International Journal of Knowledge and Language Processing
Volume 9, Number 3, **2018**

www.ijklp.org
KLP International ©2018 ISSN 2191-2734
pp.30–42

An Overview of Text Generation Technology

Zhen Xu[1] and Xinyu Wu[2]
[1]Institute of Scientific and Technical Information of China
No. 15 Fuxing Road, Haidian District
Beijing, 100038, China
xuzhen@istic.ac.cn
[2]Peking University
No. 5 Yiheyuan Road, Haidian District
Beijing, 100871, China

Received Sep 2018; revised Sep 2018

ABSTRACT. *Text generation technology refers to the generation of natural language text by computer. According to different input, text generation technology can be divided into text-to-text generation, data-to-text generation and image-to-text generation, etc.; according to different tasks, it can be divided into text summarization, image caption, dialogue generation and so on. Nowadays, there are many researches on text generation technology in the field of natural language processing and artificial intelligence, and relevant papers are published at the top conferences every year. This paper summarizes the development process and the latest results of each branch of text generation technology.*
Keywords: Text generation technology, text summarization, image caption, dialogue generation

1. **Introduction.** Automatic text generation is an important research direction in the field of natural language processing. It is a technique of converting different types of input into natural language expression. Text generation can be understood as the reverse of textual understanding: the latter requires the conversion of natural language expressions into machine-readable information, while the former is to convert the former-readable information into natural language expressions. Automatic text generation can be divided into text-to-text generation, data-to-text generation, and image-to-text generation,

depending on the input. Automatic text generation can also be divided into machine translation, text summarization, and picture caption and dialog generation according to different tasks. Since machine translation is a relatively independent research field, this article will not include this technology. It will mainly focus on summary generation, picture caption, and dialog generation techniques.

With the development of deep learning, great progress has been made in the field of automatic text generation. What's more, many scientific and technological achievements have been put into application. For example, Byte Dance's media lab has made a news robot called Xiaomingbot in collaboration with the team led by Wan Xiaojun from Institute of Computer Science and Technology of Peking University. This news robot can learn to generate news through synthesizing and sorting grammatical components. It was initially put into use in 2016. At that time, it could only write sports news, but now news generated by the robot covers dozens of categories such as technology, finance, and real estate. The Associated Press's news robot, WordSmith has been put into use since 2014. The Heliograf from the Washington Post and the WritingMaster from First Financial Corporation also brought about great changes in the journalism industry. As can be seen from these examples, the automatic text generation technology is of strong practicality and an area worthy of attention.

This article consists of five parts. The first one is introduction. The second part introduces the core text summarization technology in text generation, including the extraction method and the generation method. Then it introduces the evaluation method of automatic summarization. The third part is about picture caption technology, the combination of computer vision and natural language processing. It mainly introduces the text-picture alignment method and the model based on vector to generate indefinite length text; the fourth part is the dialogue generation technology, another branch of text generation technology, which mainly transforms abstract language into natural language and expresses it in a smooth way. It also includes a way to evaluate the quality of dialogue generation system. Finally, this paper summarizes all the automatic text generation techniques mentioned and gives a forecast of the future development of text generation technology.

2. Automatic Summarization Technology. Automatic summarization refers to the process of a computer automatically extracting abstracts from the original documents, thereby using a simple and coherent short piece of text to comprehensively and accurately reflect the central content of a certain document. In 1958, H.P. Luhn published an article entitled "The Automatic Creation of Literature Abstracts"[1], marking the prelude to the study of automatic summarization. According to Radev's definition, the abstract is "a piece of text extracted from one or more texts that contains important information in the original text that is no more than half the length of the original text."[2] Automatic text summarization is designed to automatically produce a summary that is simple, fluid, and retains key information using the computer. Automatic text summaries have many application scenarios, such as automatic report generation, news headline generation, and search result preview.

Automatic text summarization can be divided into different types. According to the number of input texts, it can be divided into single document summarization and multiple document summarization. According to the number of language, it can be divided into single language summarization and cross-language summarization. Providing the method of generation, usually it can be divided into two types-extractive summarization and abstractive summarization. Extractive summarization usually uses different methods to evaluate document structural units (sentences, paragraphs, etc.), assigns weights to each structural unit, and then selects the most important structural unit to compose a summary. The abstractive method uses advanced natural language processing algorithms to generate more concise and concise summaries through techniques such as paraphrase, synonymous substitution, and sentence abbreviations. Techniques for sentence compression and sentence fusion are included in the generative method. In short, the former is composed of fragments extracted from the original text, and the latter is the reorganization of the source content. This article focuses on the techniques of how summaries are generated.

2.1. **Text Summarization Generation Method.**

2.1.1. **Extractive Automatic Summarization.** The realization of abstract is seeing sentence as the basic research unit, then evaluating and extracting sentences of the original text by their importance. The advantage of this method is that it is easy to implement and can guarantee the summary has good readability. This method mainly includes two steps: first is to calculate the importance of the sentences in the document and sort them accordingly; second is to select the important sentences to combine into the final summary. Rule-based approach can be used to achieve the first step, for example, the importance of a sentence can be determined by the location of the sentence or the cue words contained in the sentence. Various machine learning methods (including deep learning) can be used to classify, regress or order the importance of the sentence by considering the various features of the sentence as well. For example, the HMM model can be used to judge the possibility of each sentence in the summary[3], the summary task can be treated as a sequential tagging task and the conditional random field can be used to select sentences with the most information[4], and the ranking model based on RNN is developed to rank the sentences in the multi-document summary and learn the ranking features automatically[5]. Based on the sorting results of the first step, the second step is sentence similarity calculation, repeated sentences removal and sentences sorting to obtain the final summary. The most used sentence correlation algorithm is candidate methods (measuring the similarity between candidates and the selected paragraphs and clustering method[6]). There are also many algorithms to sort sentences, such as the method according to the time of publication[7-8], the expansion sort method which puts the topics related to the content together[9] and the bottom-up method of local approximation[10]), so as to obtain the final summary. In addition, there are methods to select sentences and process redundant sentences at the same time. For example, a method based on integer linear programming to select sentences from the shortest path set generated by all clustering[11-13] and to treat the text summary as a problem of budgeted maximization of sub-modular functions[14-15].

Many domestic scientific research institutions have conducted research in the direction

of text summarization, and published their findings on related academic conferences and journals. For sentence selection in automatic summarization, Wu Xiaofeng proposes a supervised extraction method based on sequence segmentation model, which uses semi CRF to label segments. The advantage is that it can use segment as the smallest extraction unit, thereby expanding the range of features[16]. For redundant sentences, he proposes a method to judge sentence similarity by semantic role tagging information, which first judges the semantic roles of all predicates, and then judges the similarity of semantic roles. Wan Xiaojun of the Institute of Computer Science and Technology of Peking University leads his team to achieve a series of results in the field of text summarization. The topic focused multi-document summarization based on manifold sorting[17], conditional Markov walk model based on clustering and HITS model based on clustering[18] are proposed, and the compressed text summarization based on sparse optimization[19] is also proposed by the team. The experimental results of conditional Markov walk model and HITS model on DUC2001 and DUC2002 datasets show the validity of the model, and the conditional Markov random walk model is more robust when the cluster number is different. In addition, for cross-language summarization, he proposed an English-Chinese cross-language summarization method based on translation quality prediction technology[20] and a method to generate summaries using information from both ends of cross-language documents simultaneously[21]. He also proposed a restricted collaborative sorting method according to the different emphasis of news reports in different languages. Li Fang proposed a query-oriented multi-module automatic summary system, which displays four kinds of summary modes: general summary, local summary, global summary and detailed summary to meet user's requirement of summary [22].

2.1.2. **Abstractive Automatic Summarization.** Abstractive automatic text summarization can produce more human-like summaries by semantic parsing of the original document and representing the original document as a deep semantic form (such as a deep semantic map), then obtaining the deep semantic representation of the summary (such as a deep semantic sub-graph), and finally the summary text is generated from the deep semantic representation of the summary. This requires that the generative model has a stronger ability to represent, understand and generate text. At present, the generation of text summary is mainly based on machine learning and neural network, such as Seq2Seq model based on RNN[23] and Deep Reinforced (Seq2Seq+attention model for long text summary[24]. Seq2seq is a network of Encoder-Decoder structure, its input and output are both sequences. Encoder converts a variable length signal sequence into a fixed length vector representation, and Decoder converts this fixed length vector into a variable length target sequence. However, the use of Seq2Seq to deal with long text may lead to repeated phrases and incoherent phrases. Deep Reinforced model is proposed for this purpose. It introduces Attention and reinforcement learning to make the model pay attention to the generated words. It helps to solve the problem that it is easy to repeat the same word and sentence when the Seq2Seq model trying to generate long sentence, and optimizes the result for ROUGE, so that the model can get higher scores in the ROUGE evaluation.

ConvS2 model was put forward by AI Laboratory of Facebook[25]. Unlike Seq2Seq and

Deep Reinforced, it is based entirely on CNN and has the advantage of parallel computing while training. When it is implemented, word order is expressed as a distributed vector, which makes the model obtain word order and position information, simulates the perception of word order in RNN, and adds nonlinear transformation on the basis of traditional CNN, which makes the output independent of the length of input and easier to optimize.

Historically, the effect of extractive summarization is usually better than abstractive summarization. Earlier studies were mainly focused on abstractive summarization as well. With the rise of deep neural network, the abstractive text summarization based on neural network has been developed rapidly, and has achieved good results. There are several abstractive neural network models that have surpassed the best extractive models on the DUC-2004 test set.

2.2. **The Evaluation of Automatic Text Summarization.** Jones and Galliers divides the evaluation of automatic summarization into two types: one is internal evaluation, which evaluates the quality of summaries by analyzing the quality of them directly; the other is external evaluation, which considers specific purposes of summaries and evaluates the quality of summaries according to the results completed. For example, it can be evaluated according to the correct rate of retrieval when generating summaries for information retrieval[26].

Internal evaluation is the commonly used method in academia. It needs to compare the automatically generated summaries with the summary provided by experts. In other words, it needs a manually created reference summary. ROUGE is an internal evaluation method, which is based on the BLEU score of machine translation. It is a standard using the longest common substring and co-occurrence statistics of word pairs in sentences.[27] The main implementation is to use the number of n-gram matches between the reference summary and the generated summary to divide the number of n-gram phrases in the whole generated summary[28]. The result of this method has been proved to be consistent with the result of manual evaluation.

3. **Image Caption Technology.** The technology of image caption is to generate the corresponding description of a given image. It involves the knowledge of computer vision and natural language processing at the same time. These two areas, which had previously been developed separately, have more and more interactions in recent years due to the growing concerns in combining language and visual information. Image caption is also called image title generation. Its requirement is to describe a natural image in the form of natural language, that is, to translate the information in the image. It requires the system to understand the content of the image first, such as recognizing the scene in the image, a variety of objects, the object properties, the action that is taking place and the relationship between the objects; then according to the grammar rules and language structure, generates human-understandable sentences. Many methods have been proposed to solve this problem, including template-based approach, semantic migration approach, neural machine translation approach, and hybrid approach. With the continuous breakthrough of deep

learning technology, especially CNN technology in the field of language recognition and vision, the method based on neural machine translation and its mixing with other visual technologies has become the mainstream to solve this problem. This kind of methods considers that the CNN model can extract the image features which are more abstract and more expressive, and can provide reliable visual information for the subsequent language generation model.

3.1. **Image Caption Generation Method.** The key of image caption technology is image description model, that is, how to describe relevant relationship between entities and attributes and their behaviors. There are two main description models, one is to describe the retrieval in visual space, and the other is to describe the retrieval in multi-modal space.

3.1.1. **Retrieval in Visual Space.** Retrieval in descriptive visual space is to automatically generate an image description by retrieving an image similar to the query image. The general step of visual retrieval method is: 1) A given query image is represented by a specific visual feature. 2) The candidate image set is retrieved from the training set based on the similarity of features in the visual space. 3) A description of the candidate image is rearranged by further utilizing the visual and/or textual information contained in the retrieval set, or fragments of the candidate description are combined according to a particular rule or scheme. The Im2Text model is based on the visual similarity between the object regions detected in the training image and the query image, and extracts nouns and verb phrases from the descriptions in the training set. The visual similarity between the detection of the query image and training images is measured based on the appearance and geometric layout of the query to detect and collect prepositional phrases for each feature in the query image. By measuring the calculated global scene similarity between the query and training images, additional prepositional phrases are collected for each scene context detection. Finally, a description of each detected object is generated from these collected phrases by integer linear programming, in which factors such as word order and redundancy are taken into account [29]. There is also a method of extracting the visual content of the input image based solely on the text information in the description of the visually similar image, first dividing the candidate description into phrases like (subject, verb) (subject, verb, object) (verb, preposition, object), and then using the fixed template to generate the description[30]. With the introduction of deep learning, the method of description generation has changed. The generation of description becomes a problem of summary extraction. In the last step of reordering, the text information is considered to select the output description. The concrete methods include the distributed query expansion method based on the combinatorial distributed semantics of CNN[31] and the method of taking the output of CNN as the input of RNN to generate description[32].

3.1.2. **Retrieval in Muti-modal Space.** Multi-modal retrieval also regards image description as a retrieval problem. The concrete steps are as follows: 1. A training set of image description pairs is used to learn a common multimodal space for both visual and textual data. 2. Given a query, cross-modal (image-sentence) retrieval is performed using a joint representation space. In this model, images and sentences are mapped to a common space, resulting in a joint space that can be used for image search (to find the most

reasonable image for a given sentence) and image annotation (to find sentences that describe an image)[33]. Initially, the linear model was used to construct the joint space, now the neural network is used to construct vector representations of sentence and image, and then these representations are mapped to a common embedded space[34]. When generating the description, bilinear model is used to learn the common space of image features and syntactic phrases (noun phrases, verb phrases and prepositional phrases). Then, Markov model is used to generate sentences from these embedded phrases[35]. Specifically, the first step is to generate a visual semantic alignment model; the second step is to generate a multimodal RNN model for the text description of the new image. The visual semantic alignment model proposed by Andrej Karpathy et al. uses RCNN to extract image features and then uses bidirectional cyclic neural network (BRNN) to calculate the representation of words, which can avoid the loss of context meaning caused by the direct mapping of words[36]. After obtaining the image features and the text, the score function is used to calculate the text matching of each feature point in the pattern, so as to obtain the aligned image and text. The multi-modal RNN model predicts a variable-length text sequences based on the given images. During training, the model needs to predict the next word based on the given word and its context, and repeat until reaching the end symbol. When given a new image, the model will find the starting vector of the corresponding text according to the feature vector of the image, then repeat the previous process, output the text until the end symbol is met and the output is completed.

3.2. **The Evaluation of Automatic Image Caption.** The evaluation of image caption has always been a controversial topic, because it is a relatively subjective task. It is difficult to generate a set of objective and unified evaluation criteria. Whether the evaluation standards can achieve a result consistent with human evaluation is the focus of current studies. SPICE is a widely accepted evaluation method[37]. It will evaluate the generated text and reference text according to the generative semantic scene graph, with the contact ratio of tuples in two kinds of text to represent the quality of the generated text. The advantage of this method is that it can be consistent with the human judgment, but the disadvantage is that only tuples in the sentence are taken into account, and the overall structure of the sentence is not considered. On the CVPR 2018, a method of using machine learning to score the quality of the generated text is put forward. It uses a training network to evaluate the quality of the sentence, and a classifier to classify the reference text and the generated text according to the quality of the sentence. Then it uses the regression method to mark the sentences accurately[38].

4. **Dialogue Generation Technology.** The technology of dialogue generation refers to the multiple rounds of questions and answers between the computer and human beings. There is no need for the conversational system to provide answers in the first round of interaction since there can be a rhetorical mechanism to guide users to reformulate their requirements in a more acceptable way to the system. After that, users are often able to change expressions that are not standardized for the system. Question and answer generation technology has a very wide range of applications, such as online shopping automatic

customer service[39] and chat robot[40],. The development of this technology enjoys broad prospect.

Dialogue generation system mainly includes four modules: construction of knowledge base, understanding of natural language, tacking of state and generation of answer. The technology of text generation involved is to convert the abstract dialogue into fluent natural language. The traditional method is to map sentences. The input semantic symbols are mapped to the intermediate form of utterance, such as tree or template structure, and then the intermediate structure is transformed into the final answer through the surface implementation[41-42]. After the introduction of neural network, the dialogue behavior types and constraints are transformed into one-hot control vectors as input information to ensure that the generated dialogue is in accordance with expectations[43]. To avoid duplication when generating text, it is also necessary to set up a control unit to control the conversation behavior.

4.1. **Dialogue Generation Method.** Dialogue generation is to use natural language to answer users' questions. There are usually three ways to achieve the target: 1) Based on manual templates; 2) Based on the knowledge base retrieval; 3) Based on Seq2Seq model.

4.1.1. **Template-based Dialogue Generation.** The technology based on manual template sets the dialogue scene manually, and writes corresponding dialogue template for each scene. The final form of reply is to fill a template that most content has already been given, and only some concrete parameters need to be filled in.

Manual template-based dialog generation can be divided into five modules: input system, active ontology, implementation system, service system and output system. The data and models stored in active entities include: domain model, user customized information, language schema, vocabulary and domain entity database. The domain model includes the concept, entity, relation, attribute and the internal representation of instances in a vertical domain, namely the ontology. Vocabularies are used to maintain domain-specific conversation templates that are written manually. The implementation system parses the user's input into user's intention for internal use and then calls data in the active otology to assemble the answer, guide the user to input and generate the result.

Template-based dialogue generation is more suitable for specific domains and can provide precise answers within a limited range, but the disadvantage lies in its poor scalability and portability.

4.1.2. **Knowledge Base Retrieval-based Dialogue Generation.** The technical route of knowledge base retrieval-based dialogue generation is similar to that of search engine. A database called knowledge base is prepared in advance, which contains abundant dialogue materials and the problems are indexes therein. Then the NLP technology is used to analyze the problem raised by users. The most appropriate response content is found by fuzzy matching the defined knowledge base with keyword extraction, inverted index, document sorting and other methods. The core technology of such solutions is to find more data to enrich and clean the knowledge base, but it is difficult to monitor when the amount of data is too large. As a result, disjointed date may lead to poor continuity in the conversation. DeepQA is a technology based on knowledge base retrieval[44-45], which gives answers to

a question by searching and quantifying evaluation of prepared materials. In particular, the system retrieves many different resources based on different understanding of questions and types, and returns a variety of candidate answers[46]. Any answer is not immediately determined because over time the system gathers more and more evidence to analyze each answer. Then the system uses hundreds of different algorithms to analyze the evidence from different angles to get hundreds of eigenvalues or scores, which represent the degree to which some evidence supports an answer on a particular dimension. All eigenvalues or scores for each answer are combined into a single score, indicating the probability of the answer being correct. The system uses statistical machine learning method to learn a large number of data sets to determine the weight of the various eigenvalues, and ultimately will output the answer which has the highest score[47].

Knowledge base-based conversation generation is very extensible, but there may be situations where answers are not coherent and contextual.

4.1.3. **Seq2Seq-based Dialogue Generation.** Dialogue generation based on deep learning usually does not rely on a specific answer library or template, but based on the language ability learned from a large number of corpus to carry out the dialogue. A method of generating answers directly from the content of a question is defined as a generation model based on certain conditions. The Seq2Seq model described in 2.1.2 can also be used to perform this task. Google uses the Seq2Seq model to generate a question and answer set in the IT domain, and generates a dialog system upon it, which outperforms the template-based dialogue system [48]. In order to take the contextual connection into consideration, Context is introduced to form the Context+seq2seq model. There are two approaches to deal with Context, one is to replace RNN model with multi-layer feed forward neural network in Encoder part, which can encode Context and information into the middle semantic expression of Encoder-Decoder model through multi-layer feed forward neural network[49], and can avoid the excessive text length caused by the direct connection of Context and the information at the same time. Another approach is to use a hierarchical neural network[50] to encode every sentence in the Context into an intermediate semantic representation with RNN. Then the results of the first stage are encoded in order by the second level of RNN, which is called ContextRNN. In this way, the state information of hidden layer node at the tail node is the semantic encoding of all Context and current information, which is used as the output of the Decoder. This allows contextual information to be taken into account when answers are generated.

4.2. **The Evaluation of Generative Dialogue.** In view of the natural language understanding stage of dialogue generation, researchers use support vector machine (SVM)[51] and deep syntax classification[52] to classify the problem, and then use the results returned by search engines to calculate the mutual information between the each word in the problem in order to determine the keywords of the problem[53]. Good results have been achieved. There are also many studies on the evaluation of dialogue quality, such as the Ruber index based on previous answers[54], the classification of questions and answers to improve the quality of evaluation[55], the conversion of answering questions into text summary questions.

5. **Conclusions.** This paper summarizes the technology of automatic text generation and combs the cutting-edge technology in this field by selects articles of top journals and conferences. Automatic summarization is the core of automatic text generation technology, because other text generation technology can be more or less converted into the problem of summary generation. For example, Seq2Seq model is used both in text summarization and dialogue generation. Automatic text generation technology is widely used in the business field. Many automatic text generation tools have brought great changes to the industry. However, because of the complexity of human language, the current automatic text generation technology cannot replace human beings. But one can tell from the above technology that the plasticity of deep learning is very strong, therefore it can realize more potential in the text generation field. For example, the computer may learn to generate text through antagonistic neural network. Moreover, how to synthesize all kinds of text generation functions to form a reusable text generation framework is also one of future development direction of text generation technology.

REFERENCES

[1] Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development, 2(2), 159–165.

[2] Radev, D. R. et al. (2002). Introduction to the Special Issue on Summarization. Computational Linguistics, 28(4), 399–408. http://doi.org/10.1162/089120102762671927

[3] Conroy, J. M., & O'Leary, D. P. (2001). Text Summarization via Hidden Markov Models. Sigir, 406–407. http://doi.org/10.1145/383952.384042

[4] Shen D, et al. (2007). Document Summarization Using Conditional Random Fields. Presented at the IJCAI.

[5] Cao Z, et al. (2015). Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization. AAAI, 2153–2159.

[6] Carbonell, J. G., & Goldstein, J. (1998). The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. Sigir, 335–336. http://doi.org/10.1145/290941.291025

[7] McKeown, K. R, et al. (1999). Towards Multidocument Summarization by Reformulation - Progress and Prospects. AAAI.

[8] Lin, C. Y., & Hovy, E. (2001). From single to multi-document summarization (pp. 457–464). Presented at the the 40th Annual Meeting, Morristown, NJ, USA: Association for Computational Linguistics. http://doi.org/10.3115/1073083.1073160

[9] Barzilay, R., Elhadad, N., & McKeown, K. R. (2001). Sentence ordering in multidocument summarization (pp. 1–7). Presented at the first international conference, Morristown, NJ, USA: Association for Computational Linguistics. http://doi.org/10.3115/1072133.1072217

[10] Bollegala, D., Okazaki, N., & Ishizuka, M. (2006). A Bottom-Up Approach to Sentence Ordering for Multi-Document Summarization. 385–392.

[11] McDonald, R. T. (2007). A Study of Global Inference Algorithms in Multi-document Summarization. (pp. 557–564). Presented at the ECIR, Roman, Italy: Springer Berlin Heidelberg.

http://doi.org/10.1007/978-3-540-71496-5_51

[12] Gillick, D., & Favre, B. (2009). A scalable global model for summarization (pp. 10–18). Presented at the HLT-NAACL, Morristown, NJ, USA: Association for Computational Linguistics. http://doi.org/10.3115/1611638.1611640

[13] Oliveira H, et al. (2017). A Regression-Based Approach Using Integer Linear Programming for Single-Document Summarization. ICTAI, 270–277. http://doi.org/10.1109/ICTAI.2017.00051

[14] Lin, H., & Bilmes, J. A. (2011). A Class of Submodular Functions for Document Summarization. (pp. 510–520). Presented at the ACL.

[15] Lin, H., & Bilmes, J. A. (2010). Multi-document Summarization via Budgeted Maximization of Submodular Functions. (pp. 912–920). Presented at the HLT-NAACL.

[16] Wu Xiaofeng. Research on automatic text summarization method [D]. Graduate University of Chinese Academy of Sciences, 2010.

[17] Wan, X., Yang, J., & Xiao, J. (2007). Manifold-Ranking Based Topic-Focused Multi-Document Summarization. (pp. 2903–2908). Presented at the IJCAI.

[18] Wan, X, & Yang, J. (2008). Multi-document summarization using cluster-based link analysis. (pp. 299–306). Presented at the SIGIR, New York, New York, USA: ACM Press. http://doi.org/10.1145/1390334.1390386

[19] Yao, J. G., Wan, X., & Xiao, J. (2015). Compressive Document Summarization via Sparse Optimization. (pp. 1376–1382). Presented at the IJCAI.

[20] Wan, X., Li, H., & Xiao, J. (2010). Cross-Language Document Summarization Based on Machine Translation Quality Prediction. (pp. 917–926). Presented at the ACL.

[21] Wan, X. (2011). Using Bilingual Information for Cross-Language Document Summarization. (pp. 1546–1555). Presented at the ACL.

[22] Li Fang, He Tingting. Multi-mode automatic summary research for query [C]// National Youth Computing Linguistics Symposium. 2010.

[23] Nallapati R,.et al. (2016). Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. Presented at the CoNLL.

[24] Paulus, R., Xiong, C., & Socher, R. (2017). A Deep Reinforced Model for Abstractive Summarization., cs.CL.

[25] Gehring, J, et al. (2017). Convolutional Sequence to Sequence Learning.

[26] Gaizauskas, R. J. (1998). Karen Sparck Jones and Julia Galliers, Evaluating Natural Language Processing Systems - An Analysis and Review. Berlin - Springer-Verlag, 1996. ISBN 3 540 61309 9, Price DM54.00 (paperback), 228 pages. Natural Language Engineering.

[27] Lin, C. Y., & Och, F. J. (2004). Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics.

[28] Lin, C. Y. (2004). ROUGE: A Package for Automatic Evaluation of summaries (p. 10). Presented at the Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL, Barcelona, Spain.

[29] Ordonez, V., Kulkarni, G., & Berg, T. L. (2011). Im2Text - Describing Images Using 1 Million Captioned Photographs. Presented at the NIPS.

[30] Singh, N., et al. (2012). Semantic Image Retrieval by Combining Color, Texture and Shape Features (pp. 116–120). Presented at the International Conference on Computing Sciences.

[31] Yagcioglu, S, et al. (2015). A Distributed Representation Based Query Expansion Approach for Image

Captioning.
[32] Devlin, J., et al. (2015, May 8). Language Models for Image Captioning: The Quirks and What Works. arXiv.org.
[33] Hodosh, M., Young, P., & Hockenmaier, J. (2015). Framing Image Description as a Ranking Task - Data, Models and Evaluation Metrics (Extended Abstract). Presented at the IJCAI.
[34] Tai, K. S., Socher, R., & Manning, C. D. (2015, February 28). Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. arXiv.org.
[35] Lebret, R., & Collobert, R. (2015). Rehabilitation of Count-Based Models for Word Vector Representations. (pp. 417–429). Presented at the CICLing, Springer International Publishing. http://doi.org/10.1007/978-3-319-18111-0_31
[36] Karpathy, A., & Li, F. F. (2015). Deep visual-semantic alignments for generating image descriptions. Cvpr, 3128–3137. http://doi.org/10.1109/CVPR.2015.7298932
[37] Anderson, P., Fernando, B., Johnson, M., & Gould, S. (2016). SPICE: Semantic Propositional Image Caption Evaluation (pp. 382–398). Presented at the ECCV, Amsterdam, The Netherlands: Springer International Publishing. http://doi.org/10.1007/978-3-319-46454-1_24
[38] Cui, Y., Yang, G., Veit, A., Huang, X., & Belongie, S. J. (2018). Learning to Evaluate Image Captioning. Presented at the CoRR.
[39] Yan, Z., Duan, N., Chen, P., Zhou, M., Zhou, J., & Li, Z. (2017). Building Task-Oriented Dialogue Systems for Online Shopping. Presented at the AAAI.
[40] Wen, T.H, et al. (2015). Stochastic Language Generation in Dialogue using Recurrent Neural Networks with Convolutional Sentence Reranking. (pp. 275–284). Presented at the SIGDIAL Conference.
[41] Walker, M. A., Rambow, O., & Rogati, M. (2002). Training a sentence planner for spoken dialogue using boosting. Computer Speech & Language, 16(3-4), 409–433. http://doi.org/10.1016/S0885-2308(02)00027-X
[42] Stent, A., Prasad, R., & Walker, M. (2004). Trainable sentence planning for complex information presentation in spoken dialog systems (pp. 79). Presented at the the 42nd Annual Meeting, Morristown, NJ, USA: Association for Computational Linguistics. http://doi.org/10.3115/1218955.1218966
[43] Wen, T.-H., et al (2015). Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems. (pp. 1711–1721). Presented at the EMNLP.
[44] Gondek, D., et al. (2012). A framework for merging and ranking of answers in DeepQA., 56(3.4), 14:1–14:12. http://doi.org/10.1147/JRD.2012.2188760
[45] Ferrucci, D. A, et al. (2013). Watson - Beyond Jeopardy!, 199-200, 93–105.
[46] Kalyanpur, A, et al. (2012). Fact-based question decomposition in DeepQA., 56(3.4), 13:1–13:11. http://doi.org/10.1147/JRD.2012.2188934
[47] Gondek, D., et al. (2012). A framework for merging and ranking of answers in DeepQA., 56(3.4), 14:1–14:12. http://doi.org/10.1147/JRD.2012.2188760
[48] Vinyals, O., & Le, Q. V. (2015). A Neural Conversational Model. CoRR.
[49] Sordoni, A., et al. (2015). A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. Presented at the HLT-NAACL.
[50] Serban, I. V., et al. (2016). Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models. Presented at the AAAI.
[51] Zhang, D., & Lee, W. S. (2003). Question classification using support vector machines. 26.

http://doi.org/10.1145/860435.860443
[52] Li, X., & Roth, D.(2002). Learning Question Classifiers. COLING.
[53] Cui, H., Kan, M.-Y., & Chua, T.-S. (2004). Unsupervised learning of soft patterns for generating definitions from online news. Www, 90. http://doi.org/10.1145/988672.988686
[54] Tao C, et al. (2018). RUBER - An Unsupervised Method for Automatic Evaluation of Open-Domain Dialog Systems.
[55] Liu, Y., et al. (2008). Understanding and Summarizing Answers in Community-Based Question Answering Services. Presented at the COLING.