

Conditional Text Paraphrasing: A Survey and Taxonomy

Ahmed H. Al-Ghidani, Aly A. Fahmy
Computer Science Department
Faculty of Computers and Information,
Cairo University
Cairo, Egypt

Abstract—This work introduces a survey for the *Text Paraphrasing* task. The survey covers the different types of tasks around text paraphrasing and mentions the techniques and models that are regularly used when approaching towards it, alongside the datasets that are used while training and evaluating the models. Text paraphrasing has an effective impact when it is used in other applications, so, the paper mentions some text paraphrasing applications. Also, this work proposes a new taxonomy that it is called *Conditional Text Paraphrasing*. To the best of our knowledge, this is the first work that shows varieties and sub-problems of the original text paraphrasing task. The target of this taxonomy is to expand the definition of the text paraphrasing by adding some conditional constraints as features that either control the paraphrase generation or discrimination. This expanded definition opens in mind a new domain for research in Natural Language Processing (NLP) and Machine Learning. Finally, some useful applications for the conditional text paraphrasing are represented.

Keywords—*Natural Language Processing; Text Paraphrasing; Conditional Text Paraphrasing*

I. INTRODUCTION

A. Problem Definition

Text Paraphrasing is a core and challenging problem in Natural Language Processing. The problem refers to texts that convey the same meaning but with different expressions. It can be considered as a transformation for a given text while keeping its semantic meaning. These transformations may be at the level of the texts linguistics and structure. Paraphrasing differs from Entailment in the type of relationship between instances. *Entailment* occurs when one may draw necessary conclusions from an input instance. For example, *Rocky is a Dog. Rocky is an animal.* Entailment has a form of "*If A then B*", while Paraphrasing has a form of "*A is B*". Entailment is a one-way relationship. "*If A then B*" is not "*If B then A*". For instance: *Rocky is a Dog* means *Rocky is an animal*, while *Rocky is an animal* does not necessarily mean that *Rocky is a Dog* as it is maybe another type of animals. Unlike Entailment, Paraphrasing is a two-way relationship. For instance, *What is the distance between Earth and Sun?* is the same of *How many miles between Earth and Sun?*

This work demonstrates the defined tasks around the text paraphrasing problem (section 1.2) and formulates discrimination and generation tasks, alongside mentioning some existed researches and efforts that are done on both directions (sections 2.1 and 2.2). After that, some evaluation metrics are represented that are regularly used when evaluating the

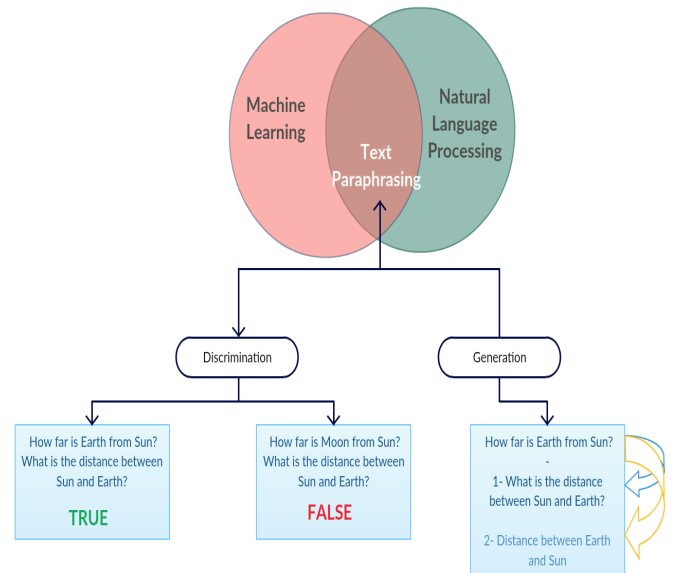


Fig. 1. Text Paraphrasing Tasks

model (section 3) and the datasets used (section 4). We talk about several applications on which text paraphrasing is used, either for data augmentation or as a module in large systems (section 5). Finally, a proposed definition for conditional text paraphrasing is introduced and its taxonomy (section 6), alongside showing some important applications on it (section 7).

B. Defined Tasks

As shown in Figure 1, text paraphrasing is a type of problems at which natural language processing and machine learning could co-operate to solve it. Text paraphrasing involves two different tasks, **Discrimination** and **Generation**. The target of the discrimination is to check if the two given texts are paraphrased texts or not. In that case, the task is considered to be a discriminative problem. In the generation, the target is to generate text(s) given a reference text, in that case, the task is a generative problem.

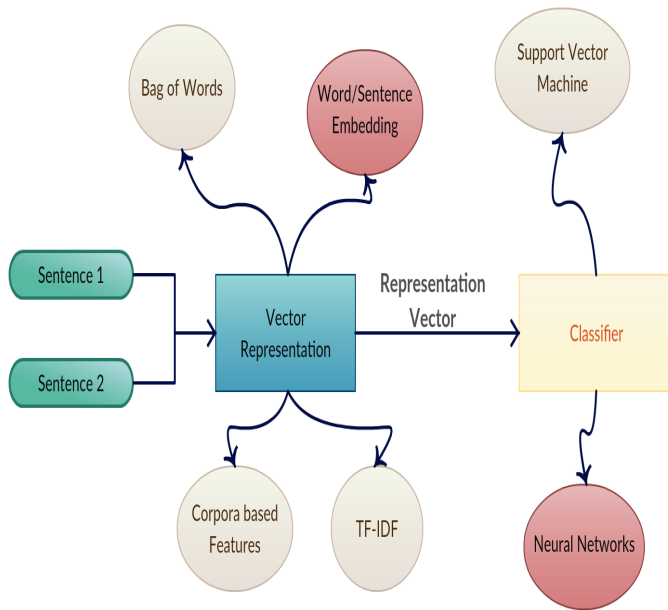


Fig. 2. Discrimination Model Pipeline

II. PROBLEM FORMULATION

A. Discrimination Task

Given two sentences (S_1, S_2), where $S_1 = \{w_1, w_2, \dots, w_n\}$ and $S_2 = \{w_1, w_2, \dots, w_m\}$, the target is to check if they are paraphrases or not.

Some researches look to this problem as a Semantic Text Similarity problem on which some distances metrics, such as Euclidean and Cosine distances are used [1], [2], alongside either using binary vectors as feature representation for sentences extracted from lexical-based features or TF-IDF representation [3]. However, the great successes of distributed words and sentences representation [4], [5], [6] altered the basic representations for texts to be used in distances measure [7].

Other researches look to the problem as a supervised learning problem. The problem is often formalized as a binary classification problem $y = \{0, 1\}$. Like in [8] on which Support Vector Machine (SVM) is used with basic features representations for sentences. The great successes of deep neural networks in fields like natural language processing, computer vision and speech recognition in both supervised and unsupervised learning problems was a motivation to build a neural-based model for such classification problem [9]. Recently, Convolutional Neural Networks (CNNs) showed remarkable results in text modeling for classification and features extraction [10], [11], so that it could be used for this task like in [12]. Other works focus on recurrent based models like in [13] on which Long Short-Term Memory (LSTM) is used to encode the sentences embedded into a Siamese network structure [14]. Shown in Figure 2, this is the general pipeline for the discrimination task.

In conclusion, currently, deep neural networks are heavily used for the identification task, alongside the currently advanced words and sentences representations.

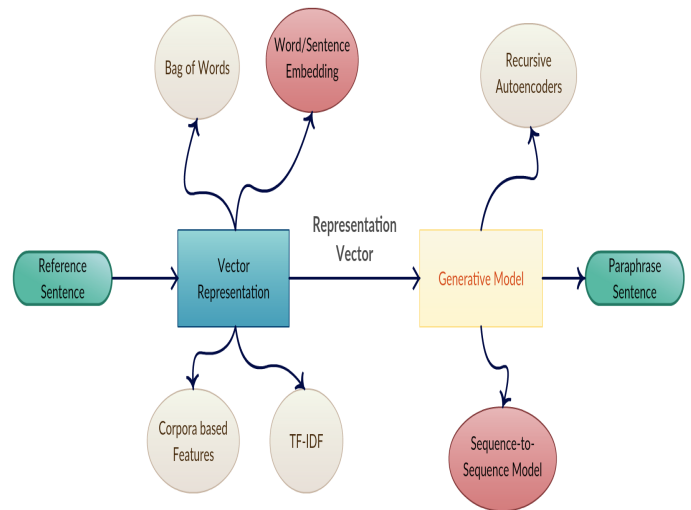


Fig. 3. Generation Model Pipeline

B. Generation Task

In the generation task, given a reference sentence S_1 , where $S_1 = \{w_1, w_2, \dots, w_n\}$, the target is to generate candidate(s) sentence(s) that are semantically equivalent to the reference sentence. It is considered to a text generation problem.

Classically, some lexical-based features and wording replacement are used to generate alternatives to the reference sentence. For instances, paraphrases are generated using templates extracted from WikiAnswers repositories like in [15] and lexical-based rules like in [16]. They make use of WordNet to get words hypernyms and synonyms for replacements, however, these techniques suffer from the generation of poor candidate paraphrases.

Recently, the great successes of the deep generative models [17] such as Variational Autoencoders (VAEs) [18] and Generative Adversarial Networks (GANs) [19] had a great impact in unsupervised learning problems. Several research worked on generation realistic texts either for task specific problems, such as machine translation [20], [21], and question generation [22]. Generic text generation has been investigated using VAEs [23] and GANs [24]. As these models depend on the hidden representation of sentences during the training, the produced texts are randomized and uncontrollable. Serious attempts were made recently to control the generated sentences [25], [26], using some conditional features such as the sentence's polarity and syntax-tree [25], [26]. Typically, this problem is considered to be sequence-to-sequence problem [27] on which the target is to generate sequence(s) of words given other sequence of words [28], [29], [30]. Shown in Figure 3, this is the general pipeline for the generation task with highlights of the most used techniques nowadays.

III. EVALUATION MEASURES

Evaluation metrics are performed on the discrimination and generation tasks. As the discrimination task is a supervised learning problem, metrics such as accuracy, precision, recall and f1-score are used to evaluate the trained models. This is

very different from the generation task on which BLEU (Bilingual Evaluation Understudy) [31], ROUGE (Recall Oriented Understudy for Gisting Evaluation) [32], METEOR (Metric for Evaluation for Translation with Explicit Ordering) [33] and Translation Error Rate (TER) [34] are used for approximate all natural language generation tasks.

IV. DATASETS

Compared to other tasks, the datasets for the text paraphrasing task aren't large. This supports the importance of having a robust and generalized text paraphrasing models to help on creating datasets with large diversity for the problem itself and other problems in general such as Sentiment Analysis and Named Entity Recognition.

A. MSCOCO

Microsoft has recently released a dataset for images captions [35]. The dataset comes with 120K images that are captioned with short and medium size texts. For each image, five captions are provided that describe the images, and these five captions are written by different five annotators. As the annotators are describing the same things, the captions could be represented as paraphrases to each other.

B. PPDB

PPDB [36], [37] is widely known dataset for paraphrase generation. It comes with wide sizes, however the most used size is PPDB 2.0 Large dataset. For some phrases, PPDB has one-to-many paraphrases.

C. Quora Questions Pairs

Quora questions pairs [38] is a dataset produced by Quora. The dataset contains approximately 400K pairs of sentences. Sentences are questions that are labeled by whether the pair is duplicate, has the same semantic meaning, or not. The duplicate questions are considered to be paraphrases if they are duplicates.

D. SNLI

The SNLI dataset [39] consists of approximately 570k sentences. These sentences are generated by human annotators and manually labeled with either they are entailment, contradiction and neutral. For the paraphrasing task, the focus is on the neutral sentences as they describe the paraphrases sentences to each other. SNLI is heavily used in natural language generation tasks and to capture the semantics of languages [40].

E. WikiAnswers

WikiAnswers [41] is a large question paraphrase corpus created by crawling the WikiAnswers website. The paraphrases are different questions, which were tagged by the users as similar questions. The dataset contains approximately 18M question pairs aligned by word.

V. APPLICATIONS

Paraphrasing has numerous applications. It is either used as a preprocessing module to increase the datasets as a *Data Augmentation* technique or embedded in an end-to-end model. For instances, paraphrasing is used in Question Answering (QA) systems [42], where several paraphrases are generated in an end-to-end neural-based model for the given question to increase the diversity and the coverage of the input.

Paraphrasing is also used to accurately evaluate the Machine Translation models [43]. This is done by generating paraphrases that are closer in wording to the translation output based on some lexico-semantic resources such as WordNet. Some researches use paraphrasing to improve the the generated logical form of sentences [44] that is known as Semantic Parsing problem and as a representation for the input queries [45] in Semantic Search.

Paraphrasing has also an important application which is plagiarism detection. In this task the target is check whether two texts are copied or altered to another or not. It is a typical paraphrasing application. It is mainly could be used for author copyrights ownership. Natural language processing suffers from a lack of resources and datasets that could be used to train the models [26], such problem decreases the model's generalization. Several researches use paraphrasing as a data augmentation technique to enrich the datasets [46], [47], [48]. Recently, the concept of paraphrasing is used to reformulate the questions to lead for better questions generation and to increase the diversity of questions intents [49].

VI. PROPOSAL: CONDITIONAL TEXT PARAPHRASING TASK

This work proposes a *Conditional Text Paraphrasing* task as an addition to the original task. As paraphrasing only focuses on the semantics of the sentences, there is a need for more paraphrasing specifications to control the recognition and generation processes. On conditional paraphrasing, the target is to either detect or generate according to specific condition or constraint. The conditions are divided to Morphology-based, Syntax-based and Readability-based categories. This work is driven by several attempts to control the text generation [25], [26]. The target of this work is to set and organize the problem as a task that is closely related to the original task. The research is represented on some ways to control the paraphrase generation by creating a taxonomy that defines the task. This may help other researches when they work on it, also, this taxonomy could be expanded for more concrete tasks if needed.

Feasible sentences generation is considered to be a challenging problem in natural language processing because of several obstacles that occur while modeling text. For instances, representing the text in a way of hidden representation that captures the semantics of text and its structure is hard because of the complexity of text and its language. Controlling the text generation means that we need to make the latent representation of sentence capture the semantics, structure and the other disentangled representation of the embedded attribute. For images, some attempts were done to control the generated image with disentangled attribute [50], [51]. In general, as the nature of text is a discrete data; training the models that depend

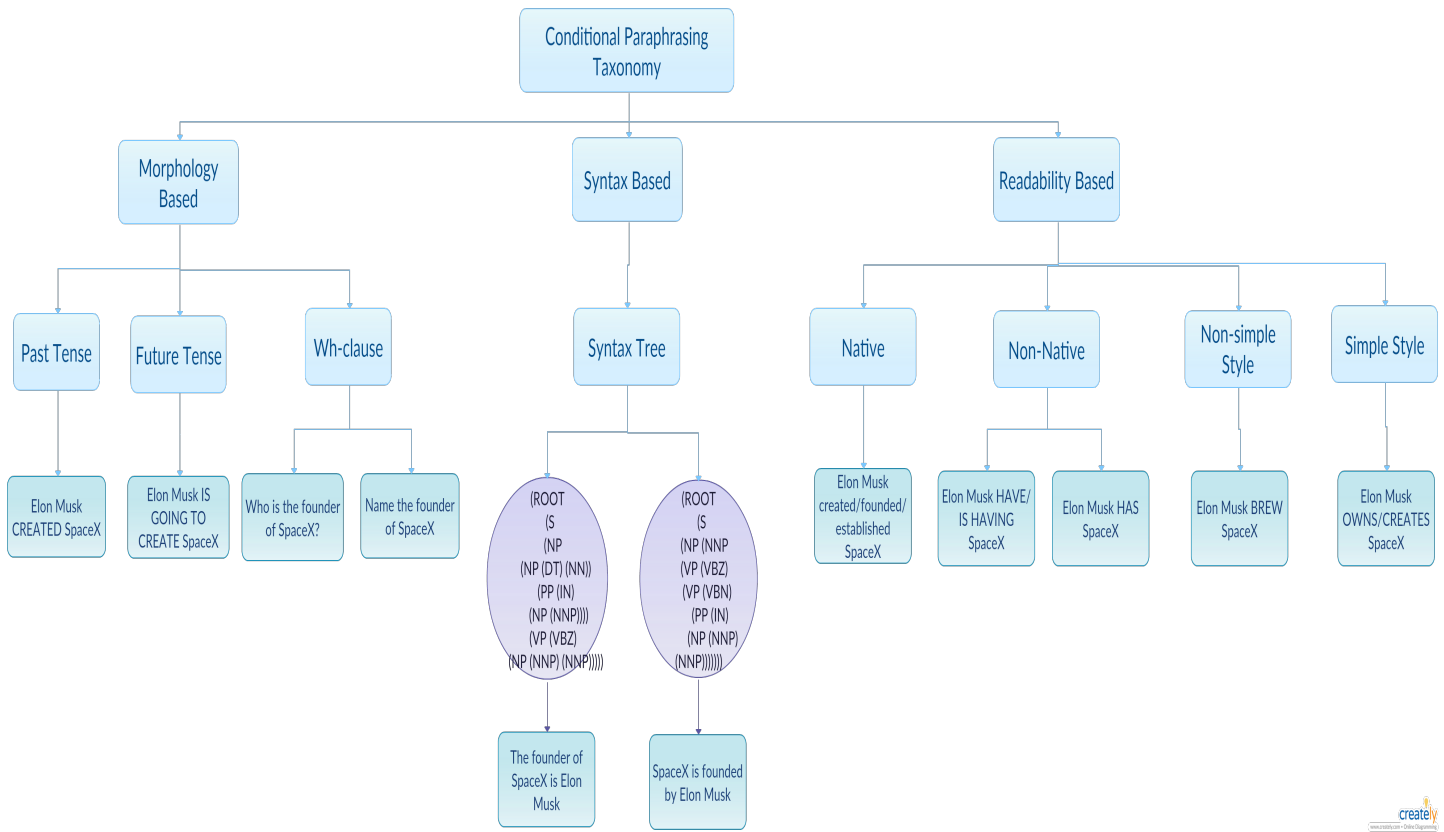


Fig. 4. Conditional Text Paraphrasing Taxonomy. Reference Text: *Elon Musk is the founder of SpaceX*

on optimization and back-propagation becomes hard to reach the global minima. The GAN-based model created by [25] tackled this problem by allocating one dimension of the latent representation to encode the disentangled attributes, such as the polarity in sentiment analysis task, and generates samples with desired sentence semantics. The research focused on disentangled representations of polarity (positive, negative) and Tense (past, present, future) attributes, however, our research believes that this could be generalized further more by expanding the additive attributes that could be conditioned to the model to force it to generate more intensive and concrete samples. These conditions are divided to Morphology-based, Syntax-based and Readability-based categories. These conditioned features control the sentence generation and also the discrimination. In the discrimination task, the conditioned features could be viewed as classes that any paraphrases could be classified to one of them. The taxonomy tree is proposed for the conditional text paraphrasing on Figure 4.

Conditional Text Paraphrasing is divided to three sub-classes. These sub-classes controls the paraphrasing of the output sentences, however, these sub-classes are all considered to be paraphrasing on their semantic meaning. In other words, they are all paraphrases that represent the same meaning, but they differ on their morphology, syntax and readability aspects. For instance, consider the given sentence *Elon Musk is the founder of SpaceX*. The following sentences are candidate paraphrases *Elon Musk created SpaceX*, *Elon Musk is going to create SpaceX*, but these sentences are in the past and future tenses respectively. Also, the sentences *The founder of SpaceX*

is Elon Musk, which syntax tree is $(ROOT (S (NP (NP (DT) (NN)) (PP (IN) (NP (NNP)))) (VP (VBZ) (NP (NNP) (NNP))))$), and *SpaceX is the founded by Elon Musk*, which syntax tree is $(ROOT (S (NP (NNP) (VP (VBZ) (VP (VBN) (PP (IN) (NP (NNP) (NNP)))))))$, are paraphrases but each of them have different structure and parse tree. These varieties are divided to three sub-classes, Morphology-based Past Tense, Future Tense, Wh-clause, Syntax-based Syntax-tree and Readability-based Native, Non-native, Simple-style, Non-simple Style. The last sub-class focuses more on the style of writing. As humans, it is easy to us to identify whether the text is complex or written by a non-native person, so, the target is to be able to either generate or detect texts that have different styles.

Formally, for the generation task in conditional paraphrasing, given a reference sentence S_1 , where $S_1 = \{w_1, w_2, \dots, w_n\}$, and in addition, a conditioned class is given as a feature C , the target is to generate candidate(s) sentence(s) that are semantically equivalent to the reference sentence and also applies the conditioning property C .

On the other hand, for the discrimination task, given two sentences (S_1, S_2) , where $S_1 = \{w_1, w_2, \dots, w_n\}$ and $S_2 = \{w_1, w_2, \dots, w_m\}$, the target is to classify whether they are paraphrases or not, further more, it is possible to check what type of paraphrases that they are conditioned on. This transforms the original classification problem to multi-class classification problem. Shown in Figure 5, which is derived from the above taxonomy, the number of parent classes is two, number of conditioning classes is four and the number of conditioning sub-classes is seven. That makes the total number

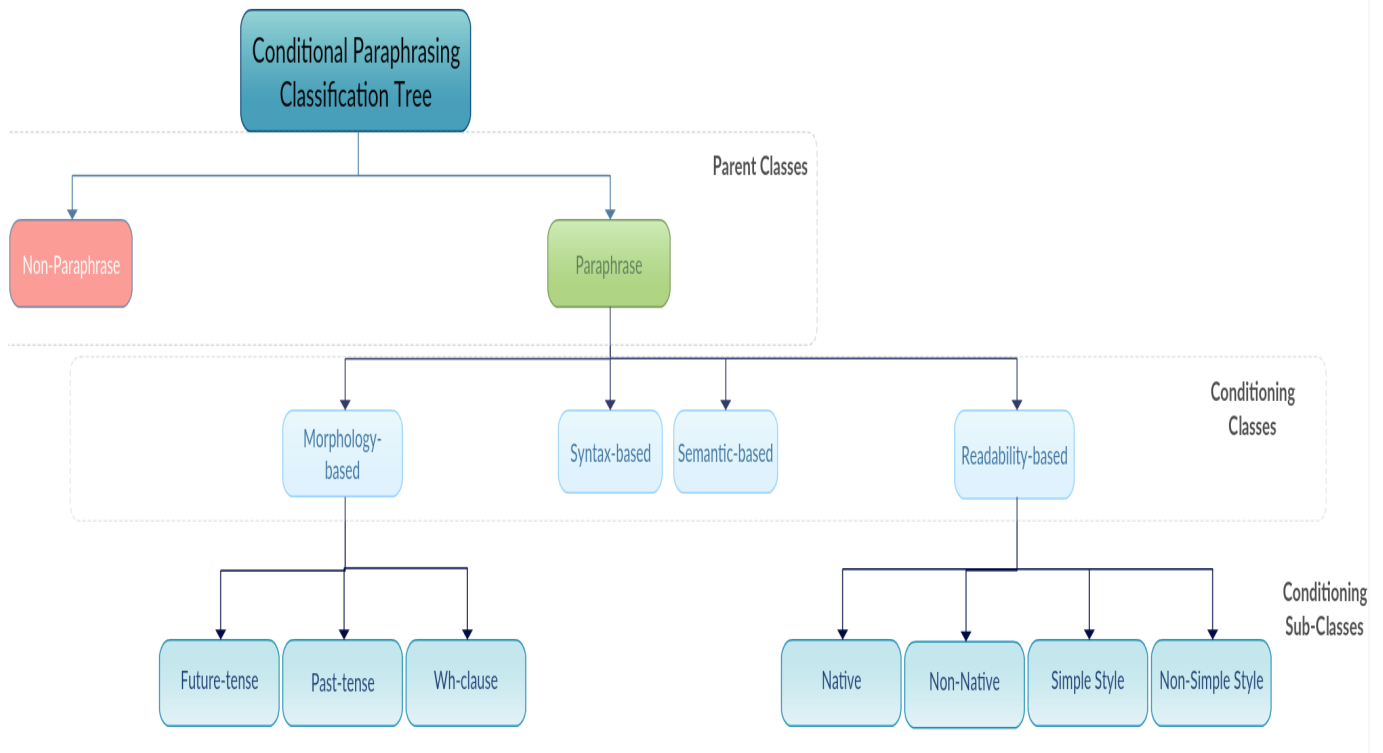


Fig. 5. Conditional Paraphrasing Classification Tree

of classes thirteen.

VII. CONDITIONAL TEXT PARAPHRASING APPLICATIONS

Conditional paraphrasing has several application at which it could be used. Mainly, it could be used in question-answer generation and domain-specific data augmentation. The constraints could be encoded and fed to the generation model as a conditional feature to handle the generated texts like such approaches [25], [26]. Regarding to the discrimination model, the problem turned to be a multi-classification problem. Given that, several transfer learning [52] and multi-task learning [53] could be applied to domain-specific objectives.

Conditional text paraphrasing could be very helpful in learning distributed representations of words and sentence. Currently, supervised learning based models dominates the field of sentences and words representation [40], so, the constraints, specifically syntax-based features, that are provided could help in much better representations for text that would effect on better natural language understanding systems.

VIII. CONCLUSION

This work showed a survey for text paraphrasing and its recent researches and efforts that work in the directions of paraphrase generation and discrimination. Also, it proposed our definition and taxonomy of conditional text paraphrasing task and how its great impact on existed applications and problems. For the future work, we are looking forward to working on datasets for conditional text paraphrasing, alongside doing several experiments for this task.

REFERENCES

- [1] S. Fernando and M. Stevenson, "A semantic similarity approach to paraphrase detection," in *Proceedings of the 11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics*, 2008, pp. 45–52.
- [2] V. Ngoc Phuoc An, S. Magnolini, and O. Popescu, "Paraphrase identification and semantic similarity in twitter with simple features," 2015.
- [3] J. Ramos *et al.*, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, vol. 242, 2003, pp. 133–142.
- [4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.
- [5] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [6] R. Kiros, Y. Zhu, R. R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler, "Skip-thought vectors," in *Advances in neural information processing systems*, 2015, pp. 3294–3302.
- [7] C. De Boom, S. Van Canneyt, S. Bohez, T. Demeester, and B. Dhoedt, "Learning semantic similarity for very short texts," in *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1229–1234.
- [8] C. Brockett and W. B. Dolan, "Support vector machines for paraphrase identification and corpus construction," in *Proceedings of the third international workshop on paraphrasing (IWP2005)*, 2005.
- [9] B. Agarwal, H. Ramampiaro, H. Langseth, and M. Ruocco, "A deep network model for paraphrase detection in short text messages," *Information Processing & Management*, vol. 54, no. 6, pp. 922–937, 2018.
- [10] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.
- [11] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, pp. 2042–2050. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2969033.2969055>

- [12] W. Yin and H. Schütze, "Convolutional neural network for paraphrase identification," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 901–911.
- [13] J. Mueller and A. Thyagarajan, "Siamese recurrent architectures for learning sentence similarity," in *AAAI*, vol. 16, 2016, pp. 2786–2792.
- [14] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 539–546.
- [15] A. Fader, L. Zettlemoyer, and O. Etzioni, "Open Question Answering Over Curated and Extracted Knowledge Bases," in *KDD*, 2014.
- [16] P. A. Duboue and J. Chu-Carroll, "Answering the question you wish they had asked: The impact of paraphrasing for question answering," in *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*. Association for Computational Linguistics, 2006, pp. 33–36.
- [17] Z. Hu, Z. Yang, R. Salakhutdinov, and E. P. Xing, "On unifying deep generative models," *arXiv preprint arXiv:1706.00550*, 2017.
- [18] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [20] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
- [21] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.
- [22] N. Duan, D. Tang, P. Chen, and M. Zhou, "Question generation for question answering," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 866–874.
- [23] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *arXiv preprint arXiv:1511.06349*, 2015.
- [24] L. Yu, W. Zhang, J. Wang, and Y. Y. Seqgan, "sequence generative adversarial nets with policy gradient. arxiv preprint," *arXiv preprint arXiv:1609.05473*, vol. 2, no. 3, p. 5, 2016.
- [25] Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing, "Toward controlled generation of text," *arXiv preprint arXiv:1703.00955*, 2017.
- [26] M. Iyyer, J. Wieting, K. Gimpel, and L. Zettlemoyer, "Adversarial example generation with syntactically controlled paraphrase networks," *arXiv preprint arXiv:1804.06059*, 2018.
- [27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.
- [28] A. Prakash, S. A. Hasan, K. Lee, V. Datla, A. Qadir, J. Liu, and O. Farri, "Neural paraphrase generation with stacked residual lstm networks," *arXiv preprint arXiv:1610.03098*, 2016.
- [29] A. Gupta, A. Agarwal, P. Singh, and P. Rai, "A deep generative framework for paraphrase generation," *arXiv preprint arXiv:1709.05074*, 2017.
- [30] Z. Li, X. Jiang, L. Shang, and H. Li, "Paraphrase generation with deep reinforcement learning," *arXiv preprint arXiv:1711.00279*, 2017.
- [31] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002, pp. 311–318.
- [32] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, 2004.
- [33] S. Banerjee and A. Lavie, "Meteor: An automatic metric for mt evaluation with improved correlation with human judgments," in *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, 2005, pp. 65–72.
- [34] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul, "A study of translation edit rate with targeted human annotation," in *Proceedings of association for machine translation in the Americas*, vol. 200, no. 6, 2006.
- [35] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [36] J. Ganitkevitch, B. Van Durme, and C. Callison-Burch, "Ppdb: The paraphrase database," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013, pp. 758–764.
- [37] E. Pavlick, P. Rastogi, J. Ganitkevitch, B. Van Durme, and C. Callison-Burch, "Ppdb 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, vol. 2, 2015, pp. 425–430.
- [38] Z. Chen, H. Zhang, X. Zhang, and L. Zhao, "Quora question pairs," 2018.
- [39] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," *arXiv preprint arXiv:1508.05326*, 2015.
- [40] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," *arXiv preprint arXiv:1705.02364*, 2017.
- [41] A. Fader, L. Zettlemoyer, and O. Etzioni, "Paraphrase-driven learning for open question answering," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2013, pp. 1608–1618.
- [42] L. Dong, J. Mallinson, S. Reddy, and M. Lapata, "Learning to paraphrase for question answering," *arXiv preprint arXiv:1708.06022*, 2017.
- [43] D. Kauchak and R. Barzilay, "Paraphrasing for automatic evaluation," in *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2006, pp. 455–462.
- [44] J. Berant and P. Liang, "Semantic parsing via paraphrasing," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2014, pp. 1415–1425.
- [45] P. Arthur, G. Neubig, S. Sakti, T. Toda, and S. Nakamura, "Semantic parsing of ambiguous input through paraphrasing and verification," *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 571–584, 2015.
- [46] R. Barzilay and K. R. McKeown, "Extracting paraphrases from a parallel corpus," in *Proceedings of the 39th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2001, pp. 50–57.
- [47] C. Bannard and C. Callison-Burch, "Paraphrasing with bilingual parallel corpora," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005, pp. 597–604.
- [48] C. Callison-Burch, "Syntactic constraints on paraphrases extracted from parallel corpora," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2008, pp. 196–205.
- [49] C. Buck, J. Bulian, M. Ciaramita, W. Gajewski, A. Gesmundo, N. Houslyb, and W. Wang, "Ask the right questions: Active question reformulation with reinforcement learning," *arXiv preprint arXiv:1705.07830*, 2017.
- [50] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [51] X. Chen, Y. Duan, R. Houthoofd, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in *Advances in neural information processing systems*, 2016, pp. 2172–2180.
- [52] S. J. Pan, Q. Yang *et al.*, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [53] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.