Unsupervise Table To Text

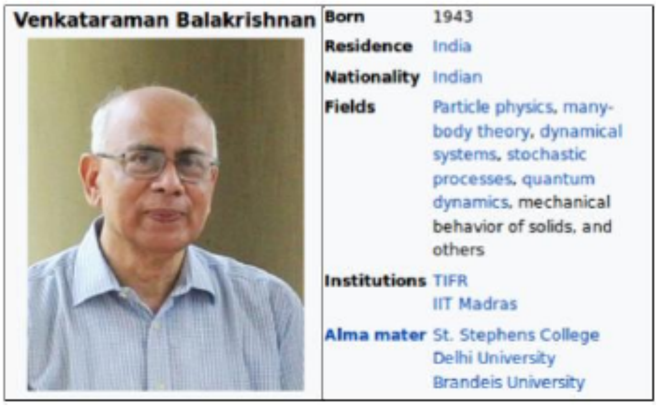
Erez Sheffi
erezsheffi@mail.tau.ac.il
066648114

The task of generating fluent utterance from structured data (i.e. tabular data) is an important natural language processing task. Usually, supervised methods are used for this task (Sha, L et al. 2018; Liu et al. 2018; Lebret et al. 2016). In this project, I investigate whether it is possible to learn to generate fluent utterance from structured data even without any parallel data. Based on the model proposed by (Lample et al., 2018), I propose a model that receives as input samples from two mono-corpora datasets (table and text datasets) and maps them into the same latent space. By learning to reconstruct in both domains from this shared feature space, the model effectively learns to generate text description from facts table without using any labeled data. I demonstrate my model on the WikiBio dataset, reporting BLEU scores of 21.35, without using even a single parallel table-text pair at training time.

I. TABLE-TO-TEXT GENERATION

A. Problem Description

Generating utterance from structured data is an important natural language processing task. It is often formulated into two subproblems: *content selection* which decides what contents should be included in the text and *surface realization* which determines how to realize the text based on selected contents. Structured data may come in various forms including databases of records, spreadsheets, expert system knowledge bases, simulations of physical systems, and so on. This project focuses on the table-to-text generation task - generating descriptions in fluent unstructured text from a corresponding table of facts, which involves comprehensive representation for the complex structure of a table. Such a table typically contains field-value pairs where the field is a property of the entity (e.g., color, fullname) and the value may be a set of possible assignments to this property (e.g., color = red) or a sequences of words (e.g., fullname = Erez Sheffi). Another example of this is the recently introduced task of generating one line biography descriptions from a given Wikipedia infobox (Lebret et al., 2016). The Wikipedia infobox serves as a table of facts about a person and the first sentence from the corresponding article serves as a one line description of the person. Figure 1. illustrates an example input infobox which contains fields such as Born, Residence, Nationality, Fields, Institutions and Alma Mater. Each field further contains some words (e.g., particle physics, many-body theory, etc.). The corresponding description is coherent with the information contained in the infobox. Note that the number of fields in the infobox and the ordering of the fields within the infobox varies from person to person. Recent table-to-text generation systems have been utilizing neural methods (Lebret et al. 2016; Mei et al. 2016; Wiseman et al. 2017). Such methods do not explicitly model any of the previously mentioned subproblems, rather they are trained in an end-to-end fashion. Up until now these models relied on supervised methods. Such models can achieve high performance only if vast amount of parallel data can be provided. Unfortunately, parallel corpora examples are costly to build and are hard to obtain for many domains. Conversely, mono-corpora datasets is much easier to obtain. In this paper I propose an unsupervised method for table-to-text generation. My model is built upon existing unsupervised methods for training NMT models (Lample et al., 2018; Artetxe et al., 2018). While not being able to compete with supervised approaches using lots of parallel resources, I show in Section IV C that my model is able to achieve meaningful results (taking into account it was trained without using even a single parallel table-text pair).



Venkataraman Balakrishnan	
Born	1943
Residence	India
Nationality	Indian
Fields	Particle physics, many-body theory, dynamical systems, stochastic processes, quantum dynamics, mechanical behavior of solids, and others
Institutions	TIFR IIT Madras
Alma mater	St. Stephens College Delhi University Brandeis University

FIG. 1: Sample Infobox with description : V.Balakrishnan (born 1943 as Venkataraman Balakrishnan) is an Indian theoretical physicist who has worked in a number of fields of areas, including particle physics, many-body theory, the mechanical behavior of solids, dynamical systems, stochastic processes, and quantum dynamics.

B. Problem Formalization

Following previous work (Lebret et al. 2016), the problem of generating text from a table is modeled by a table-conditioned language model for constraining text generation to include elements from fact tables. Thus, similarly to regular language model, table-to-text generation is formulated as the inference over a probabilistic model whose goal is - given a table T, generate a sentence $S^* = w_1, ...w_p$ such that $P(S^*|T)$ is maximized.

II. APPROACH

The task of generating text from a table can be (approximately) viewed as a language translation task (its only an approximation since structured text have different statistical characteristics than unstructured text). Inspired by recent work in NMT, specifically NMT seq2seq models who are trained in an unsupervised manner (Lample et al., 2018; Artetxe et al., 2018), I follow the same principles:

- **Language Modeling:** Given large amounts of mono-corpora data, we can train language models on both source and target languages. These models express a data-driven prior about how sentences should read in each language, and they improve the quality of the translation models by performing local substitutions and word reorderings.
- **Back-translation:** couple the source-to-target translation system with a backward model translating from the target to source. The goal of this model is to generate a source sample for each target sample in the mono-corpora datasets.
- **Sharing Latent Representations:** A shared encoder representation acts like an interlingua, which is translated in the decoder target language regardless of the input source language. This ensures that the benefits of language modeling, implemented via the denoising autoencoder objective, nicely transfer to translation from noisy sources and eventually help the NMT model to translate more fluently.
- **OOV handling:** using BPE encoding to eliminate the presence of unknown words in the output translation

Using the above framework for NMT, the system actually jointly learns to perform two tasks - translating between two languages in both directions (i.e. en-¿fr, fr-¿en). Thus in both cases the model expects a sentence (i.e. unstructured text) at its input. However, for the table-to-text generation problem the system learns to translate from a table to its corresponding text and the other way around. Hence the framework needs to be modified such that it will be able to handle inputs that can be either text or table. Additionally it should be able to generate both text and table.

To achieve the following modifications, I performed the following modifications:

1. **Encoder:** Usually each token fed to the encoder represents a word or subword from a given vocabulary. I modified the embedding layer to expect a token that represent field; content word pair, specifically, a tokens embedding is the concatenation of a *content word* embeddings with its correspond *field entity* embeddings.
2. **Decoder:** Usually the task of the decoder is to predict at each step the next word. I modified the decoder such that, at each step it predicts both next word and its corresponding field.

Encoder - Table Representation:

A table T is a unordered set of n records (aka field-value pairs) $R_1 = (f_1; v_1), R_2 = (f_2; v_2), , R_n(f_n; v_n)$. To represent the table in a meaningful way in a seq2seq framework, the entire table is transformed into a large sequence. This is achieved by tokenizing the records and serializing the produced tokens into a sequence. A record $R_i = (f_i; v_i)$ is tokenized by splitting it into $k^{(i)}$ tokens $t_1, ..., t_{k^{(i)}}$. where $t_j = (f_i; \{v_i\}_j)$, $\{v_i\}_j$ is the j's word of v_i and $k^{(i)}$ is the number of words in v_i.

Encoder - Table Representation:

In order for the embedding layer to be able to handle both input types (*Table* or *text*), each word w_i in the input sentence is converted to a token $(f; w_i)$, where f always equals to the *NULL* field entity (in order to signify to the encoder to disregard the field entity).

III. METHOD

This section describes the proposed unsupervised table-to-text method. Section III A first presents the architecture of the proposed system, and Section III B then describes the method to train it in an unsupervised manner.

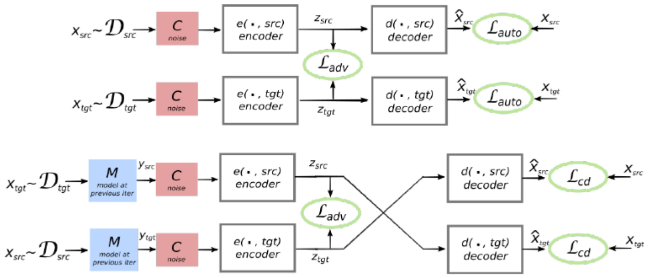


FIG. 2: Illustration of the proposed architecture and training objectives (Lample et al., 2018). The architecture is a sequence to sequence model, with both encoder and decoder operating on table/text corpus. Top (auto-encoding): the model learns to denoise sentences in each domain. Bottom (translation): like before, except that we encode from another language, using as input the translation produced by the model at the previous iteration (light blue box). The green ellipses indicate terms in the loss function.

A. Network Architecture

Figure 2 shows the general architecture of the system. The proposed system follows a fairly standard seq2seq framework with a single encoder single decoder architecture with an attention mechanism (Bahdanau et al., 2015). More concretely, I use a two-layer bidirectional RNN in the encoder, and another two-layer RNN in the decoder. All RNNs use GRU cells with 600 hidden units. I use a pre-trained embedding layer that correspond to a BPE vocabulary (Benjamin et al., 2018) which is kept fixed during training. The embeddings have dimensionality of 300 and the vocabulary size is 10000. As for the attention mechanism, the global attention method proposed by Luong et al. (2015b) is used with the general alignment function. The network also utilizes a discriminator to achieve better shared latent representation. The discriminator is a 2 layer FC network with hidden layers size of 150.

B. Training Method

The training method follows the methods introduced in related unsupervised NMT(Lample et al., 2018; Artetxe et al., 2018):

1. **Language Modeling:** language modeling is accomplished via denoising auto-encoding, by minimizing:

$$\begin{aligned} L^{lm} = \; & \mathbb{E}_{x \sim S}[-logP_{s \to s}(x|C(x))] + \\ & \mathbb{E}_{y \sim T}[-logP_{t \to t}(y|C(y)) \end{aligned}$$

where C is a noise model with some words dropped and swapped as in Lample et al. (2018). $P_{s \to s}$ and $P_{t \to t}$ are the composition of encoder and decoder both operating on the source and target sides, respectively.

2. **Back-translation:** Let us denote by $u^*(y)$ the sequence that represents a table inferred from sentence $y \in T$ such that $u^*(y) = argmaxP_{t \to s}(u|y)$. Similarly, let us denote by $v^*(x)$ the sentence in the target language inferred from table $x \in S$ such that $v^*(x) = argmaxP_{s \to t}(v|x)$. The pairs $(u^*(y), y)$ and $(v^*(x), x)$ constitute automatically-generated parallel corpora samples which, following the back-translation principle, can be used to train the two MT models by minimizing the following loss:

$$\begin{aligned} L^{back} = \; & \mathbb{E}_{y \sim T}[-logPs \to t(y|u^*(y))] + \\ & \mathbb{E}_{x \sim S}[-logPt \to s(x|v^*(x))] \end{aligned}$$

3. **Adversarial training:** constraining source and target latent representations to have similar distribution. In order to add such a constraint, I train a neural network, referred to as the discriminator, to classify between the encoding of source (table) samples and the encoding of target (sentences) samples (Ganin et al., 2016). The discriminator receives the encoder output and is expected to predict its corresponding class (table/text):

$$P_D\left(t|\text{enc}\left(x_i, t_i\right)\right)$$

To achieve the above goal, the discriminator is train by minimizing the following cross-entropy loss:

$$L^D = -\,\mathbb{E}_{(x_i,t_i)}\left[logP_D\left(t_i|\text{enc}\left(x_i, t_i\right)\right)\right]$$

Where (x_i, t_i) corresponds to input sample and type id pairs uniformly sampled from the two mono-corpora datasets.

The encoder is trained instead to fool the discriminator:

$$L^{adv} = -\,\mathbb{E}_{(x_i,t_i)}\left[logP_D\left(t_j|\text{enc}\left(x_i, t_i\right)\right)\right]$$

With $t_j =$ "$Table$" if $t_i =$ "$Text$", and vice versa.

IV. EVALUATION

In this section, I first describe the dataset I used, then I introduce the baselines I considered, and finally I report empirical validation proving the effectiveness of the suggested method.

TABLE I: BLEU-4 for unsupervised Tabe2Text model (last row), statistical language model (first four rows) and vanilla seq2seq model (fifth row)

Model	BLEU
KN	2.21
Template KN	19.80
NLM	4.17
Table NLM	34.70
Seq2seq	42.06
Unsupervised Table2Text	21.35

A. Dataset

I use the **WikiBio** dataset introduced by Lebret et al. (2016). It consists of 728,321 biography articles from English Wikipedia. A biography article corresponds to a person (sportsman, politician, historical figure, actor, etc.). Each Wikipedia article has an accompanying infobox which serves as the structured input and the task is to generate the first sentence of the article (which typically is a one-line description of the person). The same train, valid and test sets which were made publicly available by Lebret et al. (2016) are used (with minor filtering).

B. Baselines

I compared the proposed unsupervised Table-2-Text model with known results of several statistical language models and the vanilla encoder-decoder model. The baselines are listed as follows:

- **KN and Template KN** The Kneser-Ney (KN) model is a widely used language model proposed by Heafield et al. (2013). Template KN is a KN model over templates. (Lebret et al. 2016) train an interpolated Kneser-Ney (KN) language model for comparison with the KenLM toolkit. They also train a KN language model with templates.

- **NLM:** A naive statistical language model proposed by (Lebret et al. 2016) for comparison. The model uses only the field content as input without field information.

- **Table NLM:** The most competitive statistical language model proposed by (Lebret et al. 2016), which includes local and global conditioning over the table by integrating related field and position embedding into the table representation.

- **Vanilla Seq2seq:** The vanilla seq2seq neural architecture is also provided as a strong baseline which uses the concatenation of word embedding, field embedding and position embedding as the model input. The model can operate local addressing over the table by the natural advantages of LSTM units and word level attention mechanism.

C. Results

Generation Assessment:

The overall performance of text generation from fact tables is listed in Table I. An immediate observation is that the proposed model is outperformed by the supervised neural models. Still, achieving BLEU score of 21.35 without using even a single parallel table-text pair at training time is remarkable. Another observation is that (supervised) neural models outperform statistical language models.

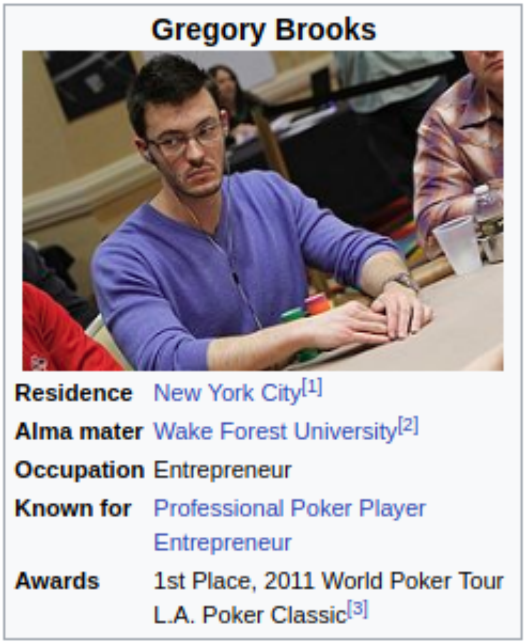
Text-to-Table:

As mentioned in Section II, the model also learns the translate in the opposite direction (i.e. generating a *Table* of facts from a /textitText description. For this task the model achieved much lower BLEU result (14.46). I speculate that improving this model will greatly improve the performance of the main task, Table-To-Text generation.

Case Studies:

TABLE II: Case study. A reference and a generated sentence by the proposed model.

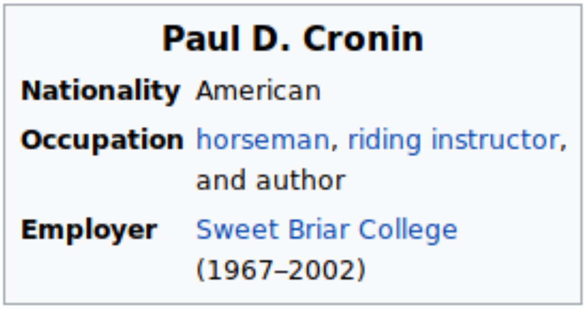
Reference	gregory brooks is an entrepreneur and former professional poker player. paul d. cronin is an american horseman , riding instructor , and author.
Text2Table	gregory brooks is an entrepreneur and professional poker player. paul d. cronin is an american horseman , riding instructor , and author sweet briar college



Gregory Brooks

Residence	New York City[1]
Alma mater	Wake Forest University[2]
Occupation	Entrepreneur
Known for	Professional Poker Player Entrepreneur
Awards	1st Place, 2011 World Poker Tour L.A. Poker Classic[3]

FIG. 3: Case study. Wikipedia infobox



Paul D. Cronin

Nationality	American
Occupation	horseman, riding instructor, and author
Employer	Sweet Briar College (1967–2002)

FIG. 4: Case study. Wikipedia infobox

V. CONCLUSION

I presented a new approach to Table-to-Text generation which follows the same methods of NMT models that are trained in an unsupervised manner using mono-corpora datasets only, without any alignment between fact tables and their correspond text description. The principle of such methods is to train jointly three types of models: Language model (by denoising auto encoder), Back translation model and adversarial model. The back-translation model is iteratively improved by the LM model which improves the reconstruction loss. Additionally a discriminator is used to align latent distributions of both the source and the target domains. The experiments I performed demonstrates that this approach is able to learn effective translation models without any supervision of any sort.

VI. FUTURE RESEARCH

Continuing the work done here, further research in unsupervised method for Table-to-Text models can try to utilize transformer layers instead of RNN layers. Additionally, it might be an interesting idea to use a NER model for the opposite direction model (Text to Table).

[1] George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014.

[2] George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Comput-*

ing and Pattern Recognition (SoCPaR), 2014 6th International Conference of, pages 312–318. IEEE, 2014.

[3] Sha, L.; Mou, L.; Liu, T.; Poupart, P.; Li, S.; Chang, B.; and Sui, Z. Order-planning neural text generation from structured data. *CoRR abs/1709.00155.*, 2017.

[4] Lebret, R.; Grangier, D.; and Auli, M. Neural text generation from structured data with application to the biography domain. *arXiv preprint arXiv:1603.07771*, 2016.

[5] Tianyu Liu, Kexiang Wang, Lei Sha, Baobao Chang, and Zhifang Sui. Table-to-text generation by structure-aware seq2seq learning. *CoRR, abs/1711.09724.*, 2017.

[6] Lebret, R.; Grangier, D.; and Auli, M. Neural text generation from structured data with application to the biography domain. *arXiv preprint arXiv:1603.07771*, 2016.

[7] G. Lample, M. Ott, A. Conneau, L. Denoyer, and M. Ranzato. Phrase-based and neural unsupervised machine translation. *arXiv preprint arXiv:1804.07755*, 2018.

[8] Liang, M. Jordan, and D. Klein. Learning semantic correspondences with less supervision. In Association of Computational Linguisitics. , 2009.

[9] Chen, D. L. and R. J. Mooney. Learning to sportscast: A test of grounded language acquisition. In International Conference on Machine Learning (ICML), *pages 128135, Helsinki.*, 2008.

[10] Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. Unsupervised neural machine translation. In International Conference on Learning Representations *ICLR*, 2018.

[11] G. Lample, A. Conneau, L. Denoyer, and M. Ranzato. Unsupervised machine translation using monolingual corpora only. In International Conference on Learning Representations. *ICLR*, 2018.

[12] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. In International Conference on Learning Representations *ICLR*, 2015.

[13] Denny Britz, Anna Goldie, Minh-Thang Luong, and Quoc V. Le. Massive exploration of neural machine translation architecture *CoRR, abs/1703.03906*, 2017.

[14] Benjamin Heinzerling and Michael Strube BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages , 2018.