisions based on information that is incomplete, fuzzy and/or with a large portion of entropy. Suppose $X = \{x_1, x_2, \dots, x_n\}$ is the initial data collection on the basis of which a decision will be made. Such an automated system has n inputs and one output. In our case, bilingual variants of a text sequence translation are the inputs, and we have only one truly correct output. Due to the specifics of the object under study, it is possible to conduct a quality assessment for any bilingual pair, and not to take into account all the available pairs at once.

As a next step, fuzzification of input variables is usually performed. That is, finding the correspondence between the crisp value of the input variable and the value of a term of a linguistic variable membership function. To create an assessment using a comprehensive quality indicator, n linguistic variables must be defined. Each variable A_i ($i = \overline{1, n}$) will be characterized by four linguistic terms α_{ij} ($i = \overline{1, n}, j = \overline{1, 4}$), which correspond to qualitative assessments "unsatisfactory", "satisfactory", "good", "excellent". Each term α_{ij} will be described using a fuzzy set $\alpha_{ij} = \left\{ \left\langle x_i, \mu_{\alpha_{ij}}(x_i) \right\rangle \right\}, i = \overline{1, n}, j = \overline{1, 4}$, where $\mu_{\alpha_{ij}}(x_i)$ is the value of membership function of x_i in the term $\alpha_{ij} \in A_i$, each of values $\mu_{\alpha_{ij}}(x_i)$ maps each element of the set $x_i \in X$ to a real number from the interval $[0, 1]$.

To determine values of membership functions, direct and indirect methods are used. Generally, direct methods are used for measurable concepts, and indirect methods are used for concepts that are difficult to measure. Experts satisfaction from the resultant translation is hard to be measured in our

case, as it can also be disputable among the experts, i.e. each person may have his or her opinion according to the correct translation. This phase requires additional study of the text subject matter even for humans. That is why formalization of the following step seems to fail in its initial intention. Moreover, prior development of rules totally diminishes translation correctness in a particular domain due to versatility of topics. Time needed for experts' knowledge presentation is enormous, and the efficiency of such a formalization is controversial.

Thus, experts' knowledge is represented in a form of fuzzy production rules $R = C \rightarrow D$, where " \rightarrow " is a symbol of fuzzy implication. For this purpose the right (antecedent) part of each rule is depicted as composite fuzzy logical statements, i.e. simple fuzzy statements are connected by fuzzy conjunctions, that is, we have i rules of the form:

$$R_i : \text{if } x_1 \text{ is } \alpha_{i1} \text{ and } \dots \text{ and } x_n \text{ is } \alpha_{in} \text{ then } y \text{ is } \beta_i, \quad (1)$$

where $i = \overline{1, m}$, m is the number of rules in the database; $x_j, j = \overline{1, n}$ are input variables; y is the output linguistic variable; α_{ij} and β_i are the membership functions defined respectively on x_j and y .

At the next step of fuzzy implication, the degrees of fuzzy premises fulfillment apply to the fuzzy conclusions for each rule. As a result, we get one fuzzy subset for the output y . The Mamdani's implication (i.e. using of min operator) can be used to resolve this issue. This actually is a "cut-off" in height of the output membership function at a level corresponding to the degree of fulfillment of the antecedent rule part. We also consider the composition according to Mamdani with the use of max operation. It means that the resulting set is obtained by taking pointwise maximums over all fuzzy subsets for output.

To obtain crisp resulting values as a consequence of defuzzification, one of the existing methods can be applied. There are sufficient number of methods for this step – center

of gravity (COG), mean of maximum (MOM), first of maximum (FOM), largest of maximum (LOM). Despite the fact that the COG and MOM methods are the most frequently used, choosing the actual method to obtain an adequate value can become an independent task that requires systematic approach [9]. The systems built under the aforementioned principle allow assessing translation quality by taking expert opinions into consideration while being a formalized knowledge base.

Another approach that can cope with the previously mentioned problems uses genetic algorithm (GA) that is an analogy to the natural selection of the best adapted units $W(x)$ from the whole range X , which may possess maximal or minimal function values $x^* = \arg \max_{x \in X} W(x)$ and $x^* = \arg \min_{x \in X} W(x)$. Here, at the initial step the population of possible solutions is created $P^1 = \{p_1^1 \dots p_n^1\}$, $p_i^1 \in X$, i.e. initial generation of cardinality $|P^1| = n$. Then, selection S is made with further units crossover C and mutation M to obtain the best suited results

$$P^{k+1} = M \times C \times S(P^k, F^k), \quad (2)$$

where $F^k = \{f_1^k \dots f_n^k\}$, $f_i^k = W(p_i^k)$ is the estimate for the next generation adaptation. The algorithm terminates when a sufficient pair is found [10]. GA is particularly efficient when dealing with word changing in a sequence for obtaining the best translation quality. The question here that arises is connected with a method for the best result estimation.

Quite effective solution can be found using self-training for a neural network approach. Its simplest presentation can be written like this one: $a(x) = \phi(\sum_{j=0}^n w_j x^j)$ which is equivalent to the linear threshold classifier. Though, multilayered network is usually considered. Training any neural network implies using an optimization algorithm to find weights that adjust inputs according to outputs in order to obtain better results for the next calculations. Such a procedure is usually implemented via the method of Gradient Descent. The following formula can be used for tuning weights at each step:

$$w_j = w_j - l_s \partial \frac{L}{\partial w_j}. \quad (3)$$

L here is the loss function defined as the sum of squared differences between the obtained and estimated results

$$L(y, y') = \frac{1}{m} \sum_{i=1}^m (y_i, y'_i)^2. \quad (4)$$

In the formula (3) above, l_s corresponds to the speed of learning per each algorithm iteration, which is multiplied by the weight derivative and subtracted from each weight. The learning speed should be taken appropriately not to trap into the local minimum or jump through it [11]. In the next

subsection we observe the proposed technique that help dealing with the issues stated here.

III. MEASUREMENTS AND COMPARISON

First of all, it should be mentioned that by a term we mean a word sequence that brings particular sense in that specific domain. Thus, domain specific search of a term translation is proposed to be performed with context consideration. As a result of such a search there appears from 2 to nearly 10 (or even more, in case of a long word sequence) variants of translation. Only one variant (or 2-3 at maximum) is truly used to denote that particular subject in the domain under analysis. In order to reduce the number of variants and finally choose the correct solution, it is proposed to use graphical data matching procedure. These graphical data hide the term visual presentation in them. Fig. 2 illustrates the algorithm of a term translation search based on the intelligent image analysis. The part of the algorithm denoted with a dashed line indicates the novel approach that is uncommon for the existing systems that operate with TMs and totally automated systems as well. Due to the proposed image matching procedure for the terms in the source and the target languages, the experts may have only to change the sentence structure and translate grammatical linkages.

For the purpose of this work, images are matched to each other with the help of Voronoi diagrams. This concept is described in [12] where its efficiency has been proved in comparison with the existing approaches. Voronoi diagrams were initially designed for geodesy, they gained their popularity lately only with their application in computer graphics, for 3D modelling in particular [13, 14]. First order Voronoi tessellations are used for the purpose of this work. The order of tessellations defines the grid of image segmentation. This grid is composed on points which should be set a priori. It should be noted that the grid contains convex polygons only.

The Voronoi diagram makes a plate decomposition (image decomposition in our case) according to the following rule: each Voronoi region connected with some key point includes all the points, the distances from which are less to their key point than to the rest key points [15, 16]:

$$v(p_i)^{(k)} = \{z \in R^2 : \max\{d(z, p_{i,h}), p_{i,h} \in v(p_i)^{(k)}\} \leq \min\{d(z, p_{i,j}), p_{i,j} \in V^{(k)} \setminus v(p_i)^{(k)}\}\}, \quad (5)$$

where k is the order of a tessellation $v(p_i)^{(k)}$, $\{\{p_{1,1}, \dots, p_{1,k}\}, \dots, \{p_{l,1}, \dots, p_{l,k}\}\}$ is an area of points that correspond with k key points $\{p_{i,1}, \dots, p_{i,k}\}$ used for the purpose of segmentation. Higher order Voronoi tessellations imply several key points for each region, in other words, each region is linked with several key points in a higher order Voronoi diagram.

The amount of key points defines the order of a diagram [17]. Increase of the diagram order influences change in the level of detailing or segmentation. For nearly maximum values of a diagram order, the detailing decreases, while it increases for small values of the order [18]. It is important to note that the threshold of detailing change into the reverse order depends on the number of key points.

IV. CONCLUSION

It is often hard even for translator experts to find an appropriate word sequence in a target language that accurately conveys meaning of the source segment. The problem of untranslatability is widely depicted in special literature on linguistics. However, the existing linguistic

approaches of calquing, i.e. modeling of words and constructions under the foreign patterns, descriptive and assimilating translation are not covered in computer assisted and fully automatic translation tools because they require totally different level of content understanding. That is why dissatisfaction when using such tools arises in translators' circles.

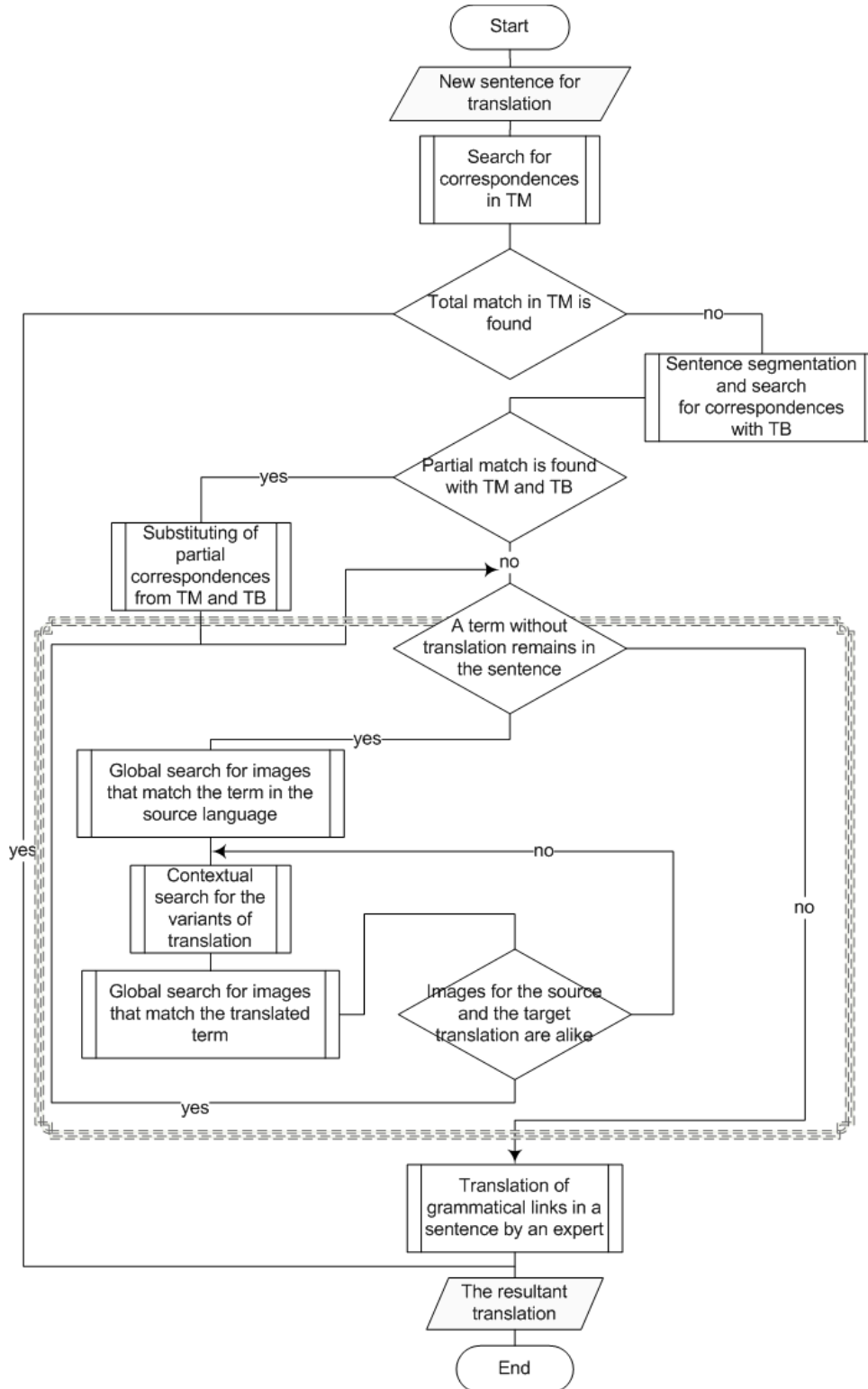


Fig. 2. The proposed algorithm of term translation search based on the intelligent image processing

In this article the authors make an attempt of solving this problem via artificial intelligent methods based on image retrieval. Rule based and genetic algorithms have shown their inefficiency, in contrast to neural networks that potentially can help at a higher level of multimedia content understanding. In our proposed technique we first associate visual objects corresponding with the source segment, and then try finding the same image in target language domain. Thus, boosting translator's experience and even perfecting his or her own language solution. Measuring difference between images is performed with the formerly proposed approach based on high-level features and Voronoi tessellations.

REFERENCES

- [1] E. D. Mihnova, "Oblachnye tehnologii v sistemah na osnove pamyati perevodov", in Zbirnik naukovih prac mizhnarodnoyi naukovo-praktichnoyi Internet-konferenciyi "Elektronni informacijni resursi: stvorenniya, vikoristannya, dostup", Vinnica: VNTU, 2015, pp. 253-258.
- [2] S. K. Abramov, V. V. Lukin, N. N. Ponomarenko, "Mera sodержaniya fona na osnove entropii dlya poiska i sortirovki izobrazhenij v bazah dan-nyh", Radioelektronni i komp'yuterni sistemi, no. 2 (21), pp. 24-28, 2007.
- [3] S. G. Antoshuk, A. A. Nikolenko, E. V. Tkachenko, "Analiz bazisnyh funkcij vevjlet-preobrazovaniya pri multimasshtabnom konturnom pred-stavlenii izobrazhenij", Elektromashinobuduvannya ta elektroobladnannya, vol. 72, pp. 15-19, 2009.
- [4] O. M. Berezkiy, K. M. Berezka, S. Yu. Popina, "Statistichne obroblyennya citologichnih zobrazhen", Visnik Hmelnickogo nacionalnogo universitetu, no. 5, pp. 161-164, 2012.
- [5] R. P. Bogush, G. A. Samoshenkov, "Analiz algoritmov obrabotki dinamicheskikh izobrazhenij v prilozenii k zadache segmentacii dvizheniya na slozhnom fone", in Aktualnye voprosy sovremennoj informatiki : mater. mezhdunar. zaochn. nauchn.-prakt. konf., Kolomna: MGOSGI, 2011, vol. 2, pp. 13-16.
- [6] B. V. Vishnyakov, Yu. V. Vizilter, A. V. Lagutenkov, "Ispolzovanie modifitsirovannogo metoda opticheskikh potokov v zadache obnaruzheniya i mezh-kadrovogo proslezhivaniya dvizhushihya obektov", Vestnik kom-pyuternyh i informacionnyh tehnologij, M.: Spektr, no. 5, pp. 2-8, 2007.
- [7] A. Divakaran, Multimedia content analysis: theory and applications. NY: Springer, 2009, pp. 198-214.
- [8] A. Kutsenko, Y. Megel, S. Kovalenko et al., "An approach to quality evaluation of embryos based on their geometrical parameters," in Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, SPIE, 2019, 11176.
- [9] Y. Megel, S. Kovalenko, S. Kovalenko, and O. Mikhnova, "Pidkhid do kilkisnoyi otsinky vplyvu viznoho turyzmu na ekonomiku krainy," in Information Processing Systems, vol. 3(158), 2019, pp. 65-72
- [10] M. Mitchell, An Introduction to Genetic Algorithms. Cambridge, Massachusetts: MIT Press, 1998, pp. 113-195.
- [11] J. C. Bezdek, J. Keller, R. Krisnapuram, N. R. Pal, Fuzzy models and algorithms for pattern recognition and image processing. NY: Springer, 2005, 516-621 p.
- [12] S. Mashtalir, O. Mikhnova, "Detecting Significant Changes in Image Sequences", Multimedia Forensics and Security, vol. 115, pp. 161-191, 2017.
- [13] H. Ledoux, C. M. Gold, "The 3D Voronoi Diagram: A Tool for the Modelling of Geoscientific Datasets", GéoCongrès, pp. 1-13, 2007.
- [14] V. Cayette, et al., "Modelling and visualisation of fish aggregations using 3D Delaunay triangulation and alpha shapes", in 8-th International Symposium on GIS and Computer Mapping for Coastal Zone Management, Santander, Spain, 2007, pp. 403-413.
- [15] Q. Du, V. Faber, M. Gunzburger, "Centroidal Voronoi tessellations: Applications and algorithms", Society for industrial and applied mathematics review, vol. 41, no. 4, pp. 637-676, 1999.
- [16] F. Hurtado, et al., "The weighted farthest color Voronoi diagram on trees and graphs", Computational Geometry: Theory and Applications, vol. 27, no. 1, pp. 13-26, 2004.
- [17] A. Okabe, et al., Spatial tessellations: Concepts and applications of Voronoi diagrams, 2-nd ed. Chichester: Wiley, 2000, pp. 317-324.
- [18] M. L. Gavrilova, Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence. Berlin: Springer, 2008, pp. 251-269.