FULL-FLEDGED SEMANTIC ANALYSIS AS A TOOL FOR RESOLVING TRIANGLE-COPA SOCIAL SCENARIOS

Boguslavsky I. M. (bogus@iitp.ru)

A. A. Kharkevich Institute for Information Transmission
Problems, Russian Academy of Sciences, Moscow, Russia;
Universidad Politécnica de Madrid, Madrid, Spain

Dikonov V. G. (sdiconov@mail.ru),
Frolova T. I. (tfrolova@gmail.com),
Iomdin L. L. (iomdin@gmail.com),
Lazursky A. V. (lazursky@mail.ru),
Rygaev I. P. (irygaev@gmail.com),
Timoshenko S. P. (nyrestein@gmail.com)

A. A. Kharkevich Institute for Information Transmission
Problems, Russian Academy of Sciences, Moscow, Russia

Text interpretation often requires common sense knowledge and reasoning. A convenient tool for developing methods of common sense reasoning are special sets of challenge problems whose interpretation requires sophisticated reasoning. An interesting example is a recently published data set called Triangle Choice of Plausible Alternatives (Triangle-COPA), which contains 100 multiple-choice problems that test the interpretation of social scenarios. Each problem includes a statement and two alternatives. The task is to identify the more plausible alternative. For processing Triangle-COPA data we use SemETAP, a general purpose semantic analyzer. We implement the full scenario of NL understanding starting from NL texts and not from manually composed simplified logical formulas, which is a common practice in logic-based approaches to common sense reasoning. We produce Enhanced Semantic Structures of the statement and both alternatives and check which alternative manifests more semantic agreement with the statement in terms of inferences.

ОДИН ПОДХОД К ОПИСАНИЮ СОЦИАЛЬНЫХ СЦЕНАРИЕВ TRIANGLE-COPA НА ОСНОВЕ СЕМАНТИЧЕСКОГО АНАЛИЗА

Интерпретация текста часто требует обращения к фоновым знаниям, основанным на логике здравого смысла, и умозаключений на основе этих знаний. Удобным инструментом для разработки методов обращения с такими знаниями являются специально составленные вопросы, ответ на которые требует определенных умозаключений. Интересным примером такого набора вопросов является недавно опубликованный набор под названием Triangle Choice of Plausible Alternatives (Triangle-COPA). Он содержит 100 заданий с многовариантными ответами, требующих интерпретации социальных сценариев. Каждое задание содержит утверждение и две связанных с ним альтернативы. Задача состоит в идентификации более правдоподобной альтернативы. Для работы с этими заданиями мы используем семантический анализатор общего назначения SemETAP. Для систем, стремящихся делать логические умозаключения, основанные на здравом смысле, типично, что они оперируют с логическими формулами, составленными вручную. В отличие от этого, наш анализатор получает на вход текст на естественном языке и реализует полный сценарий обработки этого текста. Мы строим Расширенную семантическую структуру как утверждения, содержащегося в задании, так и обеих предложенных альтернатив и стремимся определить, какая из альтернатив обнаруживает большее семантическое согласование с утверждением с точки зрения возможных умозаключений.

1. Introduction

As is known, computer text understanding includes, on the one hand, linguistic analysis of the text, and on the other hand, a large area of logical and common sense reasoning. These two lines of research are pursued to a large extent independently of one another. In most cases, text analysis is performed by means of machine learning techniques and does not include logical processing of the representation obtained (cf., for example, [Ge and Mooney 2005], [Bos 2008], [Poon, Domingos 2009]). Logical and common sense reasoning, as a rule, applies to some logical form, usually hand-authored [Mueller 2006]. Artificial Intelligence researchers who model human commonsense reasoning seek to address the following fundamental tasks: represent commonsense knowledge as formal theories for use in automated reasoning systems [Gordon et al. 2011], [Hobbs et al. 2020], [Hobbs et al. 2012], [Montazeri et al. 2011], [Montazeri et al. 2012], [Morgenstern 2001], design cognitive architectures for commonsense reasoning [Meadows et al. 2014], and evaluate automatic reasoning systems [Levesque et al. 2012]; [Roemelle et al. 2012]. There are very few integrated approaches that take an NL text at the input, construct some sort of semantic representation thereof and perform a large spectrum of logical and common sense inferences.

A notable exception is a series of publications in the framework of Ontological Semantics initiated by [Nirenburg, Raskin 2004]. Our semantic analyzer [Boguslavsky 2011], [2017], [Boguslavsky et al. 2015], [2018], [2019] is being developed in the same vein.

Over the last years, two evaluation tools have been proposed to spur progress in automated commonsense reasoning. The first one, the Choice of Plausible Alternatives (COPA), tests commonsense causal reasoning in everyday situations [Roemelle et al. 2011]. It consists of one thousand binary-choice questions; and the task is to select the more plausible cause or result of a given situation. The second approach is the Winograd Schema Challenge (WSC) [Levesque et al. 2011]. It formulates the commonsense reasoning problem as a reference resolution task. Each question consists of a pair of sentences that differ in only one or two words, which switch the way the reference should be resolved. However, neither of these evaluation tools has been instrumental in advancing logical formalization of commonsense reasoning. In both cases, the best published results come from systems based on machine learning techniques.

Later, a new set of challenge problems has been proposed, called Triangle-COPA [Maslan et al. 2015]. As opposed to previous challenge problems, it was not conceived as an evaluation tool. It was specifically designed as a development test set, rather than a held-out test set for use in competitive evaluations, which means that it was intended to aid in the logical formalization of commonsense knowledge needed to correctly answer the questions. The approach adopted in Triangle-COPA differs from that seen in COPA and WSC in two important ways. First, Triangle-COPA test contains not only NL (English) questions and answers, but also their first-order logical form. This eliminates the need for sophisticated NL processing. Second, the domain of situations described in the questions was strongly constrained. This allowed researchers to concentrate their efforts on specific areas of commonsense reasoning.

Below, we will briefly describe the Triangle-COPA test set and review the existing solutions (section 2). Then we will present our approach (section 3) and describe our experiments (section 4). We will conclude in section 5.

2. Triangle-COPA

The Triangle Choice of Plausible Alternatives (Triangle-COPA) is a set of 100 challenge problems for logical formalization of commonsense psychology. All the texts describe events occurring in a simple environment whose participants are characters of a short film authored by the social psychologist Fritz Heider [Heider and Simmel 1944]. They present some brief sequence of actions and interactions between three geometrical shapes. Two triangles and a circle perform various actions in and around a room with a door. Each text consists of a statement and two alternatives connected with the statement, where one of these two is much more plausible than the other and provides a better explanation of the statement. The task is to computationally determine the correct alternative. An example of a Triangle-COPA challenge problem is as follows:

(1) Text 83

Statement: A small triangle and a big triangle are next to each other. A circle runs by and pushes the small triangle. The big triangle chases the circle. Why does the big triangle chase the circle?

Logical form: (and (approach' e1 C LT) (push' e2 C LT) (chase' e3 BT C) (seq e1 e2 e3))

Alternative 1. The big triangle is angry that the circle pushed the small triangle, so it tries to catch the circle.

Logical form: (angryAt' e4 BT C)

Alternative 2. The big triangle and circle are friends. The big triangle wants to say hello to the circle.

Logical form: (and (friend' e5 BT C) (goal' e6 e7 BT) (greet' e7 BT C)).

Notation: e1, e2,…en—events; C—Circle, BT—Big Triangle, LT—Little Triangle.

We are aware of two approaches to Triangle-COPA test published in [Gordon 2016] and [Kalluri et al. 2017].

[Gordon 2016] uses logical and probabilistic reasoning. Commonsense knowledge of actions, social relationships, intentions, and emotions are encoded as defeasible axioms in first-order logic. For example, axioms (2a)–(2d) provide possible reasons why one character would be chasing another:

(2) (a) Chase 1: Maybe they are playing tag
        (if (and (playWith' e1 x y)
        (etcChase1 0.2 e1 x y))
        (chase' e x y))

(b) Chase 2: Maybe one is angry at the other
        (if (and (angryAt' e1 x y)
        (etcChase2 0.2 e1 x y))
        (chase' e x y))

(c) Chase 3: Maybe one is trying to rob the other
        (if (and (goal' e1 e2 x)
        (rob' e2 x y)
        (etcChase3 0.3 e1 e2 x y))
        (chase' e x y))

(d) Chase 4: Maybe one is trying to scare the other
        (if (and (goal' e1 e2 x)
        (afraid' e2 y)
        (etcChase4 0.5 e1 e2 x y))
        (chase' e x y))

Axiom (2a) claims that if x plays with y, it is probable (with the 0.2 probability) that x chases y. Axiom (2b) says that chasing is also probable, and also with the 0.2 probability, if x is angry at y. With somewhat higher probability (0.3) x may chase y, if it wants to rob it (Axiom (2c)). Finally, x will chase y with the maximal probability (0.5), if x wants to scare y.

Using these axioms, possible interpretations of action sentences are identified via logical abduction (called Etcetera Abduction), backchaining to distinct sets of assumptions that logically entail the observations. These sets of assumptions are ordered by their joint probability assuming conditional independence. The axioms, 136 in total, describe commonsense knowledge in the domain of the Triangle-COPA question set. They are oriented towards specific inferences needed to infer the correct alternative for each question. Abduction, as distinct from logical deduction or induction, is a form of logical reasoning that identifies a hypothesis that, if it were true, would logically entail the given input.

In order to calculate the joint probability of events, each axiom should contain a numerical value, which characterizes the probability of the etcetera literal. Although it is supposed that in the future the probabilities will be extracted automatically out of empirical data, in the current version they are selected intuitively, which does not always look convincing. For example, our intuition does not confirm the estimation of Axiom (2d), which maintains that, if we wish to scare somebody, 50% chances are that we will chase them.

An important advantage of the approach is that the knowledge used is explicitly formulated in the form of axioms with the probabilities. Due to this, Etcetera Abduction provides an explanation of why one solution has been preferred over a different one. On the other hand, one should not disregard the fact that the axioms on which [Gordon 2016] draws were formulated specifically for Triangle-COPA questions. Judging by the examples given in Gordon's paper, these axioms are very fragmentary and do not seek to make integral part of some balanced knowledge repository suitable for a wide range of scenarios. However, this is not a defect in the context of the paper. The goal of Gordon 2016 is to develop a new variant of abduction building on a priori probabilities and to show that, if common sense axioms are available, the method works well. As [Gordon 2016] states, "our strong performance on the Triangle-COPA test demonstrates that Etcetera Abduction is a viable approach, but our success owes much to the labor of hand-crafting the axioms necessary to solve these specific questions".

[Kalluri et al. 2017] builds a model that deploys Bayesian inference with an action affinity lexicon to infer probabilistic affinity relations characterizing the scenario. Subsequently, the model uses the inferred affinity relations to choose the more probable alternative of two variants. As opposed to [Gordon 2016], this model does not use logical axioms nor absolute prior probabilities of events—two important components of [Gordon 2016] approach. The key concept on which [Kalluri et al. 2017] relies is the concept of affinity. This is the relation between two agents that may take one of three values—Pleasant/Unpleasant/Neutral. All the knowledge the model builds on is the static probabilistic action affinity lexicon, which links actions to the relative observation distribution of affinities. For example, action argue_with is assigned 0.5 Unpleasant affinity, 0.25 Pleasant and 0.25 Neutral. The approach of [Kalluri et al. 2017] does not advance much commonsense reasoning. However, it is interesting in that it shows that relatively much can be obtained if one disposes of very lightweight knowledge.

3. Our approach

We are taking quite a different stance. Of course, one can set a specific goal of solving exclusively the Triangle-COPA problem proper and, if possible, by minimal means, as it seems to be the case in [Kalluri et al. 2017]. In this case, we could be quite justified to use any kind of knowledge, however individual and case-oriented it may be, if it helps solve the problem.

We look at the problem from a different perspective. We perceive it as a particular case of a much more general task—the task of text understanding. In this vein, we have been developing a knowledge-based general semantic analyzer for Russian—SemETAP. In previous publications mentioned above, we described its various aspects and will not repeat them here. Let it only be reminded that:

- SemETAP is an option of the ETAP-4 linguistic processor and reuses its non-semantic modules (morphological analysis, syntactic dependency parsing, and normalization).
- Semantic analysis makes use of linguistic data and extralinguistic information (background knowledge). The linguistic data are provided by the Combinatorial Dictionary and the Grammar, and the background knowledge is stored in the Ontology, Repository of Individuals and the set of inference rules SemRule.
- Inference rules is a crucial component of SemETAP. We believe that the depth of understanding is growing with the number of inferences we can draw from the text. In many cases, a decomposition of the concept meaning helps produce additional inferences and thus achieve a deeper understanding.
- Two levels of semantic structure are distinguished. Basic semantic structure (BSemS) interprets the text in terms of ontological concepts. Enhanced semantic structure (EnSemS) extends BSemS by means of a series of inferences. For the purposes of this paper, EnSemS is the most important representation, since it makes explicit all the inferences that the knowledge available permits to make from the text and the context.
- Two types of inferences are carried out: 100%-true logical entailments, and implications that implement plausible expectations.

This approach determines several distinctive features of the way we handle Triangle-COPA texts.

First, we use them as a convenient tool of formulating common sense knowledge and checking its validity. Our priority is not so much correctly answering concrete Triangle-COPA questions, but rather incorporating new knowledge into an integrated NL understanding system and testing it.

Second, in processing Triangle-COPA, we implement the full scenario of NL understanding starting from NL texts and not from manually composed simplified logical forms. This is more difficult than accepting logical forms at the input, not only because additional stages of processing are needed but also because logical forms proposed by [Maslan et al. 2015] often oversimplify the initial English text. One will clearly see that if one compares the NL text and the logical form of Alternative 1 of the text 83 above. Since our priority is developing the semantic analyzer for Russian and checking its capacity against Triangle-COPA statements, we had to translate these statements into Russian.

Third, the knowledge we acquire, both linguistic and common sense, should not be specific for the Triangle-COPA texts. We do not introduce ad hoc knowledge. The SemETAP general semantic analyzer needs knowledge that is supposed to be valid for a wide range of situations and for different purposes.

Fourth, based on this general knowledge SemETAP performs a large number of inferences, both strict and plausible, which form the Enhanced Semantic Structure (EnSemS). Such a semantic structure permits to restore implicit semantic elements and answer questions even when the answer is not directly contained in the text.

Fifth, we resolve Triangle-COPA tests as follows. For each text, we obtain En-SemSs of the Statement and both Alternatives. Then, a special algorithm selects the alternative whose EnSemS manifests higher semantic agreement with the EnSemS of the Statement. This algorithm will be explained below.

The problem of knowledge acquisition and its generality vs. specificity is worthy of some more comments, since this is where our approach differs from many other approaches to common sense reasoning. As mentioned above, our model of NL understanding proceeds from an assumption that the more inferences we can draw from the text, the deeper our understanding is. Therefore, we aim at making the maximum amount of inferences. These include both inferences motivated by the lexical meaning of words, and inferences based on world knowledge. To make such inferences, the system has to possess a large amount of knowledge. As is known, the critical bottleneck on this way is knowledge acquisition, particularly as far as general commonsense knowledge is concerned.

Many knowledge-supported projects make use of existing resources, such as Word-Net, FrameNet, FreeBase, ConceptNet or YAGO [Ponzetto and Strube 2006], [2007], [Poesio et al. 2007], [Bryl et al. 2010], [Rahman and Ng 2011], [Uryupina et al. 2011], [Lee et al. 2011]. However, these resources do not provide information needed for the inferences we strive to make. Therefore, we have to acquire by ourselves most of the knowledge we need. Since common sense knowledge acquisition is an infinitely huge task, one can only advance by small incremental steps. The critical requirement is that acquired knowledge be not oriented strictly to a particular task for which it was acquired, in our case—Triangle-COPA test. It should retain its value in a wider context and for different tasks. Only in this case one could count on the incremental growth and stability of the knowledge repository. That is why we assign special importance to making knowledge as general as possible. What follows is an example illustrating this point.

Let us recall axioms (2a)–(2d) from [Gordon 2016], which describe situations of one person chasing another. We will focus on the Axiom (2b), which proclaims a direct link between the concepts AngryAt and Chase. It reads that if *x* is angry at *y*, then *x* will probably chase him. We conjecture that this axiom was introduced in order to resolve question 83 of the Triangle-COPA set, cited above but reproduced below for convenience:

Text 83

Statement: A small triangle and big triangle are next to each other. A circle runs by and pushes the small triangle. The big triangle chases the circle. Why does the big triangle chase the circle?
Logical form: (and (approach' e1 C LT) (push' e2 C LT) (chase' e3 BT C)
(seq e1 e2 e3))

Alternative 1. The big triangle is angry that the circle pushed the small triangle, so it tries to catch the circle.
Logical form: (angryAt' e4 BT C)
Alternative 2. The big triangle and circle are friends. The big triangle wants to say hello to the circle.
Logical form: (and (friend' e5 BT C) (goal' e6 e7 BT) (greet' e7 BT C)).

To answer question 83, axiom (2b) is sufficient, but if the knowledge repository is acquired for a wider use, it should contain descriptions that give more information about the concepts and link them in a more general way. In our model, we try to include in semantic definitions all components that may be needed for inference, although of course our definitions do not pretend to be exhaustive. By way of example, we show below how concepts Chase and Anger are described in SemETAP. For readers´ convenience, we represent these definitions by means of English glosses, and not in the formal Etalog language, in which they were originally written.

(3) Chase: 'Agent **?chaser** is moving towards Agent **?victim** having the goal of coming in contact with him and perform upon him an action **?harm**, which is assessed negatively. **?victim** is moving away from **?chaser** having the goal of avoiding coming in contact with him'.

(4) Anger: 'Agent **?experiencer** is in an intensive negative emotional state towards **?stimulus**', which may be accompanied by shouting or frowning on the part of **?experiencer;** if **?stimulus** is a living being, **?experiencer** may perform upon him an action **?harm**, which **?stimulus** assesses negatively; if **?stimulus** is an event, **?experiencer** probably feels Anger with respect to the initiator of event **?stimulus**; this event takes place both when Anger takes place and not'.

Comments:
1) The concept Agent used in phrases like 'Agent ?experiencer' above does not refer to the **semantic role** Agent. It means that the domain of variable ?experiencer is the Agent **ontological class**, which includes humans, organizations, etc.
2) The last proposition of definition (4) describes the fact that Anger is a factive predicate. This means that both (a) *Mary is angry that no one agreed with her suggestion* and (b) *Mary is not angry that no one agreed with her suggestion* imply that no one agreed with Mary's suggestion.
Definitions (3) and (4) allow for many inferences. For example, if we have the text

(5) *The dog chased the rabbit*

we can answer a range of questions, such as: *Was the dog moving?* (yes). *What was the destination of its movement?* (the rabbit). *What was its goal?* (to get in contact with the rabbit). *What can be expected if the dog gets in contact with the rabbit?* (it will do something the rabbit will not like). *What was the goal of the rabbit?* (to avoid coming in contact with the dog). If sentence (5) is followed by something like …*but did not succeed,* we can safely infer that the dog did not get in contact with the rabbit.

Turning to text 83, we can see that definitions (3) and (4) can help discover a semantic agreement between the Statement and the Alternative 1. On the one hand, (3) implies that, when chasing, the big triangle wants to do something bad to the circle. On the other hand, (4) states that, being angry with the circle, the big triangle will also possibly do something bad to him.

Since no semantic agreement can be found between the Statement and the Alternative 2, the first one should be preferred. In more detail, the algorithm of checking semantic agreement will be discussed below.

4. Experiment

In order to clearly estimate the value of Triangle-COPA as a development set, we split our work in two stages.

At the first stage, we compiled the list of words used in the Triangle-COPA texts and provided their full description. This list contains about 300 elements, mostly describing actions, mental and emotional states, social relationships, and intentions. This description consists of a linguistic part and a semantic-ontological part. The former part includes syntactic and semantic features, government pattern, lexical functions and some other types of information, which are not specific for the semantic analyzer. The second part contains the semantic-ontological correlate of the given word (if it exists), correspondence between syntactic and semantic relations, and inference rules, which in their turn include semantic decomposition of concepts and other axioms. Importantly, at the first stage we did not consult the Triangle-COPA texts and did not know what kind of common sense axioms are needed for them.

At the second stage, we used the Triangle-COPA as a development set. Concept descriptions compiled at the previous stage were supplemented by new axioms. In doing that, we took care not to construct ad hoc axioms. Essentially, using the Triangle-COPA as a development set helps incorporate into the knowledge repository the axioms that escaped our notice before. This is how we introduced such axioms as "If a person touches another person, he/she probably wishes to attract his/her attention" or "If a person has a low estimation of some situation that takes place, he/she wants it to stop".

After terminating each stage of knowledge preparation, we performed an experiment on selecting alternatives on the basis of this knowledge. The selection algorithm and the results of the experiment are presented below.

The algorithm starts with two EnSemSs (of the statement and of one of the alternatives) and calculates a degree of semantic similarity between them. For the purpose of the algorithm there is no difference which EnSemS represents the statement and which one the explanation hypothesis. They are treated symmetrically to find out how many propositions can be unified between the structures.

First, we identify all nodes corresponding to events marked as true facts or plausible expectations in each semantic structure and make two lists of such event nodes. Then we take a list with fewer elements and try to match each element from this list to a corresponding element in the other list. Nodes are matched in such a way that the propositions they represent are unifiable, i. e. they do not contradict each other on the assumption that they describe the same event.

Rules for unification check are the following:

- A node corresponding to an individual constant from the ontology is unifiable only to a node corresponding to the same constant.
- Nodes of different ontological classes are unifiable only if one class is a subclass of another.
- Event nodes are unifiable only if they do not contain non-unifiable arguments. By the argument here we mean any property defined in the ontology as functional, i. e. a property that can have only one value, not multiple values.

We take each functional property of the first event in the pair and check if the second event either does not have this property or their values are unifiable by the same algorithm.

An event from one list can potentially match (be unifiable) to multiple events in the other list. In such case, we select the best match by calculating a maximum number of matched functional properties, i. e. properties presented in both events and unified.

Once the list of matched pairs is produced we calculate a number of metrics:

- M1—purely a number of matched events;
- M2—M1 divided by the minimum length of the two lists;
- M3—M1 divided by the average length of the two lists.

The ideas behind M2 and M3 are the following. The minimum length of the two lists is the maximum number of events that could match in principle. For example, if there are 4 events in the first list and 20 events in the second, then no more than 4 events could match. M2 calculates the ratio of actually matched events to this maximum. But it is insensitive to the number of events in the bigger list. M2 would not change if the second list contained 120 events instead of 20. To account for that M3 was introduced. It is sensitive to the number of events in each list. Increasing either list without increasing matches will produce a smaller M3.

Once metrics for two alternative hypotheses are calculated, we select the hypothesis with a higher metric value. Our tests showed that M2 is the best metric, so we omit the other ones in the resulting table.

At the first stage the number of correctly predicted answers was not very high—53.5%, but at the second stage we made a significant improvement reaching 80,5% of correct predictions. The system produced 76 correct answers, 15 incorrect ones, and 9 random choices.

The analysis of the experiment results shows that there are three main reasons for erroneous selection.

First, in some cases our knowledge is still insufficient for the selection.

Second, in some cases, knowledge needed to resolve alternatives was not considered general enough to be introduced to the knowledge repository. An example is

Text 68:
    Statement: A big triangle, small triangle, and circle are in the house. The big triangle and the circle each kiss the small triangle, wave, and then leave the room. How are the shapes related?
    Alternative 1. The big triangle and the circle are parents of the small triangle.
    Alternative 2. The big triangle and the small triangle are parents of the circle.

Third, an approximate manual estimation showed a better result than the real experiment run on the computer. This may suggest that the data needed for the correct selection are available in EnSemS, but the algorithm of the alternative selection may need some update, which we hope to perform in the near future.

5. Conclusion

- We obtained a reasonably good result on the Triangle-COPA test: 80.5% of questions were answered correctly. Our method implements the full scenario of text understanding, which includes morpho-syntactic and semantic analysis and a series of inferences. As opposed to other Triangle-COPA solutions, where the processing was performed on manually prepared simplified logical formulae, our analyzer applies to raw text.
- The choice of the correct alternative was done in two steps. First, Enhanced Semantic Structures of the texts to be compared were performed. These structures contain all the inferences, both strict and plausible, that can be made based on the knowledge available. Then semantic agreement between these structures was calculated, and the alternative showing higher agreement with the initial text was chosen. Importantly, the agreement is formulated in intuitively clear terms and can be checked by humans.
- Satisfactory results obtained on Triangle-COPA prove that a general scope semantic analyzer can solve specific problems, provided it is supplied with good-quality knowledge. Explicit knowledge based on the concepts meaning and common sense knowledge plays a key role.
- The knowledge repository of the semantic analyzer was incrementally enriched. The knowledge acquired during the Triangle-COPA experiments is not task-specific and can be employable in multiple scenarios.
- An important result is the discovery that common sense knowledge needed for inferences even in a very narrow domain cannot be acquired without detailed examination of the texts of the domain.

6. Acknowledgements

References

1. *Boguslavsky I.* (2011) "Semantic Analysis Based on Linguistic and Ontological Resources", Proceedings of the 5th International Conference on the Meaning—Text Theory. Barcelona. Igor Boguslavsky and Leo Wanner (Eds.). P. 25–36.
2. *Boguslavsky I., V. Dikonov, L. Iomdin, A. Lazursky, V. Sizov, S. Timoshenko.* (2015), Semantic Analysis and Question Answering: a System Under Development. In: Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2015), p. 62.

3. *Boguslavsky I.* (2017), Semantic Descriptions for a Text Understanding System. In: Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2017), p. 14–28.

4. *Boguslavsky I., Frolova T., Iomdin L., Lazursky A., Rygaev I., Timoshenko S.* (2018), Semantic analysis with inference: high spots of the football match. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018", Moscow, May 30—June 2.

5. *Boguslavsky I., Frolova T., Iomdin L., Lazursky A., Rygaev I., Timoshenko S.* (2019), Knowledge-based approach to Winograd Schema Challenge. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2019". Moscow, May 29—June 1.

6. *Bos J.* (2008), Wide-Coverage Semantic Analysis with Boxer. In Semantics in Text Processing. STEP 2008 Conference Proceedings. College Publications. pp. 277–286.

7. *Bryl V., Guiliano C., Serafini L., Tymoshenko K.* (2010), Using background knowledge to support coreference resolution. In Proceedings of the 19th European Conference on Artificial Intelligence, pp. 759–764.

8. *Ge R., Mooney R.* (2005), A Statistical Semantic Parser that Integrates Syntax and Semantics. Proceedings of the Ninth Conference on Computational Natural Language Learning. Ann Arbor, MI, pp. 9–16, June 2005.

9. *Gordon, A. S.* (2016), Commonsense Interpretation of Triangle Behavior, In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16).

10. *Gordon A. S., Hobbs J. R.* (2011), A Commonsense Theory of Mind-Body Interaction, Proceedings of the 10th Symposium on Logical Formalizations of Commonsense Reasoning, AAAI Spring Symposium Series.

11. *Heider F., Simmel M.* (1944), An experimental study of apparent behavior. The American Journal of Psychology 57(2):243–259.

12. *Hobbs J. R., Gordon A.* (2010), Goals in a Formal Theory of Commonsense Psychology, In A. Galton and R. Mizoguchi (eds.), Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010), IOS Press, Amsterdam, pp. 59–72.

13. *Hobbs J. R., Sagae A., Wertheim S.* (2012), Toward a Commonsense Theory of Microsociology: Interpersonal Relationships, In M. Donnelly and G. Guizzardi (eds.), Formal Ontology in Information Systems: Proceedings of the Seventh International Conference (FOIS 2012), IOS Press, Amsterdam, Netherlands, pp. 249–262.

14. *Kalluri P., Gervás P.* (2017), Affinity-based Interpretation of Triangle Social Scenarios. In Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017), pp. 640–647.

15. *Lee H., Peirsman Y., Chang A., Chambers N., Surdeanu M., Jurafsky D.* (2011), Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, pp. 28–34.

16. *Levesque H., Davis E., Morgenstern L.* (2012), The Winograd schema challenge. In Proceedings of the Knowledge Representation and Reasoning Conference.

17. *Maslan N., Roemmele M., Gordon A. S.* (2015), One hundred challenge problems for logical formalizations of commonsense psychology. In Twelfth International Symposium on Logical Formalizations of Commonsense Reasoning (Commonsense-2015).

18. *Meadows B., Langley, P., Emery, M.* (2014), An abductive approach to understanding social interaction. Advances in Cognitive Systems 3:87–106.

19. *Montazeri N., Hobbs J. R.* (2011), Elaborating a Knowledge Base for Deep Lexical Semantics, In J. Bos and S. Pulman (eds.), Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011), January 2011, pp. 195–204.

20. *Montazeri N., Hobbs J. R.* (2012), Axiomatizing Change-of-State Words, In M. Donnelly and G. Guizzardi (eds.), Formal Ontology in Information Systems: Proceedings of the Seventh International Conference (FOIS 2012), IOS Press, Amsterdam, Netherlands, pp. 221–234.

21. *Morgenstern L.* (2001), MidSized Axiomatizations of Commonsense Problems: A Case Study in Egg Cracking. Studia Logica, April 2001, Volume 67, Issue 3, pp. 333–384.

22. *Mueller E.* (2006), Commonsense reasoning. Elsevier, Morgan Kaufmann Publishers.

23. *Nirenburg, S., Raskin, V.* (2004), Ontological Semantics. The MIT Press. Cambridge, Mass., London, England.

24. *Poesio M., Day D., Artstein R., Duncan J., Eidelman V., Giuliano C., Hall R., Hitzeman J., Jern A., Kabadjov M., Yong Wai Keong, Mann G., Moschitti A., Ponzetto S., Smith J., Steinberger J., Strube M., Jian Su, Versley Y., Xiaofeng Yang, Wick M.* (2007), ELERFED: Final report of the research group on Exploiting Lexical and Encyclopedic Resources For Entity Disambiguation. Technical report, Summer Workshop on Language Engineering, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD.

25. *Ponzetto S. P., Strube M.* (2006), Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In Proceedings of the Human Language Technology Conference and Conference of the North American Chapter of the Association for Computational Linguistics, pp. 192–199.

26. *Ponzetto S. P., Strube M.* (2007), Knowledge derived fromWikipedia for computing semantic relatedness. Journal of Artificial Intelligence Research, 30:181–212.

27. *Poon, H., & Domingos, P.* (2009), Unsupervised semantic parsing. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Volume 1 EMNLP 09 (p.1).

28. *Rahman A., Ng V.* (2011), Coreference resolution with world knowledge. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 814–824.

29. *Roemmele M., Bejan, C. Gordon A.* (2012), Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning.

30. *Uryupina O., Poesio M., Giuliano C., Tymoshenko K.* (2011), Disambiguation and filtering methods in using Web knowledge for coreference resolution. In Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference, pp. 317–322.