POLITECNICO
MILANO 1863

MSc Computer Science and Engineering

Topic-Guided Text Generation with Adversarial Learning

Author
Federico Betti

Coordinator
– –

September, 2019

Contents

Abstract

Nowadays Deep Learning Networks are used in every field of Artificial Intelligence because they have proved to be very high performance methods on complex tasks when a huge amount of data is available. However, while there are many areas in which this technology is considered complete and applicable, researchers are very active to discover new techniques that may in the future give rise to incredible applications.

This work aims to give an in-depth survey on modern Deep Learning methods used in the Natural Language Processing with a deeper focus on text generation. There will be a comparison on different promising technologies, namely LSTM-based models, adversarial architectures and transformer-based systems. Some of these new cutting-edge technologies will be present before proposing a new architecture for text generation that leverage on very recent state-of-the-art works. The proposal consists in enhancing the generation procedure with a guidance on the topic that the generated text should talk about in order not have unconditioned generated text anymore. The goal is to insert the human inside the generation process giving it the possibility to force the generation using simple sentences as input.

In order to carry out the work, it was necessary to combine techniques and skills of Natural Language Processing and Deep Learning to understand and make the contribution in an ever-changing sector where every month ameliorative results are published.

1 Introduction

(At least 2-3 pages of introduction)

Deep Learning techniques have proven in recent years to have great performances in various areas that can be connected with the real world. Clear examples of research papers and successful applications can be found in Computer Vision (cit cit cit), Robotics ([1] cit cit), Audio signal processing (cit), autonomous driving ([2] [3]), Medicine ([4]) and, last but not least, Natural Language Processing (NLP) (cit cit).

Natural Language Generation (NLG) is a crucial topic inside the Natural Language Processing area and consists in the process of generating text starting from other sources. Now it is the center of major work that are developed by the biggest research centres in the world. In the past NLP tasks have been approached with classical grammars and ontology based methods. Now, after the deep learning revolution, the availability of big data in text, the capability of new GPU generations, NLP and the text generation research performed disruptive improvements. These new techniques have brought new life to the world of NLP research and allowed us to imagine increasingly innovative solutions with applications that in the future will be able to interact with humans through the use of language. It is therefore clear that the topic is of great relevance in scientific research with numerous conferences and journals where scientists can share the discoveries made.

The objective of the thesis is therefore to study these modern techniques of text generation in order to try to replicate them and try to propose an improvement in a specific direction. The world of Text Generation is made up of several parallel trends that use different technologies and methodologies, which have strengths and weaknesses. As can be seen from this fact, the world is still uncertain but in continuous evolution.

1.1 Context

NLG is the process of generating text starting from any type of data: starting from structured information in a database, starting from images or video frame or starting directly from other text.

This work is placed in the middle between these new Natural Language Processing approaches and brand new Deep Learning techniques. It is based on Deep Learning models

called Generative Adversarial Networks. This architecture was first proposed in 2016 by a Montreal researcher called Ian Goodfellow [5] and they have become firmly part of the scientific community that has immediately understood the potential.
These models, which will be analyzed in more detail in the work, have the ability to generate any distribution of data. This process is done thanks to models of complex neural networks that, based on real data, try to replicate these data with the greatest possible fidelity so that not even a human is able to distinguish the true samples from the generated ones. These models have shown incredible results in the field of Computer Vision and in recent years are becoming popular to model any type of data distribution, continue or discrete.

1.2 Proposed Approach

The thesis has multiple goals and objectives:

1. Provide a extensive survey on modern deep learning models used for text generation and topic modelling

2. Comparative analysis of different approaches, namely LSTM-based, Transformer-based with brute force (OpenAI) and new emerging GAN-Based methods

3. Propose a *new architecture for text generation*, capable to generate short sentences guided by a topic and a given semantics, comparing it with very recently published papers on which it is based. The new models is a combination of a GAN-based method for text generation, a topic model and modern techniques on GAN architectures taken from other sectors. Human is now integrated in the generation process: the system receives a human input, computes the main topics or arguments that are present inside these sentences and use them as guide towards the generation. This is used not only to help the generation process to create longer utterances but also to force it to build sentences that actually speak about specific topics.

4. Explore the features of the new GAN with multiple-discriminator that is useful in many different contexts.

5. Explore possible solutions and applications (possibly in Italian to leverage on the capability of training the model from scratch each time).

6. Provide a system prototype on Twitter data.

1.3 Future Applications

Although this topic is still studied a lot, it is clear which can be possible application in the future. This thesis could be considered as a small building block that goes directly to the direction of a future where these technologies, that now are just developed in highly innovative research centers, will be used commonly during daily-life activities.
Some of the possible applications that are based on Text Generation are:

- News Generation (newspapers and blog), conditional text generation wit a specific style
- Data-To-Text: create reports starting from structured data in databased or Excel sheet
- Retrieve information from text, audio, images to match content of different resources
- Man-Machine interaction for people with disabilities
- Dialogue Generation
- Machine Translation
- Text Summarization
- Image Captioning

2 Theoretical Description

This will take lot of pages because even if I don't want to speak about NN from the beginning since I think it has been done many times, I need to go into the detail to make someone who read the thesis understand 1) which are the techniques than applied and 2)that the work was done on something really deep tech. Secondo me da qui possono facilmente venire fuori 30 pagine di tesi, questo perchè non solo devo 'spiegare' le reti neurali ma devo andare nel dettaglio con tecniche moderne e complesse.

This part must be very clear with images to explain concepts and some mathematical formulas.

2.1 Natural Language Processing

Here the should be some pages of introduction on NLP concpets that will be useful during the thesis. I think that there are lot of interesting things to write here. When we started the thesis we already have a good background on that, thanks to courses and other projects I did.

2.1.1 Grammars

Qui è interessante parlare delle grammatiche usate storicamente in NLP e dei Language model probabilistici basati su un corpus e sulle occorrenze

2.1.2 Topic Models

In questo capitolo verranno analizzati nel dettaglio i Topic Model con particolare concetrazioe su LDA che per ora è quello che usaimo

LDA

2.1.3 Neural Language Models

Questo capitolo sarà abbastanza importante perchè bisogna presentare tecniche come quelle di Word2Vec che hanno creato un modello embedding della parola stessa 'concettuale'. Questo ha reso possible l'avvento di tutti i modelli futuri che quindi si basano su questa differente rappresentazione delle parole e non più utilizzando la parola semplicemente come un numero all'interno di un vocabolario.

2.1.4 Metrics

- **BLEU metric**: the idea is to measure the diversity between a human translation and machine one, however it's wildly used also in the NLG field. It is a modified version of the precision that takes into account how much a sample appears at maximum in the target text. It can also use bigram or n-gram and in that case it would measure how many times a specific sequence of n words appears also in the dataset. Using big n will decrease the overall score but it would be better from the result view point. However this metric is usually not so fast to compute and introduce bias into the model.

- **SelfBleu**: lower the better, it measures the diversity in the sentences produced. This is a very important metrics because used together with the normal Bleu score it is possible to know if the model has also new generation capabilities. Having a good Bleu score could mean that the model found a specific sentence or set of words that, if repeated, allow the Bleu score to grow. However SelfBleu would be much worse in that case. For this reason it is really important the combined metric that measures both *quality and diversity* in the generation process.

2.2 Neural Networks

Here in my opinion there should be an introduction on NN, quite fast since I think everybody will do it.
è importante presentare le Convolutional Neural Network sopratutto e come viene costruita in generale una rete profonda. Poi è importante andare nel dettaglio della Backpropagation in quanto avrà un ruolo importante successivamente. è importante specificare la funzione delle loss function e di come sono fatte normalmente le reti neurali per poi chiarificare la diversità con le GAN.

2.2.1 Memory Network

Than I would go in the direction of presenting new techniques: at the beginning simple 'Memory Network' such RNN, LSTM, Relational Memory with Self-Attention and Transformer.

2.2.2 GAN Architecture

Than there should be a focus on the type of components that were used during the developing of the thesis. So in principle I would speak about GAN Architecture. I've attended to a talk by Goodfellow (GAN inventors) back un 2016 and I have also its slides.
Mode Collapse problem:

2.2.3 GAN for text generation

Also here I think that there will be presented some of the ideas behind the GAN architecture used for text and why these are different from images **(due to the 'discrete' nature of text)**.
A big and important paragraph saying which are the main challenges in applying these technologies to text and what is the community doing to solve this. There are many different things to speak about:

- RL Based GAN:
- RL Free architecture:

3 Related Works

I decided to put this section after the Theoretical Description part since in order to better describe these related works the reader must have a good comprehension of all the basis on these models otherwise it would be completely lost.
Language models consist of systems that output the probability of a word given the context. Many times it is considered useful not only the word that can be generated but also the context itself. It, in fact, groups together very useful information that can be used for numerous activities within the world of Natural Language Processing.
How likely is it that a certain sequence of words will appear several times within a dataset or during inference? Unfortunately this event is very unlikely and for this reason models based simply on the frequency of appearance of these words within the dataset cannot obtain significant results. I think this chapter can be divided in three subsection:

3.1 From Beginning To Recurrent Neural Network Based Models

There are lot of related works and I think it would be interesting to start from the fist works that were using normal RNN to train models to predict next word [6] [7]
These models suffer of *exposure bias*: this problem was presented by Bengio[8] in 2015. This consists in the problem that training and inference are substantially different if in the training procedure you use the true token as input to generate just the next one, while in the inference you use the previously generated token to infer the next one and all the following ones. This is a big difference because in training is like you're starting always from zero instead in the inference there is an error that increases over time due to the usage of predicted words and not real ones (that trivially you don't have). Scheduled Sampling is a technique to solve it that uses a random variable in the train that states if to use a generated token or the true one, in order to make train and inference similar. Although it often increases performances, it was proved to be inconsistent by Huszar in 2015.[9] Also Professor Forcing [10] technique was proposed to solve this problem.

3.1.1 TopicRNN [11]

Anche questo modello verrà analizzato nel dettaglio visto che è rilevante all'interno del nostro modello

3.2 GAN Based

The adversarial training has been used in a comprehensive manner in numerous research and applications of Computer Vision. Also in the field of Natural Language Processing many of these techniques have been developed in recent years.
Two of the main solutions are presented in detail below:

3.2.1 SeqGAN [12]

This is an important paper inside the context of text generation using adversarial training because it introduces efficiently the usage of REINFORCE algorithm inside the generation process.
It will be analyzed more in detail.

3.2.2 RelGAN [13]

Relational Generative Adversarial Networks for Text Generation: this paper, published at the conference ICLR 2019, introduces many novelties inside this research field and achieved state of the art results. These novelties can be summerized in three points:

1. *Relational Memory Based Generator*: differently for many other models RelGAN does not use the typical RNN-based generator but authors has decided to use a Relational Memory [14]. The basic idea is to consider a steady set of memory slots and interact with these slots with a self-attention mechanism [15]. The latter [Fig 1] has achieved remarkably results in many application for its ability to learn which is the best part of the memory to take information from for the following tasks.

   Qua secondo me bisogna andare più nel dettaglio di cosa si fa effetivamente dentro una Relational Memory con anche formule etc perchè servono dopo per spiegare l'output del modello.

2. *Gumbel-Softmax Relaxation*: This is the method that RelGAN uses to sample from the output distribution of the generator. As stated before in this discrete data environment it is important to have a one-hot like vector as output and many techniques were invented to deal with this problem. This is a technique used to sample from a distribution. The problem is that sampling from a multinomial distribution can be

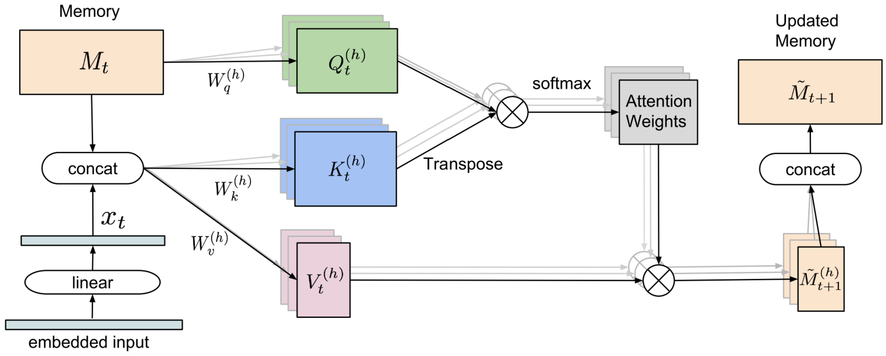

Figure 1: Relational Memory with Self-Attention Model

done using Eq. (1) .

$$y_t = one\_hot(argmax(o_t)) \tag{1}$$

However, it is not possible to pass the gradient through the operation of argmax because it is not drivable. For this reason some techniques, such as Gumbel-softmax relaxation, have been invented.

$$\begin{aligned} y_t &= \sigma(\beta(o_t + g_t)) \\ g_t &= -log(-log\ U_t) \\ U_t &\sim Uniform(0,1) \end{aligned} \tag{2}$$

where $\sigma(\cdot)$ is a softmax function which is done element-wisely on its argument.
It should also be consider the parameter β that is called *temperature* and it can control sample diversity and quality. Larger β encourages more exploration to improve the diversity of the output instead smaller values of β push to more exploitation, increasing the quality, therefore the Bleu score, conceding something on diversity.

3. *Multiple Representation in Discriminator*: A common used discriminator in such cases is a CNN-based classifier for its ability, thanks to filters of different sizes, of collecting information from different relations between elements in the input. In this case the input is a series of one-hot vector $[r_1 : ... : r_T]$ and, as proposed in previous chapter, it is important to have an embedding matrix to capture different concepts around the input sequence. It often happens that the embedding matrix is learnt

through gradient propagation by the system in order to have the best possible embedded representation. In this work they propose not have one such structure but many of them in parallel, as show in [Fig 2], in order to allow the network to learn different representation of the words in input. Thanks to the ablation study done at the end of the paper it is clear how this procedure is effective to push the generation forward keeping important information on the context and on the syntax situation of the sentence (*e.g.* if a comma or a preposition is needed)

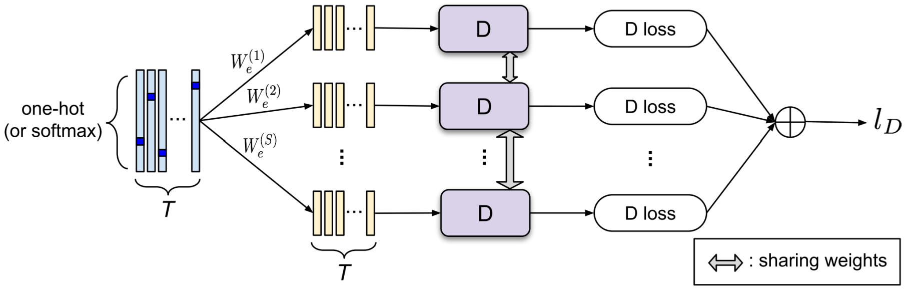

Figure 2: Multiple Discriminator

Lot of this work can be based on a Survey done 1-2 years ago [16]

There there are many other works that speaks about GAN and what I want to point out is that all these works are really new. New in the sense of the last months.

This is what I would like to show that everything is based on something really new. [17] [18] [11] [19] [13] [20] [21] [22] [12] [23]

This is not a GAN based model but it is done using Variational Autoencoders [24]

3.3 Transformer Based

Here there could be a big section related to all the works based on language models trained on big corpus like Bert and recent works on GPT by OpenAI

4 Topic-Guided GANs

Here there is the big part on the work that we have done during the thesis.
I think this should go through all the steps, the difficulties and the change we did in order to get to the result we'll have at the end.
It is important to give a meaning to the whole thesis

Many of the works presented above (TopicRNN, SeqGAN, RelGAN) are for sure interesting development for what concern pure text generation. However, if you want to take these models to a real world where there is a clear need to integrate the potential of these solutions with humans, it is mandatory to improve these architectures in order to consider a human input and then be able to manipulate the automatic generation of text according to the will of the user. With this in mind, the work aims to move away from the pure text generation towards a model of *conditioned text generation*. This is the idea on which the proposed new architecture is based.

4.1 Methods

The most promising paper, at the moment of the beginning of the work, was the RelGAN architecture published at one of the most famous and prestigious conference of Language Representation. For this reason the model is based on that solution. However it was a challenge to understand how to force the generation to create topic specific utterances.
The output of the generator at each step is a vector that represent a distribution over the number of words: assuming that the vocabulary size is V, the output logits of the generator is $o_t \in \mathbb{R}^V$. This output logits come from the relational memory output that was passed through a Linear Layer:

$$o_t = f_{\theta_1}(M_t) \tag{3}$$

The topic vector model, namely LDA

4.2 Results

Here there are the results in which it is compared also with other approaches

5 Comparison with other models

In questo capitolo verrà fatta un'analisi più dettagliata dei risultati ottenibili con altri modelli cercando di fare un paragone mostrando quali sono i punti di forza e di debolezza di ogni soluzione proposta

References

[1] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING," Tech. Rep. [Online]. Available: https://goo.gl/J4PIAz

[2] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, "DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving," Tech. Rep. [Online]. Available: http://deepdriving.cs.princeton.edu

[3] Y. Tian, K. Pei, S. Jana, and B. Ray, "Deeptest: Automated testing of deep-neural-network-driven autonomous cars," in *Proceedings of the 40th International Conference on Software Engineering*, ser. ICSE '18. New York, NY, USA: ACM, 2018, pp. 303–314. [Online]. Available: http://doi.acm.org/10.1145/3180155.3180220

[4] [Online]. Available: https://emerj.com/ai-sector-overviews/machine-learning-medical-diagnostics-4-current-applications/

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Tech. Rep. [Online]. Available: http://www.github.com/goodfeli/adversarial

[6] Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin, J. U. Ca, J. Kandola, T. Hofmann, T. Poggio, and J. Shawe-Taylor, "A Neural Probabilistic Language Model," Tech. Rep., 2003. [Online]. Available: http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf

[7] T. Mikolov, M. Karafit, L. Burget, J. Honza, and S. Khudanpur, "Recurrent neural network based language model," Tech. Rep., 2010. [Online]. Available: https://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov{_}interspeech2010{_}IS100722.pdf

[8] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks," Tech. Rep. [Online]. Available: https://papers.nips.cc/paper/5956-scheduled-sampling-for-sequence-prediction-with-recurrent-neural-networks.pdf

[9] F. Huszár, “HOW (NOT) TO TRAIN YOUR GENERATIVE MODEL: SCHEDULED SAMPLING, LIKELIHOOD, ADVERSARY?” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1511.05101.pdf

[10] A. Lamb, A. Goyal, Y. Zhang, S. Zhang, A. Courville, and Y. Bengio, “Professor Forcing: A New Algorithm for Training Recurrent Networks,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1610.09038.pdf

[11] A. B. Dieng, C. Wang, J. Gao, and J. Paisley, “TOPICRNN: A RECURRENT NEURAL NETWORK WITH LONG-RANGE SEMANTIC DEPENDENCY,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1611.01702.pdf

[12] L. Yu, W. Zhang, J. Wang, and Y. Yu, “SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient,” 2016.

[13] W. Nie, “RELGAN: RELATIONAL GENERATIVE ADVERSARIAL NETWORKS FOR TEXT GENERATION,” pp. 1–20, 2019.

[14] A. Santoro, R. Faulkner, D. Raposo, J. Rae $\alpha\beta$, M. Chrzanowski α, T. Weber α, D. Wierstra α, O. Vinyals α, R. Pascanu α, and T. Lillicrap $\alpha\beta$, “Relational recurrent neural networks,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1806.01822.pdf

[15] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention Is All You Need,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1706.03762.pdf

[16] S. Lu, Y. Zhu, W. Zhang, J. Wang, and Y. Yu, “Neural Text Generation: Past, Present and Beyond,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1803.07133.pdf

[17] L. Chen, S. Dai, C. Tao, D. Shen, Z. Gan, H. Zhang, Y. Zhang, and L. Carin, “Adversarial Text Generation via Feature-Mover’s Distance,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1809.06297.pdf

[18] C. de Masson, M. Rosca, and J. Rae Shakir Mohamed DeepMind, “Training Language GANs from Scratch,” Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1905.09922.pdf

[19] W. Fedus, I. Goodfellow, and A. M. Dai, "MaskGAN: Better Text Generation via Filling in the______," 2018.

[20] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, and J. Wang, "Long Text Generation via Adversarial Training with Leaked Information," Tech. Rep. [Online]. Available: www.aaai.org

[21] K. Lin, D. Li, X. He, Z. Zhang, and M.-T. Sun, "Adversarial Ranking for Language Generation," Tech. Rep., 2018. [Online]. Available: https://arxiv.org/pdf/1705.11001.pdf

[22] W. Wang, Z. Gan, W. Wang, D. Shen, J. Huang, W. Ping, S. Satheesh, and L. Carin, "Topic Compositional Neural Language Model," Tech. Rep., 2018. [Online]. Available: https://arxiv.org/pdf/1712.09783.pdf

[23] Y. Zhang, Z. Gan, K. Fan, Z. Chen, R. Henao, D. Shen, and L. Carin, "Adversarial Feature Matching for Text Generation," Tech. Rep., 2017. [Online]. Available: https://arxiv.org/pdf/1706.03850.pdf

[24] W. Wang, Z. Gan, H. Xu, R. Zhang, G. Wang, D. Shen, C. Chen, and L. Carin, "Topic-Guided Variational Autoencoders for Text Generation," Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1903.07137.pdf