A New Social Robot for Interactive Query-Based Summarization: Scientific Document Summarization

Marzieh Zarinbal[1], Azadeh Mohebi[1](✉), Hesamoddin Mosalli[1], Razieh Haratinik[1], Zahra Jabalameli[1], and Farnoush Bayatmakou[2]

[1] Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran
{zarinbal,mohebi}@irandoc.ac.ir, mosalli.h@gmail.com, haratinik@gmail.com, jabalameli.software@gmail.com
[2] Amirkabir University of Technology, Tehran, Iran
f.bayatmakou@aut.ac.ir

Abstract. The extractive summartization methods try to summarize a single or multiple documents based on informative sentences exactly as they appear in source(s). One method to choose these sentences is to use users' query, which could be problematic in many cases, specially in scientific context. One way to tackle this challenge is to gather more information about the user and his preferences. Therefore, in this paper we propose a novel framework to use the users' feedbacks and a social robotics platform, Nao robot, has been adapted as an interacting agent. This agent has multiple communication channels and could learn the user model and adapt to his/her needs via reinforcement learning approach. The whole approach is then studied in terms of how much it is able to adapt based on user's feedback, and also in terms of interaction time.

Keywords: Interactive robots · Social robots · Automatic text summarization · Facial expression

1 Introduction

The number of scientific documents are growing exponentially and keeping track of them, organizing and finding relevant documents have become challenging tasks in conducting research. Automatic text summarization methods could be used to tackle these challenges and the main aim of these methods is to produce a short and meaningful version of a source document. One approach to start summarization and to find relevant documents and sentences is using query. In query-based text summarization, the query is usually considered as the only source of knowledge, based on which the final summary is generated [8]. However, the query usually contains limited terms and users may not be able to express their information needs in the form of terms. This could be more complex when we are dealing with scientific documents, as the relevancy of a document or

A. Ronzhin et al. (Eds.): ICR 2019, LNAI 11659, pp. 330–340, 2019.
https://doi.org/10.1007/978-3-030-26118-4_32

a sentence is highly related to the context. Thus, the summarization methods needs to use additional sources of information. Finding the users' opinion about the retrieved results (relevance feedback) could be an additional source, but it has its own drawbacks, i.e, reluctancy of users to provide feedback and prolong their search sessions and thus, causing high computing cost and long response times [19].

To tackle this issue, we propose using a social robot, called RoboDoc, to assist researchers in generating a summary for any given query. RoboDoc is a social assistant robot designed to gather the needed information via facial expression and voice commands and to encourage its users to have more interactions. In a broader application, RoboDoc, as conceptualized by Mohebi et al. [15], can interact with researcher in order to obtain relevant additional information.

The rest of this paper is organized as follows: Sect. 2 presents a brief introduction on the concepts and the main efforts. The proposed framework is discussed in Sect. 3. Some preliminary experiments on the proposed framework are addressed in Sect. 4. Finally, conclusion is stated in Sect. 5.

2 Background

2.1 Interactive Query-Based Summarization

Automatic text summarization is defined as creating a meaningful content based on original document(s) or text(s) using computers to help users to grab the key information or general idea in a relatively short time [3]. Summarization could be abstractive or extractive; in extractive summarization, the important sentences are extracted from the source document(s) and simply put together to generate summary. However, in abstractive summarization the source document(s) is understood and then the summery is written [2]. Summarization could be begin based on user query or could be generic. In query-based summarization the summary is generated based on users' information needs, but generic summarization methods try to reflect the general idea of the source document(s) regardless of users' needs [4]. The challenge is to decide which sentence from which document is more informative and significant to be included in the summary. Thus, sentences have to be scored and the ones with the highest scores are likely to be used. Using user's query is one way to score the sentences, in which the sentences containing query expressions are given higher scores [2]. This could cause information loss in cases in which (a) the query contains limited terms, (b) users could not properly express their information needs in the form of terms, or (c) users may have special needs and backgrounds. In addition, using this approach could be challenging, because, these sentences must be analyzed and scored based on the documents' context, user needs to have more specific knowledge to identify proper query terms and he may not be able to provide his required information in single query or a single search session. Various methods have been developed to tackle these challenges such as [1,2,6,16].

One of the well-used approaches is to use an additional source of information. Interacting with users in long-term along with acquiring and evaluating their feedback could provide such information, improve the quality of the summary, and satisfy users' needs [20,22]. Relevance feedback (RF) is defined as the user's feedback on the relevance of retrieved results, i.e, user marks each result as relevant or non-relevant. Using this information, the system could re-evaluate the sentences' scores during one or more search sessions and could help the system to develop or create users model, track his evolving and changing information need and help user to refine his understanding. However, in many cases RF could cause problems; users are generally reluctant to provide feedback and prolong their search sessions, and it could cause a high computing cost and long response time [19].

To tackle these issues and in this paper, we propose using social robots. A social robot is designed to gather the needed information via any natural communication channel and to encourage users to have more interaction.

2.2 Social Robots

In social interactions we usually use various ways to express our views or emotions, such as facial expression or body language. Using social robots enables system designers to developed more effective user-adapted systems [13]. A social robot interacts with user via various communication channels (Multimodality) and learns or modifies the user model (learnability). Being able to adapt to the user's needs by changing the operational parameters automatically (adaptivity) is another important requirements of social robots.

Hegel et al. [9] defined three classes of applications for social robots: specialized applications, public applications, and individual applications. In another view, the application areas of social robots could include but not limited to assistive robots, domestic robots, healthcare robots, and General social robotics platforms such as Pepper or Nao. Social assistant robots must be able to perceive and interpret users' behavior and send signals to users in order to provide feedback and allow them to interact in a transparent manner. These robots must also operate at human interaction rates. Therefor, having embodiment, morphology and personality, modeling user and perceiving his behavior via speech, gestures, and facial recognition, showing emotions, managing dialogues and learning is essential [7].

Based on above discussions, Mohebi et al. [15] introduced a framework for a social robot as a researcher's personal assistant, named RoboDoc. RoboDoc is designed to offer several services during research, including capturing and organizing researcher's ideas, applying query expansion and enrichment, and integrating the retrived results into meaningful pieces of information [15]. In this paper and in order to improve these services we have developed a system to summarize scientific documents based on researcher's query and modify the scoring procedure using the relevance feedback. This feedback is acquired via researcher's facial expression.

3 Proposed Approach

The proposed approach for query-based summarization is a user-centric interactive approach in which the user plays a key role, and the interacting agent is social robot with sociability and learning ability. The conceptual framework for this task involves consecutive interactions between user and social assistant robot. Robot generates a simple initial summary based on user's query, and tries to improve the summary based on sequence of interaction with the user, in terms of a set of simple questions. The user gives feedback to the robot through verbal or non-verbal communication such as facial expression. In Fig. 1 the conceptual framework for the interaction is illustrated.

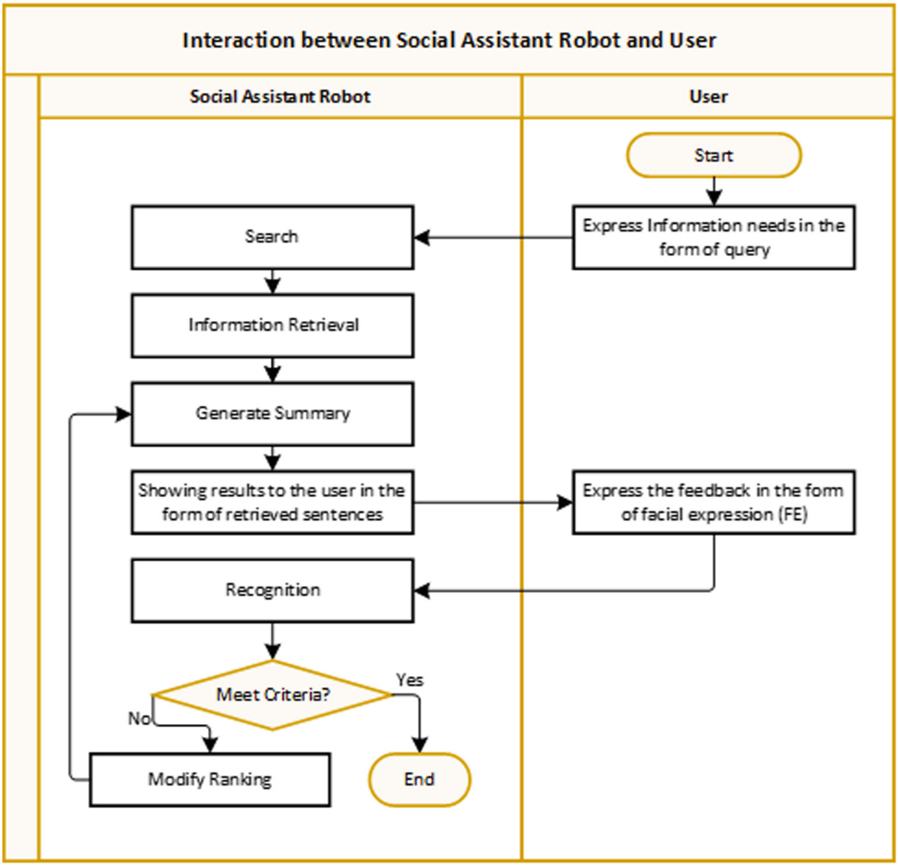


Fig. 1. Proposed conceptual framework for human-robot interaction.

Based on this conceptual framework, the user needs to give feedback based on the sentences introduced by robot and Reinforcement Learning (RL) approach is used to model this interaction. RL have been used as a practical modelling approach for extractive text summarization [10,11,17,18]. However, these studies contain some drawbacks: (a) user is not involved in the summarization task, (b) most of the RL-based methods are useful for multi-document text summarization without any given query, and (c) in previous studies, a reference summary is used as source for reward at every iteration.

In this research, however, we do not use any reference summary and instead user feedback is considered as the reward. That is, user feedback is encoded as the reward/punishment received from the environment, and the robot tries to tune the summary based on the feedbacks.

Assume that database $\mathcal{D}$ contains N distinct sentences and each sentence $x_i \in \mathcal{D}$ can be part of a summary. The problem is encoded as RL based on the following notation:

- State (S_t): summary composed of M sentences, with their scores,
- Action (a_t): action taken to modify summary S_t to S_{t+1},
- Reward (r_t): feedback received from user,
- Value function ($Q(S_t, a_t)$): gained value performing action a_t at state S_t.

So at each state, S_t, a summary composed of M different sentences is generated:

$$S_t : \quad < x_i, score^t(x_i) > \tag{1}$$

$$i = 1, \ldots, M, \quad \text{and} \quad x_i \in \mathcal{D}$$

The sentences of the summary are presented to user and user gives feedback in terms of facial expression, to express his opinion about the summary. User feedback is then used by the robot to modify the summary. This interaction continues until a specific stopping criteria is met. The action, a_t, is a re-scoring scheme to change summary S_t to S_{t+1}. The terminal state is denoted by S_F. If state S_t is the terminal state, then the action could be encoded as *Finish*.

$$a_t = \begin{cases} Rescore(\mathcal{D}) & \text{if } S_t \neq S_F \\ Finish & \text{if } S_t = S_F, \end{cases} \tag{2}$$

where, the re-scoring scheme at each state changes the score of each sentence in $\mathcal{D}$ into a new score based on user feedback. Then, the best M sentences is selected based on a policy π that impose to choose the best M sentences with maximum score at each interaction.

At S_t, the robot represents a sentence x^* to the user. The user gives feedback about this sentence as a reward r_t, and this reward is considered into the re-scoring scheme. Thus, the re-scoring scheme is defined as follows:

$$score^{t+1}(x_i) = \begin{cases} score^t(x_i) * e^{sim(x^*, x_i)} & \text{if } r_t \geq 0 \\ score^t(x_i) * e^{1-sim(x^*, x_i)} & \text{if } r_t < 0, \end{cases} \tag{3}$$

where $r_t \in \{-1, 0, 1\}$ is the feedback received from the user. r_t is 1, when user likes the sentences, -1, when user does not like the sentence, and 0 when user is neutral. The re-scoring scheme defined in (3) causes the score to increase exponentially. The incremental change in the score will not lead to any integer overflow error, since the interaction between user and robot is limited. Moreover, the score of each sentence always increases from one state to another with different rate. The exponential form in (3) amplifies the score of sentences similar to the one user liked so far, while when user did not like a sentence, the score of sentences similar to that, is increased with less rate.

At the initial stage, before receiving any feedback from the user, the most similar sentences to the query are retrieved. The similarity at this stage is calculated based the Word2Vec, CBOW learned model, as described in [14]. The Word2Vec word embedding model is a well-known vectorized representation that is widely used in calculating word and document similarity context.

The value function is defined based two distinct criteria: how much the summary is close to user's opinion and the initial query (extrinsic measure), and the quality of the resulting summary in terms of information redundancy (intrinsic measure). Thus, the value function can be defined as:

$$Q(S_t, a_t) = \sum_{x_i \in S_t} \alpha \; score^t(x_i) + (1-\alpha) \max_{x_j \in S_t \setminus x_i} sim(x_i, x_j) \tag{4}$$

where $sim(x_i, x_j)$ calculates the cosine similarity between two sentences, and α is a balancing parameter. The definition of value function is very close to the definition of summary quality measure based Maximal Marginal Relevance (MMR) [5] method, which is mainly used for query-based extractive summarization.

4 Experimental Results

To implement the proposed interactive summarization approach (Fig. 1), the used social robotics platform and database of textual documents are discussed in this section. In addition, the interactive scenario and some interaction results are reported, and finally, the proposed approach is evaluated in terms of interaction time and the effect of interaction in the summarization task.

4.1 Social Robotics Platform

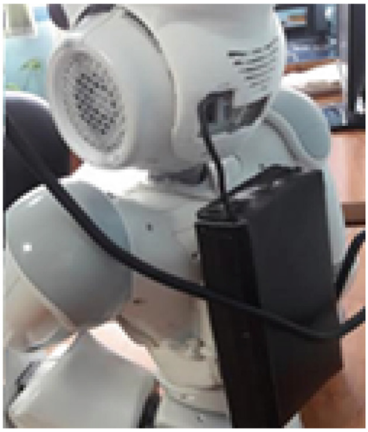
For social robotics platform, Nao robot, Robocup edition, has been adopted. As the user feedbacks are acquired via facial expression, Nao's vision system needs to recognize the expression of user. Nao's original camera is weak, so we embedded an external camera. Furthermore, in order to make the interaction more practical and effective, the robot reads the generated summary and shows it through the embedded data projector. Therefore, we modified Nao's original head and designed a new hat to embed both camera and mini data projector, as shown in Fig. 2(a).

The data projector and camera are connected with a Raspberry Pi board. This board is also attached to Nao's body and is connected to the robot through

TCP/IP connection, as shown in Fig. 2(b).The mini data projector embedded in Nao's head is "Aiptek MobileCinema i55 iphone 5/5s projector",which is connected to the Raspberry board and is automatically controlled through GPIO pins.

For face detection, we have used MTCNN face detection method based on [21] using Tensorflow Google's Deep Learning and OpenCv framework. The facial expression module is implemented based on [12], with accuracy 78 %. In addition, the robot needs to talk to the user through a text-to-speech (TTS) system and understand user's speech using a speech-to-text (STT) system. For this purpose we have used the embedded TTS system in Nao, and Google STT system.

(a) Nao's new hat with camera and data projector



(b) Raspberry Pi board attached to Nao' back

Fig. 2. Modified Nao to embed data projector, new camera, and Raspberry Pi.

4.2 Dataset for Summarization

A dataset of 16186 scientific articles (142386 sentences) obtained from Web of Science (WoS) is used for summarization. All articles are in the area of Artificial Intelligence, Robotics and Control, and Image Processing as categorized by WoS. We used the title and abstract of each article as the source of information and also for training the Word2Vec to build the vectorized representation for each word.

4.3 Test and Evaluation

The system is implemented using the modified robotic platform and the dataset described before. The text processing and retrieval tasks are handled with a Corei7 laptop connected to the robot and raspberry Pi through a TCP/IP connection. Each summary contains 10 distinct sentences. At the initial stage, the top 30 most similar sentences to the query are considered and at each interaction, the re-scoring scheme is applied. That is, after retrieving the 30 most relevant sentences to the user's query, at each interaction, robot reads a sentence with the highest score to the user, and the user gives a feedback to the robot through a facial expression (FE). The user is an expert in scientific domain and is aware that his facial expressions are being analyzed. The robot classifies the facial expression into three distinct types: 1: like the sentence, −1: dislike the sentence, and 0: neutral. Then, the robot re-scores all sentences based on this feedback, and reads another sentence with the highest score to the user. We have considered 5 interactions between the robot and the user before giving the final summary. Table 1 reports some parts of the result for the query "Image Processing".

Table 1. User-robot interaction for the query "Image Processing".

Stage	FE	Sentence read by robot
0	−1	The features were extracted from the spectrogram of the speech signal using image processing techniques
1	0	Image restoration step is important in many image processing applications
2	0	Digital image retrieval is one of the major concepts in image processing
3	1	The Weibull manifold in low level image processing: an application to automatic image focusing
4	0	Content based image retrieval plays a, significant role in the image processing field
Final summary		Digital image retrieval is one of the major concepts in image processing. Image restoration step is important in many image processing applications. The Weibull manifold in low-level image processing: an application to automatic image focusing. Content based image retrieval plays a, significant role in the image processing field. One of today's motivating medical image processing problem is registration. Image enhancement is a crucial phase in almost every image processing system

Since the process of summarization highly depends on user's query and feedback, it is not possible to compare the final summary with a reference summary. However, one can observe how much the proposed interactive approach is able to modify the initial summary. In order to show the capabilities of the proposed approach to change the initial summary and its sensitivity to user's feedback, we visualized the rank of sentences at each stage of interaction in Fig. 3. In this figure, given a query, examples of sentences ranking at each stage based on user's facial expression is illustrated. Each block belongs to a different query and each column shows the final ranking of sentences based on 5 consecutive feedbacks (encoded and shown as −1, 0 and 1 in the first row). At each column the top 10 sentences are shown as grey colors, while black/dark grey colors denotes the sentences with the highest score, and white colors are for the sentences that are not included in the summary. Through Fig. 3 we are visualizing how much user feedback is able to affect the initial ranking scheme.

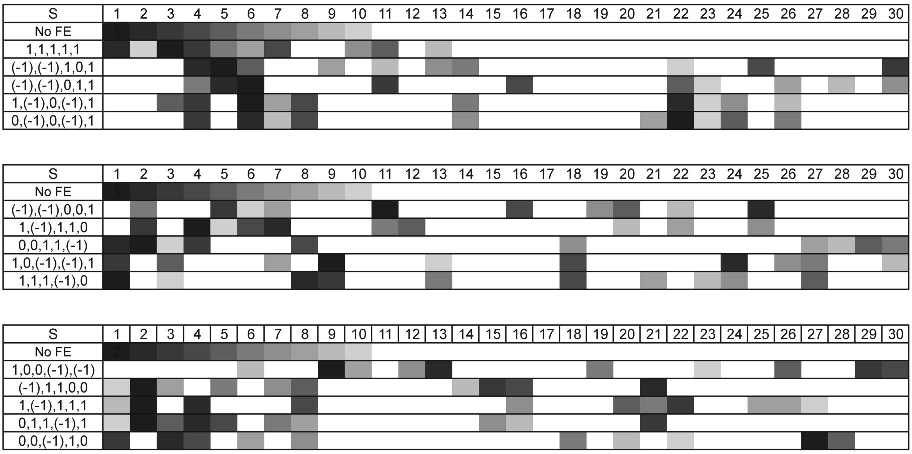


Fig. 3. Effect of user feedback on the initial summary. Each row in the *S* column corresponds to a sentence. At each column the grey cells correspond to the sentences that are chosen to be in summary, black/dark grey sentences are the ones with the highest score. The *NO FE* column corresponds to the case with no facial expression.

Duration of interaction is also a key criterion in social robotics research. The total time of interaction in the proposed approach is mainly affected by how much it takes for the system to retrieve the initial sentences. Thus, in order to study the total time of interaction, we have plotted the average time line of user-robot interaction obtained from multiple runs in Fig. 4. As it is shown, the

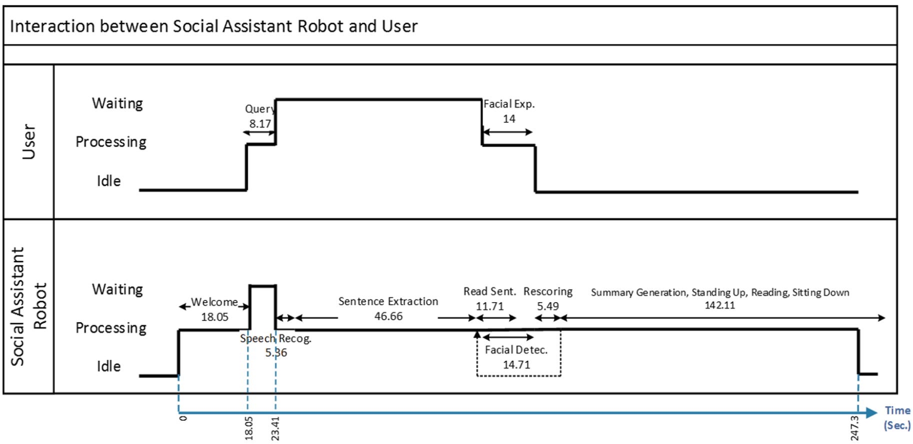


Fig. 4. Timeline of user-robot interaction.

total time of interaction is around 247 s, while 25 % of this time is dedicated to retrieving the initial set of sentences, based on user's query.

5 Conclusions

In this research, a new social robot is introduced for assisting a user in query-based text summarization. Since the query-based text summarization mainly depends on user's query, any other feedback from the user, rather than the initial query, may improve the whole process. Thus, the whole notion behind the proposed framework is to involve the user more efficiently. In this framework, the robot receives feedback from user through its vision in consequent set of interaction and at each interaction, the robot finds the best set of sentences based on user's feedback. Reinforcement learning is used for modelling the problem and reward function. The proposed framework is implemented and evaluated using Nao robot, and some features have been added to this robot in order to enhance its interaction capabilities. The whole approach is then studied in terms of how much it is able to change the initial set of sentences based on user's feedback, and also in terms of interaction time. Future research directions can be focused on improving the initial retrieval process and also evaluating the final summary based user's opinion.

Acknowledgments. This research was supported by Robodoc Human and Machine Interaction Lab at Iranian Research Institute for Information Science and Technology (IranDoc).

References

1. Abdi, A., Idris, N., Alguliyev, R.M., Aliguliyev, R.M.: Query-based multi-documents summarization using linguistic knowledge and content word expansion. Soft Comput. **21**(7), 1785–1801 (2017)
2. Andhale, N., Bewoor, L.: An overview of text summarization techniques. In: 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1–7. IEEE (2016)
3. Bhatia, N., Jaiswal, A.: Automatic text summarization and it's methods-a review. In: 2016 6th International Conference Cloud System and Big Data Engineering (Confluence), pp. 65–72. IEEE (2016)
4. Bosma, W.: Query-based summarization using rhetorical structure theory. LOT Occas. Ser. **4**, 29–44 (2005)
5. Carbonell, J., Goldstein, J.: The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–336. ACM, New York (1998)
6. Damova, M., Koychev, I.: Query-based summarization: A survey (2010)
7. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Rob. Auton. Syst. **42**(3–4), 143–166 (2003)
8. Gupta, V., Lehal, G.S.: A survey of text summarization extractive techniques. J. Emerg. Technol. Web Intell. **2**(3), 258–268 (2010)

9. Hegel, F., Lohse, M., Swadzba, A., Wachsmuth, S., Rohlfing, K., Wrede, B.: Classes of applications for social robots: a user study. In: The 16th IEEE International Symposium on Robot and Human interactive Communication, RO-MAN 2007, pp. 938–943. IEEE (2007)
10. Henß, S., Mieskes, M., Gurevych, I.: A reinforcement learning approach for adaptive single-and multi-document summarization. In: International Conference of the German Society for Computational Linguistics and Language Technology-GSCL, pp. 3–12 (2015)
11. Lee, G.H., Lee, K.J.: Automatic text summarization using reinforcement learning with embedding features. In: Proceedings of the 8 International Joint Conference on NLP (Volume 2: Short Papers), vol. 2, pp. 193–197 (2017)
12. Lopes, A.T., de Aguiar, E., Souza, A.F.D., Oliveira-Santos, T.: Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recogn. **61**, 610–628 (2017)
13. Martins, G.S., Santos, L., Dias, J.: User-adaptive interaction in social robots: a survey focusing on non-physical interaction. Int. J. Soc. Rob. **11**(1), 185–205 (2018)
14. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
15. Mohebi, A., et al.: Conceptual framework for RoboDoc: a new social robot for research assistantship. In: Agah, A., Cabibihan, J.-J., Howard, A.M., Salichs, M.A., He, H. (eds.) ICSR 2016. LNCS (LNAI), vol. 9979, pp. 808–818. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47437-3_79
16. Rahman, N., Borah, B.: A survey on existing extractive techniques for query-based text summarization. In: 2015 International Symposium on Advanced Computing and Communication (ISACC), pp. 98–102. IEEE (2015)
17. Rioux, C., Hasan, S.A., Chali, Y.: Fear the reaper: a system for automatic multi-document summarization with reinforcement learning. In: Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 681–690 (2014)
18. Ryang, S., Abekawa, T.: Framework of automatic text summarization using reinforcement learning. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 256–265. Association for Computational Linguistics (2012)
19. Schütze, H., Manning, C.D., Raghavan, P.: Introduction to Information Retrieval, vol. 39. Cambridge University Press, Cambridge (2008)
20. Yan, R., Nie, J.Y., Li, X.: Summarize what you are interested in: an optimization framework for interactive personalized summarization. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1342–1351. Association for Computational Linguistics (2011)
21. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. **23**(10), 1499–1503 (2016)
22. Zhang, Y., Wang, D., Li, T.: iDVS: an interactive multi-document visual summarization system. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) ECML PKDD 2011. LNCS (LNAI), vol. 6913, pp. 569–584. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23808-6_37