ITHEA

International Journal

INFORMATION THEORIES & APPLICATIONS

International Journal
INFORMATION THEORIES & APPLICATIONS
Volume 16 / 2009, Number 3

IJ ITA is official publisher of the scientific papers of the members of
the ITHEA® International Scientific Society

IJ ITA welcomes scientific papers connected with any information theory or its application.

IJ ITA rules for preparing the manuscripts are compulsory.
The rules for the papers for IJ ITA as well as the subscription fees are given on *www.ithea.org* .
The camera-ready copy of the paper should be received by http://ij.ithea.org.
Responsibility for papers published in IJ ITA belongs to authors.

General Sponsor of IJ ITA is the Consortium FOI Bulgaria (www.foibg.com).

ANALOGICAL MAPPING USING SIMILARITY OF BINARY DISTRIBUTED REPRESENTATIONS

Serge V. Slipchenko, Dmitri A. Rachkovskij

Abstract: We develop an approach to analogical reasoning with hierarchically structured descriptions of episodes and situations based on a particular form of vector representations – structure-sensitive sparse binary distributed representations known as code-vectors. We propose distributed representations of analog elements that allow finding correspondence between the elements for implementing analogical mapping, as well as analogical inference, based on similarity of those representations. The proposed methods are investigated using test analogs and the obtained results are as those of known mature analogy models. However, exploiting similarity properties of distributed representations provides a better scaling, enhances the semantic basis of analogs and their elements as well as neurobiological plausibility. The paper also provides a brief survey of analogical reasoning, its models, and representations employed in those models.

Keywords: analogy, analogical mapping, analogical inference, distributed representation, code-vector, reasoning, knowledge bases.

ACM Classification Keywords: I.2 ARTIFICIAL INTELLIGENCE, I.2.4 Knowledge Representation Formalisms and Methods, I.2.6 Learning (Analogies)

Introduction

Example-based reasoning is a powerful approach extensively used by humans for everyday and expert reasoning with incomplete, inaccurate, contradictory input information. For many domains and problems it is essential to rely on examples that include complex relational structures (systems of hierarchical relations). Such information is also present in some knowledge bases and ontologies. Analogical reasoning is a kind of example-based reasoning with special attention paid to the similarity of relational structures, i.e. systems of hierarchical relations between objects, situations, episodes, domains, etc. [Gentner, 1983; Hummel & Holyoak, 1997].

Analogical thinking is one of the most commonly encountered and important cognitive processes. Analogy is closely related to such phenomena as memory, remembering, perception, learning, conceptual change, knowledge formation, similarity, etc. That is why a lot of research is devoted to its study and modeling. For an introduction, see sections Processes of Analogical Reasoning and Models of Analogical Reasoning below, as well as [Holyoak & Thagard, 1989; Hummel & Holyoak, 1997; Markman, 1997; Thagard et al., 1990; Gentner & Markman, 2003; Gentner & Markman, 2000] and references therein. The work of analogy researchers continuously gives new interesting results about the peculiarities of human intellectual activity (for some recent examples, see [Gentner et.al., 2009; Taylor & Hummel, 2009]).

Historically, most advanced models employed symbolic-localist representations, see the Internal Representations and Similarity section below. Such representations are able to deal with the hierarchically structured descriptions of episodes or situations that are used as analogs in analogical reasoning. However, those models have a number of drawbacks that are often attributed to the shortcomings of the symbolic-localist representations. They are computationally complex and so do not scale well, use hand-crafted representations, have poor account of semantics, memory and generalization, are not concerned with neurobiological plausibility.

To some extent, those drawbacks could be overcome by distributed (vector) representations. However, distributed representations were considered incapable to represent structure (for a review, see [Rachkovskij &Kussul 2001] and references therein). Some time ago, structure-sensitive distributed representations have appeared where binding operations are used for structure representation, and have been immediately applied to modeling analogical reasoning [Eliasmith & Thagard, 2001; Hummel & Holyoak, 1997; Kanerva, 1998; Plate, 2003]. Previous attempts to model analogical reasoning with structure-sensitive distributed representations mainly demonstrated "proof of concept": they worked with simple analogies and used distributed representations only at some model stages, see also the Models of Analogical Reasoning section below.

In analogical reasoning, the following basic processes are usually considered and modeled: retrieval (finding in memory most close base analog given a target situation), mapping (finding correspondences between the elements of two analogs), and inference (knowledge transfer from base analog to target). In this paper, we describe a particular scheme for distributed representations and apply it to modeling of analogical mapping and inference. Mapping is established on the basis of similarity of distributed representations. We show that the distributed representation scheme of and its application to analogical reasoning are simple in use, have low computational complexity, and provide the same result as advanced traditional models.

The rest of the paper is organized as follows. In the first three sections we provide a brief survey of the analogical reasoning processes, local and distributed internal representations and similarity, and the models of analogy. Then our representational scheme is introduced and representations of analog elements are described. In the next sections we describe mapping and inference techniques on the basis of the introduced representations, and study them in the experiments. Future research directions are outlined together with the discussion.

Processes in Analogical Reasoning

Analogical reasoning theories deal with the following processes [Gentner, 1983; Kokinov & French, 2003; Falkenhainer et.al., 1989; Gentner & Markman 2003; Thagard et.al. 1990; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997].

Building representations. Analogs are considered as hierarchically structured descriptions of situations and domains.
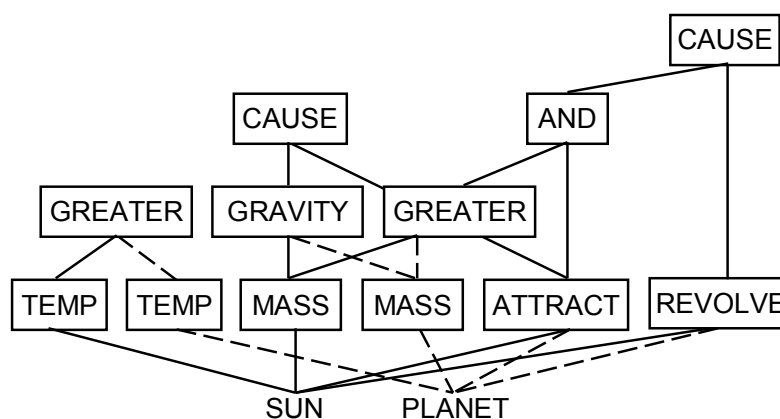

Fig. 1. Graph sketch description of the Solar System analog. Re-representation
code-vector of the Sun is constructed from the roles marked by the solid lines

Descriptions. Analog descriptions (Fig. 1) include descriptions of their elements – entities, attributes and relationships. Objects are entities of subject domains (for example, the Sun, Planet, etc.). Attributes describe

properties of the objects (for example, mass, temperature, etc.). Relations determine relationships between the elements of analogs (for example, attract, more, cause, etc.). Arguments of relations may be objects, attributes and other relations. Attributes may be considered as relations with single argument, so thereafter we will mainly speak about relations.

Are such commonly used analog descriptions appropriate for modeling human analogical reasoning and using in artificial reasoning systems? Probably, only to some degree, – if one considers a rather conservative expert estimate of recent progress in both computer reasoning systems and models of human analogical reasoning. The reason may be explained by the fact that nobody knows how such information is represented in human brain whose ability to use analogies we are trying to reproduce.

Humans describe analogs using natural language or pictures, whether in everyday life or participating in experiments. The usual way to convert verbal or visual descriptions to the formal descriptions of analogs (like in Fig. 1) is to do it manually, by human-expert. The resulting formalized description itself lacks a lot of information present in the initial "natural" descriptions. On the other hand, the formalized descriptions carry a lot of information for humans (especially for the experts who work in this area and encode those descriptions) because of the knowledge associated with the formal description in their brains. However, the system or model dealing with the description lacks such common sense knowledge, at least at the level comparable to a human. Also, the process of manual construction of the formalized description is very laborious and expensive.

So, the question is how to form at least those formalized descriptions, thus overcoming, to some degree, the knowledge acquisition bottleneck. Current efforts in this direction are dedicated to the automation of this process. [Mostek, Forbus, & Meverden, 2000] consider techniques for automatically constructing representations of analogs from knowledge bases. This work is very interesting and allows a more efficient use of knowledge bases and reasoning. However, a much more essential and difficult problem is automatic knowledge acquisition from texts.

The manual transformation of textual descriptions to the formal analog descriptions by humans-experts exploits their previous knowledge of language, of the world, of the particular analogy, of the task to be solved by analogy, and, often, of the analogy model that will work with the produced formalized description. All this knowledge is lacking in automatic systems, making the problem very challenging – one of the Holy Grails of AI. Knowledge discovery in texts or Internet became a hot topic in 90s [Hahn & Schnattinger, 1998; Schubert, 2002; Agichtein & Gravano, 2000]. At present, some progress has been made in extracting simple objects and relations, but not complex episodes. Or, a human editorial staff is used to validate knowledge as part of the process. Recent advances in this direction are known as Learning by Reading [Bobrow et al., 2009; Forbus et al., 2007; Kim & Porter, 2009], but there still is insufficient progress in the formation of the episodes at the knowledge base level of complexity. [Forbus et al., 2008] consider generating analogical representations from sketches.

Another important question is how to enhance the semantic basis of representations. Some approaches exist that create representations of objects from similarity data obtained in experiments, e.g. with humans [Navarro & Lee, 2003; Shepard, 1962; Tenenbaum, 1996]. Another approach is to form representations as "context vectors" using frequencies of joint occurrences in text corpus (e.g., [Ramscar & Yarlett, 2003; Misuno, Rachkovskij, & Slipchenko, 2005; Sahlgren, Holst, & Kanerva, 2008]. However, those approaches have several shortcomings. The representations produced may be still far from those of the brain. Also, context-dependency of representations is not considered within those approach, otherwise that single representation set is produced per experimental domain or text corpus. So far, mainly simple vector representations are formed, much simpler than the episodes used in analogical research.

Once the formalized descriptions are formed, they are usually used in the models of analogy without associated knowledge. Attempts are made to eliminate this drawback in the models of analogy as follows. WordNet is used

in [Thagard et al., 1990; Holyoak & Thagard, 1989] to estimate similarity between the entities. [Forbus et al., 2008] use Cyc for generating representations and reasoning in the domain of interactive sketch understanding. [Forbus, Mostek, & Ferguson, 2002] consider integration of analogical and first-principles reasoning.

Internal representations. Given input descriptions of analogs, they should be represented in the internal format. For human analogical reasoning, those are internal brain representations still unknown to us. For modeling, those are internal models representations. The purpose of this transformation to internal representation can be: compatibility with the format and methods of model; performance issues; a more adequate representation of some aspects poorly supported by the input description (e.g., semantic aspects).

Internal representations may differ from the input descriptions in a varying degree. For example, the relational descriptions of analogs mentioned above (Fig. 1) can immediately be used in the symbolic models. Or, they can be rather straightforwardly transformed to localist connectionist representations. Transformations to fully distributed representations are more radical (see the Internal Representations and Similarity section below). Re-representation may also occur in the process of model functioning (see the end of this section).

Retrieval and memory. When the analogical episodes are represented internally, and the target episode is given at the input, the task is to find the closest analog. Usually retrieval models find the closest analog by similarity (see [Rachkovskij & Slipchenko, under revision] and references therein). Both semantic/featural content and relational/propositional organization of analogs are important for the estimation of similarity, though the structural similarity in analogical retrieval is considered less important than in mapping. Retrieval returns the closest analog, or several candidates. Retrieval models are usually computationally expensive, since they some manner handle the whole base of known analogs (albeit with optimizations).

The issue related to retrieval is where and how the episodes are stored. In humans, the obvious answer is long term memory. Most of the analogical retrieval models ignore neurobiological aspects of memory and store isolated representations of episodes. Those representations do not change in the process of storage and recall retrieval.

Few analogical models consider generalization [Kuehne et al., 2000; Hummel & Holyoak, 2003], but again the generalized representations are stored separately and memory processes are not considered. We believe that a very promising direction is to study storage, retrieval, and generalization in a distributed associative memory [Frolov, Rachkovskij, & Husek 2002, Frolov et al., 2007]. However, we know of no work in this direction, except for [Kokinov & Petrov, 2001] that uses a kind of associative memory to context-sensitive reconstruction of the stored episodes in interaction with mapping.

Mapping. When both the base and the target analogs are known, the task is to find their corresponding elements. This is performed by mapping that is considered as the most important part of analogical reasoning. Mapping pays more attention to structure similarity, than retrieval (see also the Models of Analogy section). Structural constraints on mapping include parallel connectivity and one-to-one mapping [Markman & Gentner, 2000]. Parallel connectivity requires the arguments of corresponding predicates to be placed into correspondence. One-to-one mappings limit any element in one analog to correspond to at most one element in the other analog.

Inference. After mapping is established, knowledge from the base analog has to be transferred to the target. As an outcome of the mapping process, we obtain correspondences between the elements of analogs. So, we observe relations present in the base but absent in the target. They are the candidate inferences about the target analog projected from the base.

To make inferences, copying with substitution and generation is usually used [Holyoak, Novick, & Melz, 1994; Markman, 1997]. It takes any element in the target for which there is a corresponding element in the base and copies to the target all relational structure connected to the element in the base. In this structure, the elements

copied from the base are replaced by their corresponding elements of the target. If some entities are absent in the target but are required to complete the structure from the base, they are generated anew. All this looks as a symbolic operation, and hardly lends itself to a more natural implementations. Most of the models use some kind of implementation of this process for analogical inference [Hummel & Holyoak, 1997; Hummel & Holyoak, 2003].

Because analogy is not a deductive mechanism, these candidate inferences are only hypotheses and must be evaluated and checked (see e.g. [Gentner & Colhoun, 2010] and references therein).

Re-representation. In the process of analogy-making, some modelers of analogical reasoning allow re-representation of initial internal representations – e.g., when the initial representations are not adequate for the task [Yan, Forbus & Gentner, 2003]). In [Rachkovskij, 2004] identical representations were substituted for some mapped analogical elements in order to ensure the correct mapping of the rest of the episode. [Yan, Forbus, & Gentner, 2003] proposed a theory of re-representation in analogical mapping. They consider the problems of detecting opportunities for re-representation, generating re-representation suggestions based on libraries of general methods, and controlling the re-representation process. In "fluid analogy" models ([Kokinov & Petrov, 2001; Hofstadter, 1995; French, 1995], see also the Models of Analogy section) representations are formed depending on the context of particular episodes, and representation construction and mapping proceed in parallel and are inseparable.

As we have seen, representation and similarity are the most important factors of analogy. Let us consider them in some more detail.

Internal representations and similarity

Implementation and efficiency of operations on relational structures, which are required by the models of analogical reasoning and other cognitive models, depend essentially upon the representation scheme employed. Estimation of similarity of representations is one of the most important operations with the models. Retrieval of relevant analogs is mainly similarity-based. Mapping can also be considered as finding and purring into correspondence the most similar elements of analogs.

Processing of analogies requires taking into account both types of representation and similarity – structural and semantic. Structural similarity reflects how the elements of analogs are arranged with respect to each other. It is based on the notion of "structural consistency" [Falkenhainer et al., 1989; Gentner & Markman, 2003] or "isomorphism" [Thagard et al., 1990; Hummel & Holyoak, 1997; Eliasmith & Thagard, 2001]. Analogs are also matched by the "surface" or "superficial similarity" [Gentner 1983, Forbus et al., 1995] based on common analog's elements or a broader "semantic similarity" [Thagard et al., 1990; Hummel & Holyoak, 1997; Eliasmith & Thagard, 2001], based on some prior semantic similarity (e.g., joint membership in a taxonomic category) of those elements.

In early similarity models, objects or some other items were represented as points in a space of some dimensions [Shepard, 1962], or by feature sets [Tversky, 1977]. Similarity was estimated as the distance between those points or by the set-theoretical operations. Note that both mentioned representations can be considered as vector ones, with gradual or binary components, and so similarity of items can readily be calculated by some combination of vector(dis)similarity measures (dot product, Euclidean distance, etc.). Since the early naive versions of the vector representations did not represent relations, they were considered incompatible with the inherent structure of real world objects and situations.

Instead of repairing those drawbacks of naive vector representations, novel emerged similarity accounts were based on structured representations. They used graph-like representations, where elements of varying complexity are represented as vertices of the graph (Fig. 1). Edges provide information about the structural

organization of elements (how the elements are grouped or bound together). However, processing of graphs is complicated compared to vectors, and semantic, feature similarity of elements is rather difficult to take into account when processing graphs.

We think that the drawbacks of both feature/vector and structure/graph similarity models and representations stem from their being instances of symbolic and localist representations [Thorpe, 2003; Page, 2000]. In such representations, each element, such as a feature or attribute, entity, relation, episode, etc., has a representation in the form of symbol, node, or single vector component. Each new item or combination of items requires allocating a new node or memory location. Such representations are semantically "brittle" (all-or-none similarity). It is necessary to decide for each input item, whether it is "sufficiently" similar to (and so can be encoded by) some existing symbol or node, or requires a new representation. Thus, the system's knowledge does not generalize naturally to newly represented situations. Also, symbol or higher-level node is just a label, and so it does not carry immediate information about the structural organization or semantics of its elements. Sequential link or pointer following is needed to traverse the structure corresponding to the vertex, and to retrieve its semantic description, if any.

Comparison of structures is difficult, since it requires alignment and finding correspondences between element substructures. Popular graph similarity measures based on sub-graph isomorphism are very expensive computationally (NP-complete) and so practically inapplicable to large graphs. However, isomorphism does not take into account the semantic similarity of objects and does not meet the requirements that humans apply to analogical similarity [Markman & Gentner, 2000]. So, some heuristic measures are proposed in the models instead. Nevertheless, for localist and symbolic representations, an estimation of structure similarity requires computationally very expensive procedures that include construction and processing of virtual or real constraint-satisfaction networks (as in [Falkenhainer et al., 1989; Holyoak & Thagard, 1989; Keane et al., 1994], see also the Models of Analogical Reasoning section below).

Here, we are especially interested in fully distributed representations [Thorpe, 2003], where any item (e.g., feature or attribute, entity, relation, analog, etc.) is represented as a vector of fixed dimensionality. Each vector component does not have clear semantics, i.e. does not correspond to single item, in contrast to localist representations. It is natural to represent similar items by correlated distributed representations and to estimate their similarity by dot product. A high information capacity is provided by the possibility to represent exponentially many items by different vectors of the same dimensionality N (e.g., (M from N) binary vectors vs N for localist representations with single component per item).

However, many thought that distributed representations cannot represent nested (recursive) structures because of the superposition catastrophe (losing the information concerning item arrangements in structures, see [Rachkovskij & Kussul, 2001] for discussion and references). In fact, a number of connectionist schemes have been proposed capable of forming recursive structured representations (e.g., RAAMs [Pollack, 1990], tensor products [Smolensky, 1990], etc. – for a review and comparison, see [Plate, 2000, 2003; Browne & Sun, 2001; Rachkovskij & Kussul, 2001] and references therein). In those schemes, binding procedures (an analog of grouping brackets in symbolic representations) were proposed to avoid the superposition catastrophe. In RAAMs [Pollack, 1990], binding is realized using the weight matrix formed by training a multilayer perceptron. This requires a challenging retraining using all the items when new items to be bound are introduced. In tensor products [Smolensky, 1990], the dimensionality of the resulting tensor grows with the number of bound code-vectors. In dynamic bindings [Shastri & Ajjanagadde, 1993; Hummel & Holyoak, 1997], the representations are bound by synchronous activation, but vector measures of similarity cannot be immediately employed for resulting dynamic representations.

Some time ago, a number of schemes for structure-sensitive fully distributed representations with vectors of various formats have appeared, where binding operations that does not change dimensionality are used and bindings are produced on-the-fly (without any training). Holographic Reduced Representations (HRRs) of [Plate, 2003] use real-valued vectors and circular convolution for binding. Binary Spatter Codes (BSCs) of [Kanerva, 1996] use Boolean vectors and component-wise exclusive-or. [Gayler, 1998] uses vectors with components from {−1, +1} and component-wise multiplication. Associative-Projective Neural Networks (APNNs) use sparse binary vectors with {0,1} components and special context-dependent thinning (combination of vector component-wise AND, OR, and permutations) procedure for binding (Rachkovskij and Kussul 2001, see also this paper below). For a more detailed comparison of those binding schemes, see [Browne & Sun, 2001; Rachkovskij, 2001; Rachkovskij & Kussul, 2001].

Similarity of resulting bound representations is influenced both by the set of components and their arrangements. Thus, similar structures are encoded by similar code-vectors. So, it is not necessary to search for the match between the elements of two structures in order to estimate their overall similarity by dot product. As discussed above, this property is important for many cognitive models and systems, including those of analogical reasoning considered in the following section.

Models of analogical reasoning

Let us consider models of analogical reasoning that employ different internal representations and methods of their processing, with the emphasis on mapping models. Since, till recently, representation and processing of structures were achievable only with hierarchical symbolic or localist representations, it is natural that those representations are used in the most influential computational models of analogy.

Symbolic and localist models. Structure Mapping Theory (SMT) [Gentner, 1983] is probably the first and best-known theory of analogical reasoning that explicitly underlined the importance of structure similarity defined as the common systems of relations between analogs. In this theory, the processes involved in analogy-making – representation-building, retrieval, mapping, etc. – are separated from one another. The model of mapping in SMT is instantiated by Structure Mapping Engine (SME) [Falkenhainer et al., 1989].

At the input, SME takes propositional representations of two analogs (symbolic version of representation of the type shown in Fig.1). It uses a local-to-global alignment process to determine correspondences between their elements as follows. First, SME finds all possible local correspondences between the elements of the two representations by putting into correspondence identical relations, as well as the arguments of identical relations. Then, SME produces one or few globally consistent interpretations by integrating consistent local correspondences.

All interpretations generated by SME strictly impose the structural constraints of parallel connectivity and one-to-one mapping ([Markman & Gentner, 2000] and the Processes in Analogical Reasoning section above). Finally, SME calculates scores reflecting the quality of each interpretation using an algorithm that favors mappings preserving interconnected, higher-order relational structure (systematicity principle). The interpretation with the highest score is selected as the preferred interpretation.

SME's drawback, as a symbolic model, is a rather poor account of semantic similarity. Also, SME structure matching is computationally expensive. It makes prohibitive using it during retrieval for a structure-sensitive comparison of the input to each of (many) potential analogs stored in memory. So, SMT-based model of retrieval, MAC/FAC [Forbus et al., 1995] uses a two-stage process. The first stage MAC uses feature vectors of analogs and a computationally cheap vector similarity measures to select the candidates based on the surface similarity

only. The second stage FAC uses SME to take structure of the candidates into account. The drawback of MAC/FAC is that top-scored candidates of MAC may miss some useful analogs.

The Analogical Constraint Mapping Engine (ACME; [Holyoak & Thagard, 1989] is a localist connectionist model that determines analogical mappings using a parallel constraint satisfaction network. ACME takes the input representations of two analogs and constructs a network where each node corresponds to a possible match of analog elements, and the between-node connections represent the constraints to be satisfied. Inhibitory connections are used between competing hypotheses, while excitatory connections are used between consistent hypotheses. A spreading-activation relaxation algorithm runs until the network settles to a final state. In this state, active nodes represent winning matches corresponding to the maximally consistent mapping hypothesis.

Unlike SME, ACME relies not only on structural information, but also takes into account semantic and pragmatic constraints. Unlike strict constraints posited by SME, all those types of constraints are soft and together drive the system to decision. Pragmatic and semantic constraints, as well as representations of analogs, are hand crafted. ACME can map structures that do not involve semantic similarities, and this may be considered as its shortcoming. Also, it can produce mappings that violate the structural constraint of one-to-one correspondence, that can lead to inconsistencies in analogical inferences [Markman, 1997]. Usually, it is even more computationally expensive than SME. So, the retrieval model ARCS [Thagard et al., 1990] is a two-stage scheme, as well as MAC/FAC. It uses ACME, but with greater emphasis on semantic constraints.

Connectionist Analogy Builder (CAB) [Larkey & Love, 2003] is a localist connectionist model that takes as input two directed graphs and determines mappings via a dynamic process of interactive activation among correspondences that are represented as network nodes. Nodes are excited by other nodes that represent parallel correspondences. Nodes also compete for excitation and are inhibited by other nodes. Resulting activations establish one-to-one mappings. CAB makes time-course predictions as well as predictions concerning the role of working memory in the comparison process. Besides the common shortcomings of all localist models, its specific drawbacks are not clear.

The Incremental Analogy Machine (IAM) [Keane, Ledgeway, & Duff, 1994] is an incremental model of analogical mapping that takes into account the effects of the order of material presentation. A problem with IAM is that people's comparisons are incremental, but in a different way that in IAM [Larkey & Love, 2003]. Another limitation of IAM is that the fully serial nature of its processing prevents scaling up. Also, IAM can map unnatural analogies that do not involve semantic commonalities.

In solving analytical problems by analogy, Gladun and his colleagues [Gladun, 2000; Gladun et al., 2000] used structural-attributive models. They employ localist connectionist representations, such as growing pyramidal networks, where generalization of analogs is used as a prerequisite for inference. Classification of new objects is implemented by comparison of their representations to the class concepts. The model is mainly elaborated for classification tasks, where only the class label is transferred, not relational systems.

Unlike the models discussed above, the models such as CopyCat [Hofstadter & Mitchell, 1994; Hofstadter, 1995], Tabletop [French, 1995], Associative Memory Based Reasoning (AMBR) [Kokinov, 1988; Kokinov, 1994; Kokinov & Petrov, 2001] consider representation-building and reasoning as parallel, interacting sub-processes. In the process of analogy-making, the interactions between micro-agents construct in working memory context-sensitive representations of the episodes using semantic knowledge in memory and the current problem task. Elements of analogs and the whole episodes are represented by coalitions of agents that may be considered as a kind of distributed representation.

Copycat solves proportional letter-string analogies such as "abc is to abd as mrrjjj is to ?" (most people answer either mrrkkk, mrrjjk, or mrrjjjj). It gradually builds its own representations and simultaneously explores different interpretations of analogy. The drawback is that Copycat does not provide a domain-general account of analogy.

Each new micro-domain requires encoding new system knowledge. The model extension to more complex types of analogies is also unclear.

AMBR integrates retrieval, mapping and transfer. These processes run in parallel and can influence each other. In the system's long-term memory coalitions of agents allow for distributed representations of objects, concepts and episodes. Mapping is performed by gradually building and relaxing a constraint satisfaction network that is built incrementally and in a distributed way by the independent operation of many agents which make their decisions using only local information. The AMBR model demonstrates insertions from general memory and blending of episodes, priming order and context effects. Some of those effects were met in the experiments with humans.

Analogy models including distributed internal representations. The quest to enhance a semantic basis of representations, their scaling, and degree of neurological relevancy, led to the attempts to augment the models of analogy with some share of distributed representations.

Learning and Inference with Schemas and Analogies (LISA) is an integrated theory of analogical access and mapping [Hummel & Holyoak, 1997, 2003], as well as similarity [Taylor & Hummel, 2009]. It is a connectionist model that stresses the importance of working memory capacity limitations. Elements of analogy are represented in long-term memory as hierarchies of localist nodes. However, entities are semi-distributed and represented by micro-features of the bottom level semantic nodes. The more similar two entities are - the more semantic nodes they share. Those distributed representations come to work in working memory. Structure sensitivity is achieved through dynamic binding by synchronous oscillation of distributed representations. LISA posits limitations on the number of different oscillating activation patterns that can be simultaneously active in working memory, thus modeling its span. The drawback of LISA is its limited ability to scale up to human performance in processing analogies involving large representations.

STAR2 [Gray et al., 1997] uses the tensor-product representations to represent and to store relational structures of analogs. Tensor representations [Smolensky, 1990] can potentially represent items in fully distributed fashion. Their dimensionality grows exponentially vs the number of relational arguments that is used by the authors to limit complexity of reasoning according to the pattern demonstrated by humans. For mapping, localist constraint satisfaction networks are built and evolve in parallel, similar to ACME. Sequential focusing on different parts of analogs allows working with more complex analogies than proportional analogies of the previous model versions.

 Distributed Representation Analogy MApper (DRAMA) [Eliasmith & Thagard, 2001] uses distributed representations of analogs – HRRs of [Plate, 2003]. However, those representations are only used at the first stage to construct and initialize a "semantically driven mapping network" based on ACME that then produces the mappings.

So, application of distributed representations in the described models is not systematic. At some stage, they use structure handling techniques conceptually similar to localist networks. Emerged structure-sensitive fully distributed representations mentioned in the Internal Representations and Similarity section have been applied to modeling of analogical retrieval and mapping.

HRRs [Plate, 2000; 2003] and APNNs (Rachkovskij, 2001; Rachkovskij & Slipchenko, under revision] were tested on the analogical retrieval tasks. It was shown that for analogical episodes with different similarity types of (see the Experiments section below), the retrieval results obtained by vector similarity measures are consistent with the experimental results observed for people and modeling results reported for MAC/FAC and ARCS (where different episodes were used). In [Rachkovskij & Slipchenko, under revision], a modified representation scheme was proposed that provided results comparable to MAC/FAC using MAC/FAC's knowledge base of analogs.

The models of mapping based on HRRs [Plate, 2000; 2003] and BSCs [Kanerva, 1998; 2009] were proposed that used the technique based on "unbinding" (binding with the inverse of a distributed representation) operations. In [Rachkovskij, 2001], similarity of distributed APNN representations of analogical elements was used for their mapping. However, those techniques worked only for the most straightforward mappings cases. In [Rachkovskij, 2004], a number of approaches for mapping with the APNN distributed representations were proposed (including direct similarity mapping, re-representation by substitution of identical code-vector, parallel traversing of structures, using higher-level roles) and some of them were demonstrated on complex analogies. However, the methods were rather complex and used sequential operations. In [Gayler & Levi, 2009] an interesting scheme was proposed for finding graph isomorphism with HRRs and associative memory.

In the rest of the paper, we consider the scheme for binary distributed representation of analog elements that allows a simple mapping by similarity of rather difficult and complex analogies and consider mapping and inference using this scheme.

Distributed representations of relational structures

Let us consider APNN-style of analogs. Each item x (element of analog – attribute, object, relation) is represented by a code-vector X ($x \rightarrow X$). Code-vector is a form of vector representation that is binary ($X \in \{0,1\}^M$) and sparse (the fraction of non-zero vector components M in code-vector X with dimensionality N is small: $M/N \ll 1$). Similar items (in context of the application problem) should have similar code-vectors, whereas items with undefined similarity should have dissimilar code-vectors.

Similarity of items $sim(x,y)$ is estimated by similarity of their code-vectors $sim(X, Y)$. Code-vector similarity is defined using dot product, which for binary code-vectors is equal to the number of common unit components ("1s"). Similarity $sim(X, Y)$ is a relative overlap of code-vector:

$$sim(x,y) = sim(X, Y) = \sum_{i=1,N} X_i Y_i / \sum_{i=1,N} X_i = |X \wedge Y| / |X|, \qquad (1)$$

where X_i is a component of X, $\wedge$ is component-wise conjunction, $|Z|$ is the number of non-zero components in Z.

We consider relational structures of analogs or episodes in declarative knowledge bases as labeled directed acyclic graph [Frasconi, Gori, & Sperduti, 1998], where child vertices (arguments) of parent vertices (relations) can be objects (entities) or relations (Fig. 1). Previously [Rachkovskij & Kussul, 2001; Rachkovskij 2001], we have proposed schemes for representation of relations $R(A,B,....)$, (where R is the label of relation, $A,B....$ are its arguments), that corresponds to the *role-filler* schemes traditionally used in symbolic representations. In the scheme we use bellow, code-vector of the relation is formed as $R(A,B,....) \rightarrow \langle \langle R_a, A \rangle, \langle R_o, B \rangle, ... \rangle$, where A, B, ... are code-vectors of the arguments (fillers), R_a, R_o, ... are code-vectors of the roles (agent, object,...), and $\langle ... \rangle$ denote the binding procedure for code-vectors.

We consider the binding procedure as a functional analog of grouping brackets traditionally used in symbolic notation. In the examples of relation given above, $\langle R_a, A \rangle$ should preserve the information that R_a is bound with A, not with B or R_o. We will use the Context Dependent Thinning (CDT) binding procedure described in [Rachkovskij & Kussul, 2001]. It is implemented as follows.

First, code-vector Z is formed as component-wise disjunction $\vee$ of element code-vectors X_i:

$$Z = \vee_i X_i, \qquad (2)$$

For the example above, $Z = R_a \vee A$ (and also another $Z = R_o \vee B$). Then the result $\langle Z \rangle$ of binding $X_1, X_2 ,..., X_S$ (for the example above, $\langle Z \rangle = \langle R_a \vee A \rangle$ as well as $\langle Z \rangle = \langle R_o \vee B \rangle$) is formed as

$$\langle Z \rangle = Z \wedge \vee_{k=1,K} Z^*(k). \qquad (3)$$

Here $\wedge$ is component-wise conjunction, $Z^*(k)$ is Z with permuted components. For each k, random independent permutation is used, fixed for this k (there can be versions with single fixed permutation iteratively applied to Z^* [Kussul et.al., 2006]. The number of 1s in $\langle Z \rangle$ is controlled by K. Dimensionality of $\langle Z \rangle$ and X_i is the same. Subset of 1s of each component code-vector X_i preserved in $\langle Z \rangle$ depends on Z and therefore on each and all X_i, thus preserving information on the particular subset of elements that produced it, and so providing binding property. Also, similar code-vectors X_i produce similar $\langle Z \rangle$.

So, the dimensionality of all code-vectors (roles, fillers, relations) in the considered representation scheme is the same. Therefore, its recursive application allows forming code-vectors of episodes with hierarchical relational structures, where higher order relations (relations over relations) are present.

To construct the code-vectors of episode's elements and episode itself, it is necessary to have the code-vectors of its terminal elements, such as roles, attributes, names, labels, constants. In this paper, randomly and independently generated and memorized code-vectors are used for the terminal elements. The same code-vector is always used for the initial representation of any occurrence of the particular terminal element. From those code-vectors, other elements are constructed using the role-filler scheme described above.

Note that the code-vector of the whole episode is just the disjunction of code-vectors of its top-level relations (i.e., those relations that are not arguments of other relations), see Fig 2.

$$
\begin{aligned}
&\textbf{SOLAR\_SYSTEM} = \\
&\langle\ \textbf{CAUSE\_1} \vee \langle \textbf{GRAVITY\_1} \vee \langle \textbf{MASS} \vee \textbf{SUN} \rangle\ \rangle \vee \langle \textbf{GRAVITY\_2} \vee \langle \textbf{MASS} \vee \textbf{PLANET} \rangle\rangle\ \rangle \\
&\vee \langle \textbf{CAUSE\_2} \vee \langle \textbf{ATTRACTS\_1} \vee \textbf{SUN} \rangle \vee \langle \textbf{ATTRACTS\_2} \vee \textbf{PLANET} \rangle\ \rangle \\
&\vee \\
&\langle\ \textbf{GREATER\_1} \vee \langle \textbf{TEMPERATURE} \vee \textbf{SUN}\ \rangle\rangle \\
&\vee \langle\ \textbf{GREATER\_2} \vee \langle \textbf{TEMPERATURE} \vee \textbf{PLANET} \rangle\ \rangle \\
&\vee \\
&\langle\ \textbf{CAUSE\_1} \vee \\
&\quad\langle\ \textbf{AND} \vee \langle \textbf{GREATER\_1} \vee \langle \textbf{MASS} \vee \textbf{SUN} \rangle\rangle \vee \langle \textbf{GREATER\_2} \vee \langle \textbf{MASS} \vee \textbf{PLANET} \rangle\rangle\ \rangle \\
&\quad\ \vee \langle\ \textbf{AND} \vee \langle \textbf{ATTRACTS\_1} \vee \textbf{SUN} \rangle \vee \langle \textbf{ATTRACTS\_2} \vee \textbf{PLANET} \rangle\rangle\ \rangle \\
&\vee \langle\ \textbf{CAUSE\_2} \vee \\
&\quad\langle\ \textbf{REVOLVE-AROUND\_1} \vee \textbf{PLANET} \rangle \vee \langle \textbf{REVOLVE-AROUND\_2} \vee \textbf{SUN} \rangle\ \rangle
\end{aligned}
$$

Fig. 2. The role-filler code-vector representation of the Solar System analogical episode

If different episodes include some identical elements, and therefore identical code-vectors, the code-vectors of the episodes will be similar. The more identical relations and their arguments the episodes have, the more is the similarity of their code-vectors. Note also, that this approach allows the episode code-vectors to be similar not only if they include identical code-vectors, but also for similar code-vectors of similar elements, such as, e.g., produced as context code-vectors mentioned in section Processes in Analogical Reasoning above.

The role-filler representation scheme has been used earlier both for analogical retrieval [Rachkovskij, 2001; Rachkovskij & Slipchenko, under revision] and mapping [Rachkovskij, 2001, 2004]. For analogical retrieval by similarity of code-vector episodes as wholes, the representation schemes of this type appeared adequate. However, [Rachkovskij, 2001, 2004] showed that mapping using the direct similarity of code-vectors of analog's elements obtained by this scheme is not always correct This is due to the fact that in the simple role-filler scheme the element codevecor includes only code-vectors of its sub-elements. However, the mapping also depends, and to a greater degree, on the similarity of relational systems to which the mapped elements belong.

So, in order to develop mapping techniques based on the similarity of element code-vectors, we developed a scheme for re-representation of analog's elements (based on the ideas from [Rachkovskij, 2004; Slipchenko & Rachkovskij, 2009]). The re-representation idea we employ consists in combining the element code-vector

obtained by the simple role-filler scheme described above, with the code-vectors of the higher-levels roles in that element. Thus, using the code-vectors of analog elements, generated using the role-filler scheme (that are adequate for retrieval and always the same for identical elements), for mapping we construct re-representations that depend on the episode.

Re-representation code-vector Cx* of analog element *x* is the combination of its "lower" representation, that is the code-vector CxLower of *x* generated by the role-filler scheme, and its "higher" representation CxHigher, that contains the information about the roles r in the higher-level relations to which x belongs (immediately as their filler, or recursively through other higher level relations):

$$Cx^* = CxLower \vee CxHigher, \quad CxHigher = \vee_r C_r. \tag{4}$$

Here C_r may be the role code-vector itself, or binding with the initial role-filler code-vector of *x*, i.e., CxLower.

Example of the roles used for the re-representation of the Sun element of the Solar System – Atom analogy is shown in Fig. 1 by the solid lines,. and the re-representation code-vector is formed as follows (repeated code-vectors are omitted):

$$SUN^* = SUN \vee TEMPERATURE \vee MASS \vee ATTRACT\_1 \vee REVOLVE\_1 \vee GREATER\_1 \vee \tag{5}$$
$$GRAVITY\_1 \vee GREATER\_1 \vee CAUSE\_1 \vee CAUSE\_2 \vee AND.$$

Depending on implementation, here RELATION_1 means RELATIONagent or ⟨RELATIONagent ∨ ELEMENT⟩.

Mapping and inference techniques

Mapping. Mapping using the re-representation code-vectors is done as follows.

Step 1. For each top-level relation $t \in T_{top}$ (i.e., those relations that are not arguments of other relations) of the target analog *T*, the best mapping b'(t) is found by the maximal similarity between its code-vector C_t^* and code-vectors C_b^* of the elements of the base analog $b \in B$:

$$b'(t) = \text{argmax}_{b \in B} \; sim(C_t^*, C_b^*). \tag{6}$$

Step 2. Taking into account the parallel connectivity constraint (the arguments of corresponding relations must correspond), for each element pair mapped at step 1 similarity of their sub-elements of all levels is calculation as follows. All pairs (ti, bj) of children of the mapped elements are considered, where ti $\in$ T are children of the target element t, bi $\in$ B are children of the base element b, and similarity of their re-representation code-vectors (4) is calculated by (1). All pairs whose similarity is less that the threshold value (i.e. similarity of random code-vectors with the same number of 1s) are ignored. So, at the output we get the list of remained triples (t, b, sim(t,b)), where x and y are elements of the target and the base, correspondingly, and sim is their similarity.

Note that since each element can belong to different top-level relations identified at step 1, there could be multiple triples (*t*, *b*, *sim(t,b)*) for the same pair (*t*, *b*).

Step 3. For each element of the target t the corresponding mapping b' is obtained as follows.

$$b'(t) = \text{argmax}_{b \in B} \; \Sigma \, sim(t,b), \tag{7}$$

where summation stands for the triples that occur several times after step 2.

Note that strict one-to-one mapping can be achieved, if necessary, by elimination of the already mapped elements with maximal similarity.

This mapping technique takes into account both the similarities of the elements themselves reflected in the similarity of their corresponding code-vector representations ("lower" ones), as well as the similarity of roles of those elements in relations or attributes to which the elements belong ("higher" code-vector representations).

Analogical Inference. As many analogical reasoning models, we consider inference as an extension of mapping and use copying with substitution and generation (section Processes in Analogical Reasoning and [Markman 1997]) for its implementation. The inference is realized based on the mapping results. Specifically, the elements of the base which have not been mapped are identified and transferred to the target. Those are hypotheses, the quality of which is evaluated as follows. For each element transferred from the base, starting from the highest-level elements, similarity of its code-vector is calculated to both the code-vectors of the target sub-elements (i.e., child elements or arguments or fillers) of its transferred mapping hypotheses and to the code-vector of the entire target episode. The code-vectors used here are the "lower" ones, i.e., the initial role-filler representations. If the similarity to the code-vector of the whole episode is larger than the similarity to each of its sub-elements, then the hypothesis is accepted, otherwise it is rejected. The same procedure is repeated for the lower-level hypotheses. Computational complexity of the proposed mapping method is $O(n2M)$, where n is the number of analog elements, and M is the average number of 1s in the element code-vectors (M may be considered constant). The developed techniques were implemented and studied experimentally on analogical episodes that are commonly used for model testing.

Experiments

Experiment 1. In this experiment, we consider the episodes adapted by [Plate 2000, 2003] from [Thagard et al., 1990]. The following entities are involved: dogs (Fido, Spot, Rover), people (Jane, John, Fred), a cat (Felix), a mouse (Mort). Relations are bite, flee, cause. The probe (base) episode P is "Spot bit Jane, causing Jane to flee from Spot". Other episodes (Table 1) have the same relations as the probe, but different types of similarity - mainly according to Gentner's classification [Gentner, 1983; Forbus et al., 1995].

Table 1. The animal analogical episodes with various types of similarity to the Probe

Similarity type	Episode						Comments (description)
Probe P	Spot	bit	Jane	causing	Jane	to flee from	Spot	all episodes are described compared to Probe
Literal similarity LS	Fido	bit	John	causing	John	to flee from	Fido	both structural and superficial similarity
Surface features SF	John	fled from	Fido	causing	Fido	to bite	John	superficial but not structural similarity
Cross-mapped CM	Fred	bit	Rover	causing	Rover	to flee from	Fred	structural and superficial but entities switched
Analogy AN	Mort	bit	Felix	causing	Felix	to flee from	Mort	structural but not superficial similarity
First order relations FOR	Mort	fled from	Felix	causing	Felix	to bite	Mort	neither structural nor superficial similarity

Dimensionality of code-vectors was $N=10^5$ if not stated otherwise. The average number of 1s M(attribute) = M(object) = 1000 was chosen as in [Rachkovskij, 2001], whereas M(role) = 2000 > M(attribute) = M(object)=1000 was chosen to reflect importance of relations. Instantiations of those parameters and the thinning factor 0.2 (defined as $|\langle Z \rangle|/|Z|$, see (2) and (3)) were selected from the interval that provided correct retrieval (not mapping yet!) scores for the animal episodes, as in [Rachkovskij & Slipchenko, 2009; under revision].

Correct mapping of the Probe (Base) episode to each of the other episodes is shown in columns of the Table 1. It may seem rather non-evident, but it is ensured by the same roles in the higher-order relation "cause". If we use the naive mapping by the direct similarity of code-vectors of the initial role-filler representation schemes (used for retrieval and represented by the "lower" code-vector of (2)), results of [Rachkovskij 2001, 2004] show that only the LS episode can be correctly mapped to the Probe, and for the AN episode, the mapping was correct for all elements, despite the entities. For the SF, the CM, the FOR episodes, the correct mapping can not be established by the similarity of the "lower" code-vectors. On the other hand, the results obtained using the re-representation code-vectors and the techniques described in this paper are correct. This is illustrated for the Probe-FOR episodes in Table 2 that shows the similarity values calculated for the re-representation code-vectors of the elements of those episodes. The correct mapping requires the values at main diagonal to be the largest, and that is actually observed.

Table 2. The similarity matrix for the element code-vectors (re-representation) of the Probe and the FOR episode

	Probe	Bite	Flee	Spot	Jane
E_FOR	0.48	0.38	0.39	0.64	0.65
Flee	0.31	0.47	0.30	0.66	0.67
Bite	0.30	0.30	0.47	0.66	0.66
Mort	0.23	0.31	0.31	0.81	0.45
Felix	0.24	0.31	0.31	0.45	0.81

Experiment 2. We investigated our mapping and inference model using rather complex analogical episodes that are usually used to test analogical models, such as various versions of "Water Flow – Heat Flow", "Solar System – Atom", "Schools", etc. The code-vector parameters and representation scheme are the same as in Experiment 1.

In figures below, dashed lines show the found correspondences between the elements of the base and the target analogs, the numbers on the arrows indicate the similarity value of the mapped elements, and the dotted analog elements of the target show the hypotheses.

In the experiment with the Water Flow – Heat Flow analogy (Fig. 3), the similarity between clean(beaker), liquid(water), flat-top(water), greater(diameter(beaker), diameter(vial)) and any element of the other analog is less than similarity of random code-vectors, so they do not map. The mappings for the other elements of the Heat Flow analog established by the similarity of corresponding code-vectors, as described in the techniques proposed in this paper, are one-to-one and correct The similarity between the unmapped cause (...) code-vector from the base analog and the elements of the target analog flow(…) and greater(…) is less than its similarity to the code-vector of the target episode, so the hypothesis of cause (...) transfer to the target becomes the valid inference.
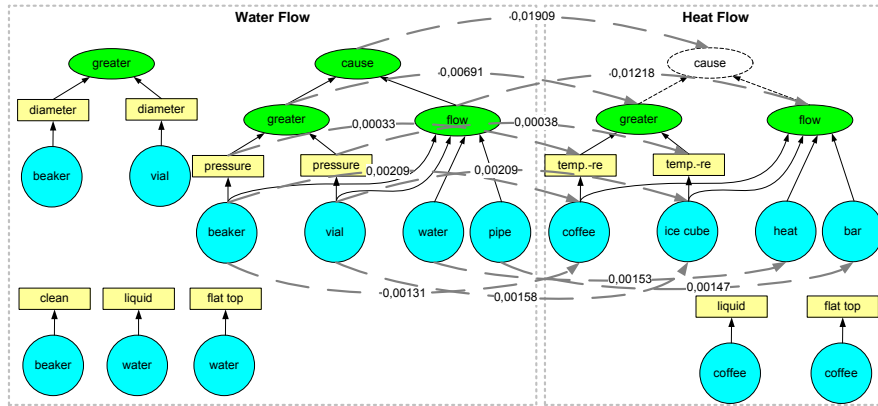
Fig. 3. Mapping and inference for the Water Flow – Heat Flow analogy

For the "Solar System – Atom" analogy (Fig. 4), the algorithm generates and checks hypotheses at several levels (see cause(…) and and(…) in the base). Unlike the previous example, where there was only one group of mappings, and all non-essential mappings and inferences were ignored due to the low similarity value, here after generation of inference hypotheses we observe three mapping groups.

The first group establishes the correspondence between the difference of mass, gravity and rotation of bodies. In this group, the causal relation between the attraction of the nucleus and the electron, and the rotation of the electron around the nucleus has been transferred to the Atom from the Solar System as a hypothesis and confirmed by similarity of appropriate code-vectors.
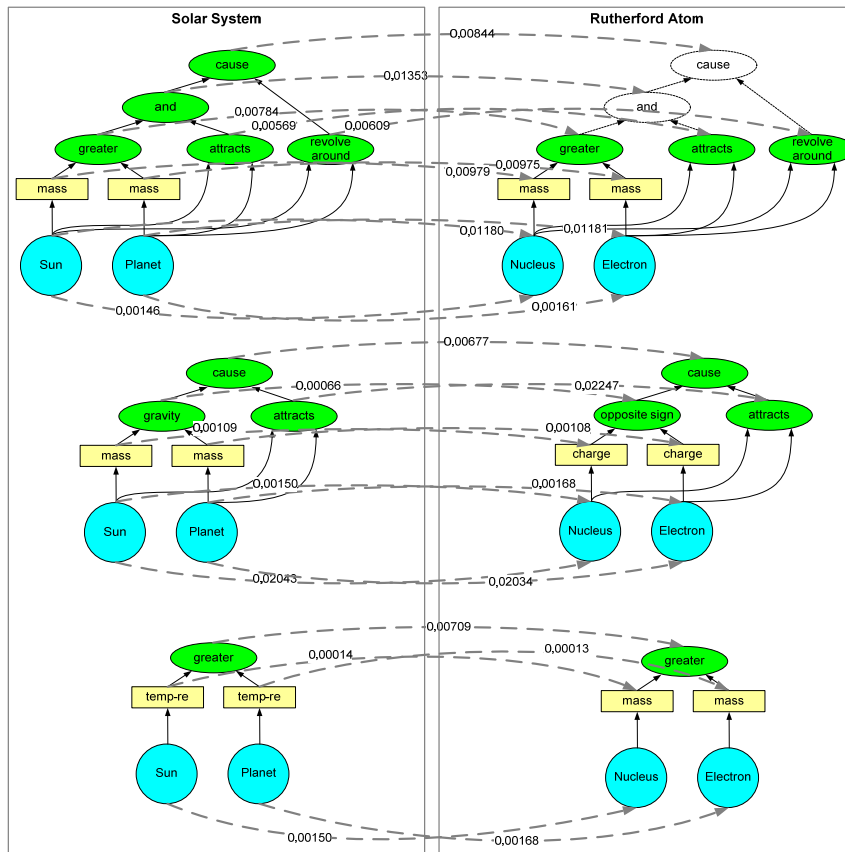

Fig. 4. Mapping and inference for the Solar System – Atom analogy

The second group matches the causation between the gravity and attraction in the Solar System analog and the causation between the charge sign difference and the attraction in the Atom analog.

The third group matches the temperature differences between the Sun and the Planet, as well as the mass difference of the Nucleus and the Electron, but the combined weight of those matches is very small (less than a random overlap), so they should be ignored.

For the "Schools" analogy (Fig. 5) [Markman 1997] all four cause() relations present in the Old School and Absent in the New School were correctly transferred to the target, and were accepted as valid inferences.

Thus, the proposed approach allows processing of fairly complex analogs and creates mappings and inferences consistent with the results of psychological tests and with the results of other analogical models.

```
"Old-School"
; English Department
(CAUSE (OBTAIN ENGLISH-FACULTY GRANTS) (HIRE ENGLISH-FACULTY RESEARCH-
ASSISTANTS))
(CAUSE (INFIGHTING ENGLISH-FACULTY) (AVOID ENGLISH-FACULTY OFFICES))
; Biology Department
(CAUSE (EXCELLENT BIOLOGY-FACULTY TEACH) (OVERSUBSCRIBED CLASSES))
(CAUSE (SMALL-NUMBER BIOLOGY-FACULTY) (NOT (PERFORM BIOLOGY-FACULTY ADVISING)))

"New-School"
; Computer Science Department
(OBTAIN CS-FACULTY GRANTS)
(INFIGHTING CS-FACULTY)
; Music Department
(EXCELLENT MUSIC-FACULTY TEACH)
(SMALL-NUMBER MUSIC-FACULTY)
```

Fig. 5. The Old School – New School analogy.

Experiment 3. The computational complexity of the implementation of the proposed code-vector-based approach to analogical mapping and inference largely depends on the code-vector dimensionality. We investigated the quality of mapping and inference for varying N at constant $p = N / M$ (equal to 0.01 or 0.02 for some terminal code-vectors as described above). Fig. 6. shows the percentage of correct mappings vs N for mapping of the animal episodes. Fig. 7 shows the standard measures of recall, precision, and integrated F1-measure of inference vs N for the Schools analogy. For each N, the results reported were averaged over 100 instances of random terminal code-vectors used to construct code-vectors of analogs. Both figures show reliable results at $N >$ 1000...10000.

In general, the experiments have shown the adequacy of the proposed approach to modeling mapping and inference based on the similarity of re-representations of analog elements and a lower computational complexity of its implementation compared to the traditional symbolic methods.
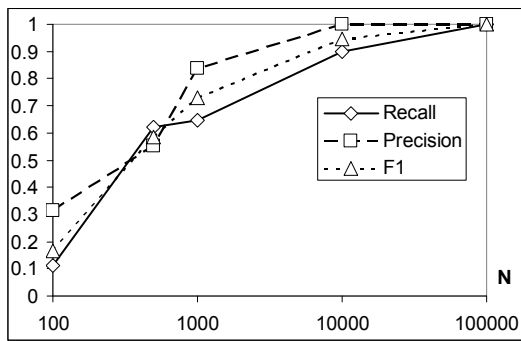
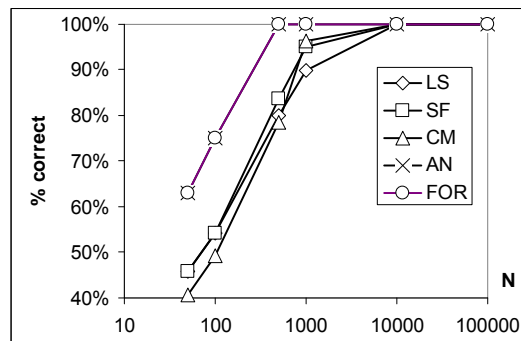Fig. 6. Mapping quality vs code-vector dimensionality N

Fig. .7. Inference quality vs code-vector dimensionality N

Discussion

Analysis of analogical models developed on the basis of symbolic or localist neural representations (SME, ACME, etc.) shows that, although such models are well suited for processing and comparison of complex hierarchical structures required for analogical reasoning, estimation of analogical similarity requires mapping of their elements implemented by computationally expensive procedures. Computational complexity of SME varies from $O(n^2)$ to $O(n!)$, and that of ACME [Holyoak & Thagard, 1989] is $O(n^4)$.

In quest for a more adequate account of analog semantics, increasing the neurobiological relevance of analogy models and their scaling potential, some new models of analogical reasoning employed distributed representations. However, in previous models the use of distributed representations was fragmented or inconsistent, or they were able to work only with simple analogies.

We propose a scheme for constructing distributed re-representations (code-vectors) of analog elements that allow finding correspondences between the elements (mapping) and justification of new knowledge (relations) transfer to the target (inference validation) to be done based on the similarity of the re-representation code-vectors. Code-vectors are binary distributed vectors that are similar for similar elements in terms of vector similarity measures based on dot-product. Those re-representation code-vectors are constructed from two parts – the "lower" part (that may be considered as more semantically grounded) and the "upper part" (mainly reflecting structure). This creates a context-dependent representation of analogs' elements – the same element gets different representations depending on the systems of relations in which it participates (reflected in the "upper" representation code-vector). Also, the same relational system gets different contexts to which it applies. (a kind of grounding reflected in the "lower" representation code-vector).

Experiments with the developed mapping and inference techniques using analogies that were previously used for testing known symbolic models (SME, ACME, etc.) and psychological experiments showed appropriate results and have confirmed the adequacy of the proposed approach and methods of analogical reasoning.

The computational complexity of the proposed mapping method is $O(n\ n')–O(n^2)$, where n' is the number of elements for which mapping should be established, n is the number of elements of another analog, with whom mappings are established. This is at the level of the lowest known computational complexity of mapping, and lower than the complexity of traditional methods $O(n^4)–O(n!)$. So the proposed methods are promising for mapping fragments of knowledge bases consisting of a large number of elements.

The proposed combined representation consisting of two parts lends itself to further elaboration. For example, the lower part can be made more grounded, if the terminal code-vectors are taken not random, but are formed by and

reflect the semantic basis of the outside world and the semantic similarity of corresponding objects and relations. The upper part can be made more structure-sensitive by using not only the set of higher-level roles (which may not distinguish different relational structures including identical relations in different arrangements) but also binding them with each other. We think that such representations would be useful not only in models of analogy, but also in other (cognitive) models, such as structure-sensitive models of similarity [Taylor & Hummel, 2009], etc.

We think that interesting and promising directions of future research of structure-sensitive distributed representations might include the following:

- The mechanisms for various possible mappings of the same analogical element.

- Making mapping with code-vector similarity more dynamic by introducing a process by which alternative mappings compete with each other, e.g. in the spirit of [Gayler & Levy, 2009].

- Mapping of large knowledge base fragments consisting of many elements.

- Possible code-vector representations in long-term memory. Whether only code-vector representations for retrieval are stored in memory and after retrieval are deployed and re-represented for mapping, or, if re-re-representations and mappings are stored in memory.

- How storing code-vector distributed representations in associative memory influenced retrieval, generalization, and other processes of analogical reasoning.

Bibliography

[Agichtein & Gravano, 2000] E.Agichtein & L.Gravano (2000). Snowball: Extracting relations from large plain-text collections. Proceedings of the fifth ACM conference on Digital libraries, San Antonio, Texas, US, pp. 85 - 94.

[Bobrow et al., 2009] D.G.Bobrow, C.Condoravdi, L.Karttunen, & A. Zaenen (2009). Learning by reading: normalizing complex linguistic structures onto a knowledge representation. AAAI Symposium 2009: Learning by Reading and Learning to Read; 2009 March 23-25; Stanford, CA.

[Browne & Sun, 2001] A.Browne & R.Sun (2001). Connectionist inference models. Neural Networks, 14(10): 1331-1355.

[Eliasmith & Thagard, 2001] C. Eliasmith & P. Thagard (2001). Integrating Structure and Meaning: A Distributed Model of Analogical Mapping. Cognitive Science, 25(2), 245-286.

[Falkenhainer et al., 1989] B. Falkenhainer, K.D. Forbus, & D. Gentner (1989). The Structure-Mapping Engine: Algorithm and Examples. Artificial Intelligence, 41, 1-63.

[Forbus, Mostek, & Ferguson, 2002] K.D.Forbus, T.Mostek, & R.Ferguson (2002). An analogy ontology for integrating analogical processing and first-principles reasoning, Proceedings of the 14th conference on Innovative applications of artificial intelligence, p.878-885, July 28-August 01, 2002, Edmonton, Alberta, Canada.

[Forbus et al., 1995] K.D.Forbus, D.Gentner, & K.Law (1995). MAC/FAC: a model of similarity-based retrieval. Cognitive Science, 19(2):141-205.

[Forbus et al., 2007] K.Forbus, C.Riesbeck, L.Birnbaum, K.Livingston, A.Sharma, & L.Ureel (2007). Integrating Natural Language, Knowledge Representation and Reasoning, and Analogical Processing to Learn by Reading. Proceedings of AAAI-07: Twenty-Second Conference on Artificial Intelligence, Vancouver, BC.

[Forbus et al., 2008] K.Forbus, J.Usher, A.Lovett, K.Lockwood, & J. Wetzel (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. In the Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling. Annecy, France.

[Frasconi, Gori, & Sperduti] P.Frasconi, M.Gori, & A.Sperduti (1998). A general framework for adaptive processing of data structures. IEEE Transactions on Neural Networks, 9(5):768-786.

[French, 1995] R.M.French (1995). The Subtlety of Sameness. Cambridge, MA: The MIT Press, 320 pages.

[Frolov, Rachkovskij, & Husek] A.A.Frolov, D.A.Rachkovskij, & D.Husek (2002). On Information Characteristics of Willshaw-Like Auto-Associative Memory. Neural Network World , Issue 2 , p. 141-157.

[Frolov et.al, 2007] A.A.Frolov, D.Husek, I.P.Muraviev, & P.Y.Polyakov (2007). Boolean factor analysis by attractor neural network. IEEE Trans. Neural Netw. 18(3), pp. 698-707.

[Gayler, 1998] R. Gayler (1998). Multiplicative binding, representation operators, and analogy. In Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences. Edited by K. Holyoak, D. Gentner, and B. Kokinov. New Bulgarian University. Sofia, Bulgaria, pp. 405.

[Gayler & Levy, 2009] R.W. Gayler & S.D. Levy (2009). A distributed basis for analogical mapping. In Proceedings of the Second International Analogy Conference. NBU Press, Sofia, Bulgaria, pp. 165-174.

[Gentner, 1983] D. Gentner (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, pp. 155-170.

[Gentner & Markman, 2003] D.Gentner & A.B.Markman (2003). Analogy-based reasoning and metaphor. In The Handbook of Brain Theory and Neural Networks. Edited M.A Arbib. The MIT Press, Cambridge, MA, pp. 106-109.

[Gentner & Colhoun, 2010] D.Gentner & J.Colhoun (2010). Analogical processes in human thinking and learning. In A. von Müller & E. Pöppel (Series Eds.) & B. Glatzeder, V. Goel, & A. von Müller (Vol. Eds.), On Thinking: Vol. 2. Towards a Theory of Thinking. Springer-Verlag Berlin Heidelberg. p. 35-48.

[Gentner et.al., 2009] D.Gentner, J.Loewenstein, L.Thompson, K.D.Forbus (2009) Reviving Inert Knowledge: Analogical Abstraction Supports Relational Retrieval of Past Events. Cognitive Science 33 (2009), pp.1343–1382

[Gladun, 2000] V.P.Gladun (2000). Partnership with Computers. Kiev: Port-Royal, 128 p. (in Russian).

[Gladun et al., 2000] V.P.Gladun, V.Ju.Velichko, N.N.Kiselyova, N.N.Moskalkova (2000). Inference of Hypothesis on Composition and Properties of Objects on the Basis of Analogy. Artificial Intelligence, 1, pp.44-52 (in Russian).

[Gray et al., 1997] B.Gray, G. S. Halford, W. H. Wilson, & S.Phillips, (1997). A Neural Net Model for Mapping Hierarchically Structured Analogs, In Proceedings of the Fourth conference of the Australasian Cognitive Science Society. University of Newcastle, NSW, Australia.

[Hahn & Schnattinger, 1998] U. Hahn & K. Schnattinger (1998). Towards Text Knowledge Engineering, In: AAAI'98 - Proceedings of the 15th National Conference on Artificial Intelligence. Madison, Wisconsin, USA, July 26-30, 1998. AAAI Press - MIT Press, 1998, pp.524-531.

[Hofstadter, 1995] D.Hofstadter (1995). Fluid Concepts & Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought, NY: Basic Books.

[Holyoak & Thagard, 1989] K.J. Holyoak, & P. Thagard (1989). Analogical mapping by constraint satisfaction. Cognitive Science, 13, 295-355.

[Hummel & Holyoak, 1997] J.E. Hummel & K.J Holyoak (1997). Distributed representations of structure: A theory of analogical access and mapping. Psychological Review, 104, pp. 427-466.

[Hummel & Holyoak, 2003] J. E.Hummel & K.J.Holyoak (2003). A symbolic-connectionist theory of relational inference and generalization. Psychological Review, 110, 220–264.

[Kanerva, 1996] P.Kanerva (1996). Binary Spatter-Coding of Ordered K-tuples. In C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, & B. Sendhoff (Eds.). Proceedings of the International Conference on Artificial Neural Networks - ICANN'96, Bochum, Germany. Lecture Notes in Computer Science, 1112, 869-873. Berlin: Springer.

[Kanerva, 1998] P.Kanerva (1998). Dual role of analogy in the design of a cognitive computer. In: K. Holyoak, D. Gentner, and B. Kokinov (Eds.), Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences (Proc. Analogy'98 workshop). NBU Series in Cognitive Science, Sofia, Bulgaria: New Bulgarian University, pp. 164-170.

[Kanerva, 2009] P.Kanerva (2009). Hyperdimensional computing: An introduction to computing in distributed representation with highdimensional random vectors. Cognitive Computation, 1:139-159.

[Keane, Ledgeway, & Duff, 1994] M.T.Keane, T.Ledgeway, & S.Duff (1994). Constraints on Analogical Mapping: A comparison of Three Models. Cognitive Science, 18: 387-438.

[Kim & Porter, 2009] D.S.Kim & B.Porter (2009). Kleo: A Bootstrapping Learning-by-Reading System. Symposium 2009: Learning by Reading and Learning to Read; 2009 March 23-25; Stanford, CA.

[Kokinov, 1988] B. Kokinov (1988). Associative memory based reasoning: How to represent and retrieve cases. In Artificial intelligence III: Methodology, systems, applications. Edited by T. O'Shea and V. Sgurev. Elsevier Science Publishers B.V., Amsterdam, North Holland, pp. 51-58.

[Kokinov & French, 2003] B.Kokinov & R.M.French (2003). Computational Models of Analogy-making. In Nadel, L. (Ed.) Macmillan Encyclopedia of Cognitive Science. Vol. 1, pp.113 - 118. London: Nature. Publishing Group.

[Kokinov & Petrov, 2001] B.Kokinov & A.Petrov (2001). Integrating memory and reasoning in analogy-making: The AMBR model. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), The analogical mind: Perspectives from cognitive science (pp. 59-124). Cambridge, MA: MIT Press

[Kuehne, 2000] S.Kuehne, K.Forbus, D.Gentner, & B.Quinn (2000). SEQL: Category Learning as Progressive abstraction using structure mapping Proc. 22nd Ann. Conf. Cognitive Science Soc. (Cogsci 00), Lawrence Erlbaum Associates, 2000, pp. 770–775.

[Kussul, 1992] E.M.Kussul (1992) Associative neuron-like structures. Kiev: Naukova Dumka. 144 p. (In Russian).

[Kussul et al., 2006] E.M.Kussul, T.N. Baidyk, D.C. Wunsch, O. Makeyev, & A. Martin (2006). Permutation coding technique for image recognition systems. IEEE Transactions on Neural Networks, 17(6):1566-1579.

[Larkey & Love, 2003] L.B.Larkey & B.C.Love (2003). CAB: Connectionist Analogy Builder [find similar] [try Google] Cognitive Science, 27(5), pp. 781-794, 2003.

[Markman, 1997] A.B. Markman (1997). Constraints on Analogical Inference. Cognitive Science, 21(4), 373-418.

[Markman & Gentner, 2000] , A.B.Markman & D.Gentner (2000) Structure mapping in the comparison process. American Journal of Psychology, 113, 501-538.

[Markman et al., 2003] A.B.Markman, D.A.Rachkovskij, I.S.Misuno,& E.G.Revunova (2003). Analogical reasoning techniques in intelligent counterterrorism systems // Informational Theories & Applications, 10(2), pp. 139-146.

[Misuno, Rachkovskij, & Slipchenko, 2005] I.S. Misuno, D.A.Rachkovskij, & S.V.Slipchenko (2005). Vector and distributed representations reflecting semantic relatedness of words. Mathematical Machines and Systems, Issue 3, p. 50–66 (in Russian).

[Mostek, Forbus, & Meverden, 2000] T. A.Mostek, K. D. Forbus,.& C.Meverden (2000). Dynamic case creation and expansion for analogical reasoning. In Proceedings of the 17th National Conference on Artificial Intelligence.and Twelfth Conference on Innovative Applications of Artificial Intelligence, pp. 323 - 329.

[Navarro & Lee, 2003] D.J. Navarro & M.D. Lee (2003). Combining Dimensions and Features in Similarity-based Representations, Advances in Neural Information Processing Systems. Ed. by S. Becker, S. Thrun, K. Obermayer. N.Y.: MIT Press, P. 6774. No. 15.

[Plate, 2000] T.A.Plate (2000). Analogical Retrieval and Processing with Distributed Vector Representations. Expert Systems: The International Journal of Knowledge Engineering and Neural Networks, 17(1), pp. 29-40.

[Plate, 2003] T.A. Plate (2003). Holographic Reduced Representation: Distributed Representation for Cognitive Structures. Stanford: CSLI Publications, 300 p.

[Pollack, 1990] J. B.Pollack (1990). Recursive distributed representations. Artificial Intelligence, 46: 77-105.

[Rachkovskij, 2001] D.A. Rachkovskij (2001). Representation and processing of structures with binary sparse distributed codes. IEEE TKDE 13(2), pp. 261-276.

[Rachkovskij, 2004] D.A. Rachkovskij (2004). Some Approaches to Analogical Mapping with Structure Sensitive Distributed Representations, Journal of Experimental and Theoretical Artificial Intelligence, 16(3), pp. 125-145.

[Rachkovskij & Kussul, 2001] D.A. Rachkovskij & E.M. Kussul (2001). Binding and Normalization of Binary Sparse Distributed Representations by Context-Dependent Thinning. Neural Computation, 13(2), pp. 411-452.

[Rachkovskij & Slipchenko, under revision] D.A.Rachkovskij & S.V. Slipchenko. Similarity based retrieval with structure-sensitive sparse binary distributed representations

[Ramscar & Yarlett, 2003] M.Ramscar & D.Yarlett (2003). Semantic grounding in models of analogy: an environmental approach. Cognitive Science 27(1): 41-71.

[Sahlgren, Holst, & Kanerva, 2008] M.Sahlgren, A.Holst, & P.Kanerva (2008). Permutations as a Means to Encode Order in Word Space. Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci'08), July 23-26, Washington D.C., USA.

[Schubert, 2002] L.K. Schubert (2002). Can we derive general world knowledge from texts?, M. Marcus (ed.), Proc. of the 2nd Int. Conf. on Human Language Technology Research (HLT 2002), March 24-27, San Diego, CA, pp. 94-97.

[Shastri & Ajjanagadde, 1993] L.Shastri, & V.Ajjanagadde (1993). From simple associations to systematic reasoning: connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. Behavioral and Brain Sciences, 16: 417-494.

[Shepard, 1962] R. N. Shepard (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part I. Psychometrika, 27, 125–140. Part II. Psychometrika, 27, 219–246.

[Slipchenko & Rachkovskij, 2009] S.Slipchenko & D.Rachkovskij (2009). Mapping and analogical inference based on neural distributed representations. 15-th Intern. Conf. Knowledge-Dialogue-Solution, Varna, Bulgaria. 9, p. 95–101 (in Russian).

[Smolensky, 1990] P.Smolensky (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. Artificial Intelligence, 46: 159-216.

[Taylor & Hummel, 2009], E.G.Taylor, J.E.Hummel (2009). Finding similarity in a model of relational reasoning Cognitive Systems Research, Vol. 10, No. 3, pp. 229-239.

[Tenenbaum, 1996] J.B.Tenenbaum (1996). Learning the structure of similarity. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), Advances in Neural Information Processing Systems (Vol. 8, p. 3-9). Cambridge, MA: MIT Press.

[Thagard et al., 1990] P.Thagard, K.J.Holyoak, G.Nelson, D.Gochfeld (1990). Analog Retrieval by Constraint Satisfaction Artificial Intelligence 46, pp. 259-310.

[Thorpe, 2003] S.Thorpe (2003). Localized versus distributed representations. In The Handbook of Brain Theory and Neural Networks. Edited by M.A Arbib. The MIT Press, Cambridge, MA, pp. 643-646.

[Tversky, 1977]. A. Tversky (1977). Features of Similarity Psychological Review 84(4), pp. 327--352.

[Yan, Forbus, & Gentner, 2003] J.Yan, K.Forbus, & D. Gentner (2003). A theory of rerepresentation in analogical matching. In R. Alterman & D. Kirsch (Eds.), Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Authors' Information

Serge V. Slipchenko – *PhD, Senior Research Scientist, International Research and Training Center for Information Technologies and Systems, NAS and MES of Ukraine; Pr. Acad. Glushkova, 40, Kiev, 03680, Ukraine; e-mail: serge.slipchenko.irtcits (at) gmail.com*

Major Fields of Scientific Research: Computational and Artificial Intelligence

Dmitri A. Rachkovskij – *PhD, DSc, Leading Research Scientist, International Research and Training Center for Information Technologies and Systems, NAS and MES of Ukraine; Pr. Acad. Glushkova, 40, Kiev, 03680, Ukraine; e-mail: dar (at) infrm.kiev.ua*

Major Fields of Scientific Research: Computational and Artificial Intelligence