Semantic Comprehension System for F-2 Emotional Robot

Artemy Kotov[1(✉)], Nikita Arinkin[1], Alexander Filatov[2], Liudmila Zaidelman[1], and Anna Zinina[1]

[1] National Research Center "Kurchatov Institute", pl. Kurchatova, 1, Moscow, Russia
kotov_aa@nrcki.ru
[2] Samsung R&D Institute Rus, Dvintsev 12/1, Moscow, Russia

Abstract. Within the project of F-2 personal robot we design a system for automatic text comprehension (parser). It enables the robot to choose "relevant" emotional reactions (output speech and gestures) to an incoming text – currently in Russian. The system executes morphological and syntactic analysis of the text and further constructs its semantic representation. This is a shallow representation where a set of semantic markers (lexical semantics) is distributed between a set of semantic roles – structure of the situation (fact). This representation may be used as (a) fact description – to search for facts with a given structure and (b) basis to invoke emotional reactions (gestures, facial expressions and utterances) to be performed by the personal robot within a dialogue. We argue that the execution of a relevant emotional reaction can be considered as a characteristic of text comprehension by computer systems.

Keywords: Natural language comprehension · Syntactic parser · Text analysis

1 The Problem of Text Comprehension

Since the Chinese room argument [1] the problem of automatic text comprehension became one of the cornerstone questions in Computer Linguistics and Artificial Intelligence. In his original publication Searle has argued that no artificial computer can understand natural text in a way people do. Since then – numerous critics have suggested the architectures of "understanding", which apply to human comprehension and also can be implemented on a basis of a computing machine. The development of robot companions has opened another view into the problem: human infers, or rather – feels being understood, basing on replies and emotional reactions from the interlocutor. This view has very little to do with internal architecture of the software. As suggested by our communication with dogs – humans feel being understood, basing solely on the behavioral responses from dogs, without any speech interaction. Following these observation we develop the project of F-2 companion robot with the emotional reactions as the main

Design of the syntactic parser is supported by RFBR grant 16-29-09601 ofi_m, and design of the F-2 robot is executed within the research program of "Kurchatov Institute".

A.V. Samsonovich and V.V. Klimov (eds.), *Biologically Inspired Cognitive Architectures (BICA) for Young Scientists*, Advances in Intelligent Systems and Computing 636,
DOI 10.1007/978-3-319-63940-6_17

component to maintain contact with humans. Unlike dogs, F-2 processes natural speech and suggests emotional comments and reactions. To enable this capacity we develop a syntactic parser, which constructs shallow semantic representation, suitable to select a relevant emotional reaction and suggest the distribution of emotional roles in a situation.

2 Approaches to Text Comprehension and Sentiment Analysis

The task of extracting emotional evaluations from text is usually solved using the bag-of-words method. The text (as a whole or each sentence separately) is represented as a set of word forms or lexemes; the position of the word in the text or in the sentence is not taken into account. This approach is presented in a number of papers [2–5]. In order to minimize numerous shortcomings of this approach researchers often use the bag-of-n-grams – an unordered set of tuples consisting of *n* consecutive words [2, 6]. Dialogue systems also often focus on individual words or n-grams to choose their answers. However, intelligent systems for text analysis and user interaction should determine the role of a particular character of the text in a situation, attribute the character to a certain syntactic or semantic valence and recognize the situation frame.

Apart from a complete syntactic analysis of the text there are several alternatives that extract fact structure with the help of partial analysis. For example, in [7] text is parsed into so-called *T-expressions*: three element tuples <*subject, relation, object*>. T-expressions are used as the basis both for sentiment analysis task and for other applications such as automatic question answering.

A similar approach is used in [8]. The suggested system divides each fact into four parts: (a) an object from the thesaurus, (b) the type of syntactic relationship between the object and the member of the sentence syntactically associated with the object, (c) member of the sentence associated with the object, and (d) presence of a negation. Syntactic relationships are extracted by Tomita parser – a tool for context-free grammars – equipped with 50 rules. The resulting syntax group of four elements is a subject to further analysis and evaluation.

ABBYY Compreno parser [9] for each sentence in the source text constructs a tree, whose nodes have not only grammatical characteristics, but also attributed semantic classes from the ontology. The resulting tree is used as a basis for facts extraction. Once a rule is triggered, the proposition associated with the object is extracted. Thus, the semantics of the sentence is presented as a set of propositions related to given objects. The extracted collection of facts enriches the tree with new characteristics, which in turn triggers other rules, allowing new propositions to be extracted.

In Sentilo project [10], a complex linguistic model that includes a variety of linguistic resources and tools is used to perform sentiment analysis. Evaluation of the sentence is computed from RDF graph. The graph nodes are syntactic elements and their ontological classes, the edges – are the relations derived from several of linguistic theories and ontologies. Emotional evaluation is calculated for significant actants and for the whole predicative structure. The averaged positive and negative evaluations are used as the final sentiment score for the whole sentence.

We can see that the main tendency in the field is to construct complete syntactic trees and use extracted facts as reference structures. In our work we separate syntactic and semantic representations, as well as process semantic representations by numerous scripts, responsible for emotional arousal and robot's reactions.

3 Parser Architecture

Parser is designed to construct semantic representation, suitable to apply an emotional reaction towards a specific participant. In particular, some situations can be considered as 'terrible' with possible aggression to be conveyed to *A*, or as 'pity' with possible compassion to be expressed to *B*. To construct the representation the parser implements morphological, syntactic and semantic processing of the incoming text as suggested by theoretical linguistic models, e.g. [11]. The parser is written in C#, the grammar is in syntXML format and the dictionary is stored in SQL database. On each step of processing the parser may upload the results of analysis to an SQL database or transfer them to other software components, e.g. to F-2 robot (see Fig. 1).

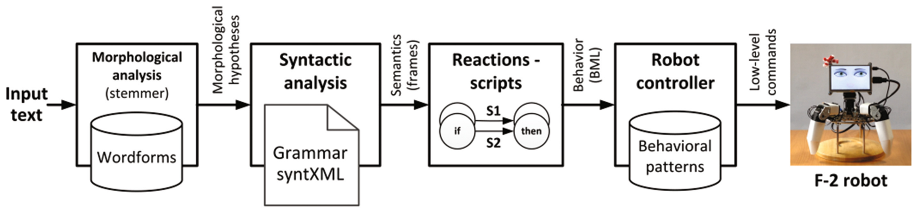

Fig. 1. Architecture of the text analysis and reaction transfer to F-2 robot

3.1 Morphological Processing: Stemmer and Dictionary

Stemmer relies on a database, which keeps all wordforms and grammemes for 48,000 lemmas. The dictionary is based on OpenCorpora project [12]. 28,000 words in the dictionary are annotated by semantic markers (from 1 to 18 markers per word, average 2). Markers are assigned (a) basing on hyperonyms – and represent the semantic class of the word, and (b) basing on the sensitive semantic features, for example, 'intensity' can be emotionally relevant in phrases like *Why do you push me?* [13]. Unlike traditional ontologies, a word may keep semantic markers from different classes: *bank* has the markers for 'organization', 'building' and 'abstract container'. This polysemy allows us to simulate "situational effects", where a word meaning may be shifted depending on the situation or by the emotion to be invoked by semantics – top-down emotional processing [14]. So different reference frames of scripts (units for inferences and emotional reactions) may address different focal markers in the semantics of a word. In addition to polysemy, we describe lexical homonymy: markers can be assigned to several meanings of a word (like $bank_1$ – *financial institution, building* vs. $bank_2$ – *river bank*). A script will also select the meaning, which fits better to its reference frame.

3.2 Syntactic Processing: syntXML

We develop a dependency parser with left-to-right approach. After the morphological component assigns morphological hypotheses to each segment (token) – the segment is added to a syntactic stack, and the syntactic component tries to reduce the stack head with the grammar rules. A rule is defined as a possible reduction, where the right-hand side can be reduced to the left-hand head $\boldsymbol{h}$ (1). Head $\boldsymbol{h}$ can also be a member of the right-hand side and subordinate all other right-hand side segments (2).

$$\boldsymbol{h} \rightarrow < a, b, \ldots n > \tag{1}$$

$$\boldsymbol{h} \rightarrow < a, b, \boldsymbol{h}, \ldots n > \text{ or } < a, b, \boldsymbol{h}^{\text{head}}, \ldots n > \tag{2}$$

Right-hand side of the rule may have a variable number of segments (1 or more) as well as optional segments. The grammar contains 490 rules, written on a specially designed syntXML language [15]. Application of each rule is evaluated, scores are calculated on the basis of SynTagRus treebank [16] – total score is calculated for a stack. Once a rule is applied, it may assign a semantic role to a segment. We rely on this list of semantic roles, suggested in [17]: *ag* (agent), *pat* (patient) etc. The predicate is assigned to *p* semantic role. This procedure locates clauses in a tree – where each clause consists of a predicate *p* and a number of its actants.

For each type of homonymy appearing within a stack – lexical or morphological ambiguity, ambiguity of rules application – the stack is duplicated. So on each step the parser works with n stacks with highest total scores (for standard tasks we set $n = 1000$), stacks with lower scores are discarded.

3.3 Semantic Processing: Scripts

Each tree is subdivided by syntactic rules into clauses – a predicate and a list of actants. For each actant – semantic markers of the head word (noun, verb) and subordinate words (adjectives, adverbs) are extracted from the dictionary and assigned to the semantic role of the actant. This constructs a semantic frame, representing a single clause (Table 1).

Table 1. Semantic representation (frame) for the utterance *A real man is always interested in the life of the beloved girl*

p (predicate)	ag (agent)	pat (patient)
think, pay-attention, frequently	object, somebody, man, positive	abstract, time-period, existence, object, somebody, woman, of-minimal-age, positive

As suggested by M. Minsky [18], artificial agents may have numerous models for drives and emotions – proto-specialists, which compete to control the agent behavior. Further A. Sloman [19] has suggested CogAff architecture, where scripts, responsible for emotional processing, compete with scripts, devoted to rational (deliberative) procedures and meta-management (reflexive thinking). We rely on the list of scripts for emotional processing represented in [20]. It includes 13 scripts for negative situations: DANGER, APPROPR ("Appropriation"), SUBJV ("Subjectivity" – e.g. 'all he thinks about is himself!') etc., and 21 scripts for positive situations: CONTROL, CARE, COMFORT, ATTENTION (e.g. 'they all adore you!'), APPROVAL (e.g. 'you did it like a real man!') etc. The semantic representation in Table 1 activates the following scripts:

- PLAN: Somebody plans something frightening against me – 'man makes some evil plans against woman'
- SUBJV ("Subjectivity"): Somebody is narrow-minded, thinks only about one thing – 'all men think about are women'.
- ATTENTION: Subject is pleased, because somebody pays an attention at him – 'woman is happy because of the men's attention'.
- APPROVAL: Somebody acts like a hero, does something right – 'real men do it right to pay attention'.

Although APPROVAL and ATTENTION are more relevant, we do not consider the activation of PLAN and SUBJV as false positive. These reactions can be used (a) to generate latent behavioral patterns of the robot (where it is happy but afraid to attract attention), (b) to express mood – where a "depressive" robot prefers negative reactions, (c) by the mechanism of irony to generate sarcastic responses and simulate the sense of humor [21].

Scripts are also helpful to solve syntactic homonymy: if numerous trees are exported by the syntactic analysis, the semantic processor chooses the tree, which has the highest degree of similarity with reference frames of the scripts – which corresponds to a more standard situation or is more likely to invoke emotions.

4 Robot and the Transfer or Behavior

As suggested in [22] the development cycle for a computer agent should include (a) collection of human behavior into a multimodal corpus, (b) design of the behavioral model, (c) implementation of the model in a computer agent, simulation of the observed behavior and (d) test of the simulated behavior. In our studies we collect and annotate records of the multimodal behavior within the project of the Russian Emotional Corpus – REC [23]. We also design software to operate a robot companion and F-2 robot – as a demonstrator of the software. We observe behavioral patterns – gestures and facial expressions, typical to express certain communicative functions [24], draw the patterns in 3D model and save them in a library to be accessed by the robot. Each script is assigned to one or several behavioral reactions: utterance pattern and a BML record – Behavior Markup Language [25]. Sematic analysis of an input text activates one or several scripts, which send their BMLs to robot for execution. BMLs can compete for the robot actuators, which results in richer and more compound behavior, where numerous reactions are expressed at the same time.

As shown, text comprehension can be an important component within the design of a robot companion, which maintains emotional contact with a human. Text understanding here is implemented by the construction of a shallow semantic representation and the selection of a relative emotional reaction. This representation can also serve as a basis for knowledge extraction and semantic search.

References

1. Searle, J.: Minds, brains, and programs. Behav. Brain Sci. **3**, 417–424 (1980)
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79–86. Association for Computational Linguistics (2002)
3. Su, F., Markert, K.: From words to senses: a case study of subjectivity recognition. In: Proceedings of the 22nd International Conference on Computational Linguistics, vol. 1, pp. 825–832. Association for Computational Linguistics, Manchester (2008)
4. Chetviorkin, I.I.: Testing the sentiment classification approach in various domains — ROMIP 2011. In: Computational Linguistics and Intellectual Technologies, vol. 2(11), pp. 15–26. RSUH, Moscow (2012)
5. Fang, L., Huang, M.: Fine granular aspect analysis using latent structural models. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, vol. 2, pp. 333–337. Association for Computational Linguistics, Jeju Island (2012)
6. Poroshin, V.: Proof of concept statistical sentiment classification at ROMIP 2011. Computational Linguistics and Intellectual Technologies, vol. 2(11), pp. 60–65. RSUH, Moscow (2012)
7. Katz, B.: From sentence processing to information access on the world wide web. In: AAAI Spring Symposium on Natural Language Processing for the World Wide Web (1997)
8. Mavljutov, R.R., Ostapuk, N.A.: Using basic syntactic relations for sentiment analysis. In: Computational Linguistics and Intellectual Technologies, vol. 2(12), pp. 91–100. RSUH, Moscow (2013)
9. Anisimovich, K.V., Druzhkin, K.J., Minlos, F.R., Petrova, M.A., Selegey, V.P., Zuev, K.A.: Syntactic and semantic parser based on ABBYY Compreno linguistic technologies. In: Computational Linguistics and Intellectual Technologies, vol. 2(11), pp. 91–103. RSUH, Moscow. (2012)
10. Recupero, D.R., Presutti, V., Consoli, S., Gangemi, A., Nuzzolese, A.G.: Sentilo: frame-based sentiment analysis. Cogn. Comput. **7**, 211–225 (2014)
11. Melčuk, I.A.: The experience of the theory of linguistic models "MEANING <=> TEXT" (in Russian). Languages of the Russian Culture, Moscow (1999)
12. Bocharov, V.V., Alexeeva, S.V., Granovsky, D.V., Protopopova, E.V., Stepanova, M.E., Surikov, A.V.: Crowdsourcing morphological annotation. In: Computational Linguistics and Intellectual Technologies, vol. 12(19), pp. 109–114. RSUH, Moscow (2013)
13. Apresian, V.Yu.: Implicite aggression in the language (in Russian). In: Kobozeva, I.M., Laufer, N.I., Selegey, V.P. (eds.) Computational Linguistics and Intellectual Technologies, pp. 32–35. Nauka, Moscow (2003)
14. Clore, G.L., Ortony, A.: Cognition in emotion: always, sometimes, or never? In: Lane, R.D., Nadel, L. (eds.) Cognitive Neuroscience of Emotion, pp. 24–61. Oxford University Press (2000)
15. Kotov, A., Zinina, A., Filatov, A.: Semantic parser for sentiment analysis and the emotional computer agents. In: Proceedings of the AINL-ISMW FRUCT 2015, pp. 167–170 (2015)

16. Boguslavsky, I.M., Iomdin, L.L., Chardin, I.S., Kreidlin, L.G.: Development of a dependency treebank for russian and its possible applications in NLP. In: Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-2002), vol. III, pp. 852–856. European Language Resources Association, Paris (2002)
17. Fillmore, C.J.: The case for case. In: Emmon, B., Harms, R. (eds.) Universals in Linguistic Theory, pp. 1–68. Holt, Rinehart & Winston, New York (1968)
18. Minsky, M.L.: The Society of Mind. Touchstone Book, New-York, London (1988)
19. Sloman, A., Chrisley, R.: Virtual machines and consciousness. J. Conscious. Stud. **10**, 133–172 (2003)
20. Kotov, A.A.: Mechanisms of speech influence (in Russian). Kurchatov Institute, M. (in print)
21. Kotov, A.: Accounting for irony and emotional oscillation in computer architectures. In: Proceedings of International Conference on Affective Computing and Intelligent Interaction ACII 2009, pp. 506–511. IEEE, Amsterdam (2009)
22. Rehm, M., André, E.: From annotated multimodal corpora to simulated human-like behaviors. In: Modeling Communication with Robots and Virtual Humans, pp. 1–17 (2008)
23. Kotov, A., Budyanskaya, E.: The Russian emotional corpus: communication in natural emotional situations. In: Computational Linguistics and Intellectual Technologies, vol. 11(18), pp. 296–306. RSUH, Moscow (2012)
24. Kotov, A.A., Zinina, A.A.: Functional analysis of the nonverbal communicative behavior (in Russian). In: Computational Linguistics and Intellectual Technologies, vol. 1(14), pp. 299–310. RSUH, Moscow (2015)
25. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., Vilhjálmsson, H.: Towards a common framework for multimodal generation: the behavior markup language. In: Intelligent Virtual Agents, pp. 205–217 (2006)