ADVERSARILY-TRAINED NORMALIZED NOISY-FEATURE AUTO-ENCODER FOR TEXT GENERATION

Xiang Zhang
Courant Institute of Mathematical Sciences, New York University
Element AI
xiang@cs.nyu.edu

Yann LeCun
Courant Institute of Mathematical Sciences, New York University
Center for Data Science, New York University
Facebook AI Research, Facebook Inc.
yann@cs.nyu.edu

ABSTRACT

This article proposes Adversarially-Trained Normalized Noisy-Feature Auto-Encoder (ATNNFAE) for byte-level text generation. An ATNNFAE consists of an auto-encoder where the internal code is normalized on the unit sphere and corrupted by additive noise. Simultaneously, a replica of the decoder (sharing the same parameters as the AE decoder) is used as the generator and fed with random latent vectors. An adversarial discriminator is trained to distinguish training samples reconstructed from the AE from samples produced through the random-input generator, making the entire generator-discriminator path differentiable for discrete data like text. The combined effect of noise injection in the code and shared weights between the decoder and the generator can prevent the mode collapsing phenomenon commonly observed in GANs. Since perplexity cannot be applied to non-sequential text generation, we propose a new evaluation method using the total variance distance between frequencies of hash-coded byte-level \(n\)-grams (NGTVD). NGTVD is a single benchmark that can characterize both the quality and the diversity of the generated texts. Experiments are offered in 6 large-scale datasets in Arabic, Chinese and English, with comparisons against \(n\)-gram baselines and recurrent neural networks (RNNs). Ablation study on both the noise level and the discriminator is performed. We find that RNNs have trouble competing with the \(n\)-gram baselines, and the ATNNFAE results are generally competitive.

1 INTRODUCTION

Learning high-level, abstract representations of text or other discrete structures is a task that may have many applications in NLP, including text generation, translation and general understanding. This article makes 4 contributions: (1) a new class of model and objective functions called Adversarially-Trained Normalized Noisy-Feature Auto-Encoder (ATNNFAE) that is suited for encoding and generating sequence of symbols, such as text; (2) a recursive convolutional architecture for the encoder and decoder/generator that is designed to represent texts of any length at the byte level; (3) a measure of performance for byte-level text generators called \(n\)-Gram Total Variation Distance (NGTVD) that compares statistics of hash-coded \(n\)-grams; (4) experimental results on text generation by training on very large text corpora in multiple languages.

The basic architecture of ATNNFAE, shown in figure 2, consists of an auto-encoder where the internal code is normalized on the unit sphere and corrupted by additive noise. The AE is trained to reconstruct the input while eliminating the effect of noise. This effectively regularizes the information content of the code and forces the AE to maximize the distance between the codes of training samples. Simultaneously, a replica of the decoder (sharing the same parameters as the AE
decoder) is used as the generator and fed with random latent vectors, uniformly sampled on the unit sphere. An adversarial discriminator is trained to distinguish training samples reconstructed from the AE from samples produced through the random-input decoder replica, making the entire generator-discriminator path differentiable for discrete data like text. The combined effect of noise injection in the code and shared weights between the decoder and the generator can prevent the mode collapsing phenomenon commonly observed in GANs (Goodfellow et al., 2014).

The auto-encoder architecture we used is a byte-level recursive convolutional auto-encoder (Zhang & LeCun, 2018). This choice is made because convolutional networks have been shown to have better auto-encoding accuracy compared to recurrent neural networks (RNNs) at both word (Zhang et al., 2017b) and byte (Zhang & LeCun, 2018) levels. As a result of this choice, our model becomes a non-sequential (or non-autoregressive (Gu et al., 2018)) text generator. Since perplexity or bits-per-character cannot be directly applied to non-sequential text generation, we propose an evaluation method using the $n$-gram total variation distance (NGTVD). NGTVD can capture both the quality and the diversity of generated texts, since in either case it will result in a mismatch on the $n$-gram frequencies. Experiments are offered in 6 large scale datasets in Arabic, Chinese and English, with comparisons against $n$-gram baselines and recurrent neural networks (RNNs).

There are numerous attempts in text generation with or without GANs that merit discussion in this article. We discuss the difference between these ideas in section 2. ATNNFAE is introduced in section 3. The NGTVD evaluation method is introduced in section 4. Section 5 offers the experimental results, with comparisons against $n$-gram models and RNNs. Ablation study on the necessity of the discriminator and the denoising process is also included, which prompts us to do a hyper-parameter search on the level of noise. Furthermore, we showed additional improvements for RNNs and $n$-gram models via output selection, and for ATNNFAE models via $n$-gram correction. Before concluding this article, we also show some generated examples by interpolating in the feature space.

2 RELATED WORK

The challenge of applying GAN to text lies in the gap between the discrete nature of text data and the continuous nature of the discriminator. Most solutions can be classified into 3 categories.

1. The discriminator accepts a discrete sample. Because it is not differentiable with respect to the generator, some other solutions are required to provide gradients to the generator.
2. The discriminator accepts some intermediate representation in the generator. It is differentiable with respect to the sub-network in the generator that produces this representation.
3. The discriminator accepts a continuous sample in some transformed space. Some network is required to transform a discrete sample to this space, but the entire path is differentiable.

In the case that the discriminator accepts a discrete output, a few different approaches have been proposed. The idea of SeqGAN proposed by Yu et al. (2017) uses policy gradient (Sutton et al., 2000) to provide gradients to the generator, by casting the problem as a sequential decision making process. On the other hand, MaskGAN (Fedus et al., 2018) uses a discriminator that accepts a discrete word with its surrounding context, using the same policy gradient method in an actor-critic framework (Sutton & Barto, 1998) (Degris et al., 2012). Beyond reinforcement learning approaches, MaliGAN (Che et al., 2017) uses the maximum likelihood principle by assuming the discriminator has achieved optimum with respect to the current generator.

There are numerous attempts to apply the discriminator to the some intermediate representation of the generator. Professor forcing (Goyal et al., 2016) was proposed to use GAN on the hidden units to ensure generator stability, which improves the quality of long samples. Adversarial feature matching (Zhang et al., 2017a) was an idea to improve RNN generators using a convolutional discriminator on the hidden units. Adversarially regularized auto-encoder (ARAE) (Zhao et al., 2018) makes the generator match the feature from the encoder.

Our approach – Adversarially-Trained Normalized Noisy-Feature Auto-Encoder (ATNNFAE) – is one that belongs in the realm of letting the discriminator operate in some transformed sample space. Previously, Kusner & Hernandez-Lobato (2016) proposed to use a Gumbel-softmax distribution on the output of an RNN while the samples are provided as one-hot vectors. This approach could
collapse at large-scale, because the discriminator could easily distinguish between one-hot encoding and the generator’s output. Instead, we use an auto-encoder to transform a one-hot encoded sample to an unnormalized log probability space.

Beyond using GANs, an alternative approach is to use the variational auto-encoder (VAE) framework (Kingma & Welling, 2013). However, previous attempts such as (Bowman et al., 2016) have shown limited success. In VAE, the normalized feature from the encoder is optimized towards constant values, making it easy for the model to ignore the encoder. In ATTNFAE, the feature is corrupted with additive noise, and its strength is controllable via a hyper-parameter.

Similar to our approach, the generator in parallel WaveNet (van den Oord et al., 2017) maps from a sequence of random vectors to samples. It has an implicit sequential dependence via inverse-autoregressive flows (IAF) (Kingma et al., 2016). However, the parallel WaveNet paper (van den Oord et al., 2017) only experimented on supervised tasks in speech synthesis, and it is unknown whether an unconditional generative model is possible.

Finally, none of the discussed approaches can prevent mode collapsing of GANs, while our method can do so via denoising in a normalized feature space. In addition, using GAN for non-sequential text generation is a necessity, which is in contrast with RNNs, for which the maximum likelihood principle (“teacher forcing” (Williams & Zipser, 1989)) already exists for training.

3 Adversarially-Trained Normalized Noisy-Feature Auto-Encoder (ATNNFAE)

This section introduces the different components in ATNNFAE, using byte-level recursive convolutional auto-encoders (Zhang & LeCun, 2018). Additionally, the hyper-parameters used for training are detailed.

3.1 Normalized Noisy-Feature Auto-Encoder (NNFAE)

The NNFAE architecture in this article is the byte-level recursive convolutional auto-encoder (Zhang & LeCun, 2018), chosen for its better accuracy compared to RNNs. Good auto-encoding accuracy is required because its output is used as the target to the discriminator. Malik et al. (2018) offered improvements by removing the linear layers and using a fixed number of recursion groups, which give better results for long byte sequences. Following them, we use an NNFAE that has a fixed number of recursion groups without linear layers.

Figure 1 illustrates the NNFAE architecture in this article. All of the layers operate in 1 dimension, and ReLU (Nair & Hinton, 2010) is used as the non-linearity. Residual connections (He et al., 2016) are used in between every 2 layers. The encoder – denoted as $f$ – consists of a prefix group, a recursion group and a postfix group. The prefix contains $k$ convolutional layers with feature size 256 and kernel size 3. The recursion group contains $k$ convolutional layers with the same configuration, plus a max-pooling layer of size 2. Every time the recursion group is applied, the feature length is reduced by a factor of 2. All recursion groups share parameters. The postfix consists of $k$ convolutional layers and a normalization layer, making each feature vector norm 1.

The decoder is a reverse mirror of the encoder, denoted as $g$. The same normalization layer is used to normalize again after adding noise, which is part of the prefix that has $k$ convolutional layers. The recursion group contains $k$ convolutional layers in either the encoder or the decoder.
Figure 2: Adversarially-Trained Normalized Noisy-Feature Auto-Encoder (ATNNFAE) combines Normalized Noisy-Feature Auto-Encoder (NNFAE) and GAN. Note that the NNFAE decoder and the GAN generator are the same model $g$. ATNNFAE learns by alternating between 3 objectives. (1) The NNFAE objective $L_{NNFAE}$ optimizes the encoder $f$ and the decoder $g$ to reconstruct the sample $y$ from the feature corrupted by the additive noise $\eta$. (2) The discriminator objective $L_d$ optimizes the discriminator $d$ to distinguish between the reconstructed output $g(f(y)+\eta)$ from the NNFAE and the generator output $g(z)$, in which $z$ is set of vectors uniformly sampled from the unit sphere. (3) The generator objective $L_g$ optimizes the generator $g$ to “fool” the discriminator by making $d(g(z))$ approach the same target used for $d(g(f(y)+\eta))$ in the discriminator loss $L_d$.

convolutional layers, in which the first layer expands the feature length by a factor of 2 using sub-pixel convolution (or pixel shuffling) (Shi et al., 2016). All recursion groups share parameters. A postfix of $k$ convolutional layers follows, whose output is the unnormalized log-probabilities of bytes.

In both the encoder and the decoder, the number of recursion groups is fixed to 4. As a result, the feature has a length equal to $1/2^4 = 1/16$ of the input. For any input of size $s$, we tail-pad it to $16 \times \lceil s/16 \rceil$ using zero vectors to make the feature length exactly $\lceil s/16 \rceil$. The maximum input length is set to 1024 during training. A Gaussian noise with distribution $\mathcal{N}(0, \sigma^2)$ is used in the normalized feature space. The NNFAE is similar to the denoising process used for images by Doi & Lewicki (2005).

The NNFAE optimization problem looks like the following

$$\min_{f,g} L_{NNFAE} = \text{cross-entropy}(\text{softmax}(g(f(y)+\eta)), y),$$

in which $y$ is an one-hot encoded byte sample and $\eta$ is a random noise vector sampled from $\mathcal{N}(0, \sigma^2)$. Since $y$ is a one-hot vector, the cross-entropy (Solla et al., 1988) loss in $L_{NNFAE}$ degenerates to a negative-log likelihood at each position.

3.2 Generator and Discriminator

The decoder $g$ is also used as the generator. To generate a sequence of bytes, we sample $t$ vectors uniformly from the 256-D unit sphere as the feature. This corresponds to at maximum $16t$ bytes. The output from the generator $g$ is treated as a sequence of unnormalized log-probabilities, and the maximum is chosen at each position. $t$ is sampled from the length distribution in the training data. The end-of-sequence is determined by either the zero (NULL) byte, or the maximum length $16t$.

The discriminator – denoted as $d$ – has the same design as the encoder but does not share its parameters. It also does not contain the normalization layer. The scalar value required to form the adversarial objectives is obtained by simply averaging over the output values. We use a variant of HingeGAN (Miyato et al., 2018), which was the first GAN loss form that worked. The use of a Hinge loss for GAN can also be seen in energy-based GAN (EBGAN) (Zhao et al., 2016). The HingeGAN objectives are bounded, which can stabilize the training process. Other loss variants we tried include the original GAN (Goodfellow et al., 2014), the Wasserstein GAN (Arjovsky et al., 2017) and the Least Squares GAN (Mao et al., 2016). The paper by Lucic et al. (2017) suggests that different GAN loss forms perform similarly well for image generation, therefore we did not experiment with more after knowing HingeGAN works.

The adversarial training objectives look like the following,


\[
\begin{align*}
\text{minimize } & \quad \mathcal{L}_d = \max \{0, m - d(g(f(y) + \eta))\} + \max \{0, m + d(g(z))\}, \\
\text{minimize } & \quad \mathcal{L}_g = \max \{0, m - d(g(z))\},
\end{align*}
\]

in which \(y\) is a one-hot encoded byte sample and \(z\) is a sequence of random vectors sampled from the unit sphere. \(m\) is the margin of the Hinge loss. \(\mathcal{L}_d\) attempts to make the discriminator \(d\) give a value larger than \(m\) for the NNFAE’s output \(g(f(y))\), and give a value smaller than \(-m\) for the generator’s output \(g(z)\). Meanwhile, \(\mathcal{L}_g\) attempts to let the generator \(g\) “fool” the discriminator by making \(d(g(z))\) a value larger than \(m\). Compared to Miyato et al. [2018] and Zhao et al. [2016], there is also a margin in \(\mathcal{L}_g\), further stabilizing training. Furthermore, we find it necessary to use the feature noise in \(\mathcal{L}_d\) to prevent mode collapsing.

The adversarial optimization objectives are required because the NNFAE objective \(L_{\text{NNFAE}}\) is not enough to cover the entire feature space with acceptable output byte sequences. On the other hand, the adversarial objectives \(\mathcal{L}_d\) and \(\mathcal{L}_g\) are not enough to ensure the generator can output a diverse set of acceptable samples. Theoretically, if \(f\), \(g\) and \(d\) all have sufficient representation capacity, it would have been okay for \(g\) to output only one acceptable sample for all \(z\), with \(\mathcal{L}_g\) having achieved the minimum and \(\mathcal{L}_d\) stationed in the equilibrium.

In other words, GAN attempts to make the support of the generator’s output distribution a subset of the support of the sample distribution, which seems to be the reason for mode collapsing. The denoising process during auto-encoding could encourage diversity, since it “pushes away” the values in the feature space for different samples. When there are many samples, the prior knowledge that there are distant values in the feature space corresponding to acceptable samples is sufficient to prevent mode collapsing. Section 5 offers an ablation study between the discriminator and \(\sigma\).

### 3.3 Training Hyper-parameters

The entire optimization process is simply an alternating direction method by iterating through objectives 1, 2, and 3. The choice of margin \(m\) depends on the balance between the NNFAE objective and the adversarial objectives. Auto-encoding should perform well before adversarial training kicks in, which means that \(m\) should be small. We find \(m = 0.001\) works well. The model parameters are initialized using distribution \(\mathcal{N}(0, \sqrt{2/\tau}/1000)\) for the weights and 0 for the biases. \(\tau\) is the number of output units each input unit connects to. It is 1000 times smaller than the value suggested by He et al. [2015], which we find working well when used with residual connections [He et al., 2016] without the need for batch normalization [Ioffe & Szegedy, 2015].

The training algorithm proceeds by repeating 10 steps for each of the objectives using stochastic gradient descent (SGD) with momentum 0.9 [Polyak, 1964] [Sutskever et al., 2013]. Whenever a sample \(y\) is needed, it is randomly chosen from the training dataset with replacement. The learning rate begins with 0.001, and is halved for every 10,000,000 steps for each objective until the training stops at 40,000,000 steps.

### 4 Evaluation using n-Gram Total Variation Distance (NGTVD)

The most frequently used benchmark for text generation is perplexity. Unfortunately computing perplexity for a non-sequential generator is intractable in closed form and infeasible via Monte Carlo approximation (see appendix A). Therefore, we need to seek a new benchmark method.

The MaskGAN paper [Fedus et al., 2018] suggests that perplexity alone is not enough to characterize the quality of the generated text. They propose to use whether a generated word-level \(n\)-gram has appeared in the data as the benchmark. It was inspired by the Bilingual Evaluation Understudy (BLEU score) [Papineni et al., 2002]. However, as a benchmark for machine translation, BLEU score is applied on a per-sample basis and the aggregated value is able to characterize the distribution of \(n\)-grams. The mere 1 or 0 on whether an \(n\)-gram appears in the data could not take into consideration the frequency of \(n\)-grams. For large-scale datasets, this is misleading because a large number of infrequent \(n\)-grams and a small number of frequent \(n\)-grams would be considered equal.
Table 1: Datasets. Numbers in both articles and paragraphs are shown. Paragraphs are used as training or testing samples, making each dataset contain tens of millions of samples. They span 3 languages – Arabic, Chinese and English. The allgiga dataset is a combination of argiga, engiga and zhgiga, which forms a multi-modal distribution in the space of byte sequences.

| NAME  | ARTICLE TRAIN | ARTICLE TEST | PARAGRAPH TRAIN | PARAGRAPH TEST | LANGUAGE  |
|-------|---------------|--------------|-----------------|----------------|-----------|
| enwiki| 7,634,438     | 850,457      | 41,256,261      | 4,583,893      | English   |
| hudong| 1,618,817     | 180,278      | 53,675,117      | 5,999,920      | Chinese   |
| argiga| 3,011,403     | 334,764      | 27,989,646      | 3,116,719      | Arabic    |
| engiga| 8,887,583     | 988,513      | 116,456,520     | 12,969,170     | English   |
| zhgiga| 5,097,198     | 567,179      | 38,094,390      | 4,237,643      | Chinese   |
| allgiga| 16,996,184  | 1,890,456   | 182,540,556     | 20,323,532     | Multi-lingual |

Instead, we propose to use the total variation distance on the frequency of byte-level \( n \)-grams between generated data and validation data.

\[
\text{NGTVD} = \frac{1}{2} \sum_i |p(u_i) - q(u_i)| , \tag{4}
\]

in which \( p(u_i) \) and \( q(u_i) \) are frequencies of the \( n \)-gram \( u_i \) from generated data and validation data respectively. In practice, these values are computed over multiple generated samples as

\[
p(u_i) = \frac{\text{count}(u_i)}{\sum \text{count}(u_i)} . \tag{5}
\]

One problem of the benchmark above is that we could not use very large \( n \) because it would exhaust computational resources. Therefore, we also propose to use a hash table on the \( n \)-grams.

\[
\text{NGTVD}[N, M] = \frac{1}{2} \sum_{i=1}^{M} |p(i) - q(i)| , \tag{6}
\]

in which \( N \) is the maximum length of a byte \( n \)-gram, and \( M \) is the number of bins in the hash table. \( p(i) \) and \( q(i) \) are frequencies of the hash table entries from generated data and validation data respectively. The hope is that when \( M \) is large, it could capture the \( n \)-gram distribution well while still allowing a large \( N \). This is inspired by the success of the hashing trick (Weinberger et al., 2009) for various \( n \)-gram based models in NLP (for example, Vowpal Wabbit (Weinberger et al., 2009) and fastText (Joulin et al., 2017)). In this article, we use \( N = 256 \) and \( M = 1,000,000 \) on 1,000,000 generated samples from each model, denoting the benchmark as NGTVD[256, 1e9]. This benchmark is in the range \([0, 1]\) and can be applied to both sequential and non-sequential text generation models.

NGTVD is capable of capturing both the quality and the diversity. If the generated texts are not similar to the training data (quality), or if just a few acceptable texts can be generated (diversity), it will both result in a mismatch between the \( n \)-gram frequencies of the generated texts and the validation data.

5 Experiments and Analysis

For all of the experiments, we use the same datasets as in [Zhang & LeCun (2018)]. All of these samples are at the level of paragraphs, and all the texts are treated as sequences of bytes encoded in UTF-8. These datasets each have tens of millions of samples. Table 1 is a summarization.

5.1 Comparison with \( n \)-Gram Models and Recurrent Neural Networks (RNNs)

The simplest byte-level \( n \)-gram model defines a sequential generator constructed from the formula

\[
\text{Pr} [y_i|y_1, y_2, \ldots, y_{i-1}] = \frac{\text{count}(y_{i-n+1}y_{i-n+2}\ldots y_i)}{\sum_{y_i=1}^{256} \text{count}(y_{i-n+1}y_{i-n+2}\ldots y_i)} . \tag{7}
\]
Table 2: Results of $n$-gram models, RNNs, and ATNNFAEs on enwiki. NGTVD$[256, 1e9]$ can be computed for all models. Byte-level perplexities for sequential models are shown, and so are auto-encoding errors for ATNNFAE. We also have varying model sizes for both ATNNFAE and RNNs. ATNNFAE achieved better NGTVD$[256, 1e9]$ than either the $n$-gram models or the RNNs. In all cases, the larger the models are, the better the results.

| MODEL              | NGTVD$[256, 1e9]$ | PERPLEXITY | ERROR |
|--------------------|------------------|------------|-------|
|                    | TRAIN | TEST | TRAIN | TEST | TRAIN | TEST |
| ATNNFAE $k = 2$, $\sigma = 0.1$ | 0.0895 | 0.0942 | - | - | 28.71% | 28.71% |
| ATNNFAE $k = 4$, $\sigma = 0.1$ | 0.0885 | 0.0932 | - | - | 20.27% | 20.29% |
| ATNNFAE $k = 8$, $\sigma = 0.1$ | 0.0865 | 0.0913 | - | - | 20.08% | 20.09% |
| Simple 5-gram      | 0.1035 | 0.1071 | 4.2603 | 4.2478 | - | - |
| Complex $n$-gram   | 0.0975 | 0.1013 | 4.0045 | 3.9939 | - | - |
| Plain RNN level 1  | 0.2864 | 0.2864 | 6.3597 | 6.3540 | - | - |
| Plain RNN level 2  | 0.2708 | 0.2708 | 6.1451 | 6.1988 | - | - |
| LSTM level 1       | 0.1851 | 0.1877 | 4.5779 | 4.5740 | - | - |
| LSTM level 2       | 0.1747 | 0.1763 | 4.2945 | 4.2915 | - | - |
| GRU level 1        | 0.1823 | 0.1847 | 4.5063 | 4.5071 | - | - |
| GRU level 2        | 0.1665 | 0.1688 | 4.3207 | 4.3507 | - | - |

However, in practice if $n$ is small, the generated texts have low quality due to the lack of long-term dependency. On the other hand, if $n$ is large, the existence of long byte $n$-grams becomes sparse and text generation is frequently interrupted. Therefore, we define a new $n$-gram model as

$$\Pr[y_i | y_1, y_2, \ldots, y_{i-1}] = \frac{\sum_{Q=Q_R}^{R} \text{count}(y_{i-n+1}y_{i-n+2} \ldots y_i)}{\sum_{Q=Q_R}^{R} \sum_{y_{i-n+1}^{256}} \text{count}(y_{i-n+1}y_{i-n+2} \ldots y_i)}, (8)$$

which uses the sum of the counts of $n$-grams from size $Q$ to $R$. We could therefore set $R$ to be a large number to encourage long-term dependency. In practice, we use $Q = 5$ and $R = 64$, and consider all of the grams that have appeared more than 256 times in the training data. This modified $n$-gram model turns out to be a competitive baseline in both NGTVD$[256, 1e9]$ and perplexity. We name the model defined in equation 7 the “simple $n$-gram” model, and equation 8 the “complex $n$-gram” model. Appendix B presents some samples generated by the complex $n$-gram model.

In this article we also offer comparisons against multi-level stacked recurrent neural networks (RNNs), using 3 cell variants including the standard plain variant with linear cells, the long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997), and the gated recurrent unit (GRU) (Cho et al., 2014). They all have 1024 hidden units. They are trained using the maximum likelihood principle at each sequential step with the correct byte-sequence history, also called the “teacher forcing” algorithm (Williams & Zipser, 1989). The optimization algorithm used is SGD with momentum (Polyak, 1964) (Sutskever et al., 2013), using the same hyper-parameter settings as the ATNNFAE models. At test time, text generation proceeds by sampling one byte at a time and it is fed back to the model for the next step.

The results of $n$-gram models, recurrent networks and convolutional ATNNFAE models are presented in table 2. For any $k$, the number of parameterized layers in an ATNNFAE model is $18k$, because there are $6k$ convolutional layers in the encoder, the decoder/generator and the discriminator. Therefore, the network depth values in table 2 are 36, 72, and 144. The first conclusion from the table is that the ATNNFAE models achieved better NGTVD$[256, 1e9]$ than both $n$-gram models and RNNs, with better results as the models get deeper. Furthermore, RNNs actually struggle to compete with the $n$-gram models for sequential text generation in both NGTVD$[256, 1e9]$ and perplexity, suggesting that $n$-gram models are strong baselines.

5.2 Output Selection for $n$-Gram Models and RNNs

The results from RNNs in table 2 are somewhat unexpected in the sense that they are far worse than the baseline $n$-gram models. Besides the usual argument that RNNs lack the ability to model
Figure 3: The length histogram of generated texts on enwiki. The ATNNFAE model is the one with $k = 8$ and $\sigma = 0.1$, which matches with the length distribution of the dataset. All $n$-gram and RNN models strongly favor generating shorter texts, and RNNs prefer even shorter texts than both the simple and the complex $n$-gram models.

long-term dependencies due to gradient vanishing (Bengio et al. 1994) (Hochreiter et al. 2001), the other reason could be that RNNs prefer generating shorter texts. This can be visually observed from the text samples shown in appendix C for LSTM. Figure 3 also shows the length histograms of generated samples from RNNs, the $n$-gram models and an ATNNFAE with $k = 8$ and $\sigma = 0.1$ against the enwiki training data. The ATNNFAE model shows an advantage in matching with the length distribution from the training data.

To provide additional comparison without the influence from the difference between sample length distributions, we performed selection on the generated samples so that the filtered length distribution matches that of the training data, for $n$-gram models, LSTM and GRU. In practice we find it infeasible to do output selection for plain RNNs because its output length distribution is skewed too much. The results are presented in table 3 in which significant improvements are observed for $n$-gram models and RNNs. That said, the ATNNFAE results in table 2 still compare better than that of RNNs with output selection.

5.3 Ablation Study on the Discriminator and the Noise

To provide an ablation study on whether the discriminator is necessary in ATNNFAE, we compare between using NNFAE only and using ATNNFAE for a $k = 4$ model in table 4. Improvements from adding the discriminator can be observed for $\sigma \geq 0.1$, whereas for $\sigma \leq 0.05$ the discriminator has an adverse effect due to mode collapsing.

The results in table 4 suggest that there is a balance between the discriminator and the noise standard deviation $\sigma$ in ATNNFAE. On one hand, the discriminator attempts to make sure that all the outputs from the generator look like the NNFAE’s output; on the other hand, the noise is necessary to prevent mode collapsing. In order to improve the quality of generated text, we would prefer a small $\sigma$ so that the NNFAE’s output is accurate. However, we could not make the noise too small either, since the use of discriminator will result in a mode-collapsed model that lacks diversity. In this case, the encoder’s feature is concentrated on a small region in the space of $z$, which can still give good accuracy for auto-encoding.

| MODEL          | TRAIN | TEST |
|----------------|-------|------|
| Simple 5-gram  | 0.0743| 0.0795|
| Complex $n$-gram| 0.0643| 0.0703|
| LSTM level 1   | 0.1055| 0.1087|
| LSTM level 2   | 0.1233| 0.1261|
| GRU level 1    | 0.0950| 0.0986|
| GRU level 2    | 0.1294| 0.1321|

Table 3: Improved NGTVD[256, 1e9] for $n$-gram models and RNNs by selecting output samples to match the length distribution of the training data. Significant improvements over the results in table 2 observed. The results for $n$-gram are improved so much that they become the best numbers among all models in this article. The NGTVD[256, 1e9] results for ATNNFAE are still better than RNNs with output selection.
Table 4: Results between NNFAE and ATNNFAE, using $k = 4$. Comparing between the rows, ATNNFAE suffers from mode collapsing when $\sigma \leq 0.05$. When $\sigma \geq 0.1$, mode collapsing no longer happens, while the quality of generated texts degrades as $\sigma$ becomes larger because the auto-encoding errors are higher. Comparing between the NNFAE and ATNNFAE columns, when mode collapsing is prevented for $\sigma \geq 0.1$, the use of adversarial training with a discriminator improves ATNNFAE’s results over that of NNFAE. The last row is a result by performing a hyper-parameter search on $\sigma \in \{0.055, 0.06, 0.065, 0.070, 0.075, 0.08, 0.085, 0.09, 0.095\}$.

| $\sigma$ | NNFAE NGTVD$^{[256, 1e9]}$ | ATNNFAE NGTVD$^{[256, 1e9]}$ | $\sigma$ | NNFAE ERROR | ATNNFAE ERROR |
|----------|-----------------------------|-------------------------------|----------|--------------|---------------|
|          | TRAIN TEST                  | TRAIN TEST                  |          | TRAIN TEST   | TRAIN TEST    |
| 0.01     | 0.0960 0.1007 0.05% 0.05%   | 0.6241 0.6243 0.18% 0.18%   |
| 0.02     | 0.0955 0.1002 0.11% 0.12%   | 0.5626 0.5628 0.35% 0.35%   |
| 0.05     | 0.0918 0.0966 2.23% 2.24%   | 0.9943 0.9943 3.24% 3.24%   |
| 0.1      | 0.0932 0.0978 18.85% 18.85% | 0.0885 0.0932 20.27% 20.29% |
| 0.2      | 0.1050 0.1097 56.08% 56.07% | 0.1008 0.1055 57.09% 57.06% |
| 0.5      | 0.1819 0.1855 78.43% 78.39% | 0.1768 0.1805 79.46% 79.41% |
| (0.085)  | 0.0929 0.0972 16.27% 16.26% | 0.0874 0.0921 17.33% 17.56% |

As far as the models in this section is concerned, 0.1 is the smallest acceptable $\sigma$ that could make ATNNFAE work for enwiki. However, the auto-encoding accuracy at $\sigma = 0.1$ is not good enough to provide the best targets to the discriminator. This explains why there are frequent occurrences of “invented” words in appendix D. That said, from appendix C we could see that RNNs also “invent” words when trained on English data. The next section offers a method to improve the appearance of generated text by combining ATNNFAE with an $n$-gram model.

To achieve a better balance between $\sigma$ and the discriminator in ATNNFAE, we performed a hyper-parameter search on $\sigma$ for $k = 4$. As suggested by table 4, the best choice for $\sigma$ is somewhere in between 0.05 and 0.1. Therefore, we trained $k = 4$ ATNNFAE models with $\sigma \in \{0.055, 0.06, 0.065, 0.070, 0.075, 0.08, 0.085, 0.09, 0.095\}$. Then, we choose the smallest $\sigma$ that can obtain an ATNNFAE model without mode collapsing. The mode collapsing phenomenon is quite obvious by just inspecting the generated samples during training, therefore the hyper-parameter selection can be done without involvement of the testing data. We find that the best choice is $\sigma = 0.085$, and its result is presented as the last row in table 4.

5.4 $n$-Gram Correction for Better Text Appearance

In spite of the better NGTVD$^{[256, 1e9]}$ result for ATNNFAE, the text samples in appendix D appear noisy at the level of bytes. This demonstrates that text generation is challenging in terms of achieving smoothness at the level of bytes, while at the same time shows ATNNFAE’s potential in learning better high-level structure of the text. We want to point out that word-level text generation will not have such a intra-word smoothness problem by construction, and applying our models at the level words is also scalable and feasible. Even at the level of bytes, the scale of generated texts in our model is unprecedented, in the sense that the current practical limitation is 1024 bytes – corresponding to around 200-300 words on average for English. This is in addition to the fact that we can prevent mode collapsing via noise injection in the NNFAE.

That said, in this section we also explore one simple approach to improve the appearance of text – especially the intra-word smoothness for English – combining an ATNNFAE with the complex $n$-gram model. This is done by using the formula

$$\Pr[y_i|z, y_1, y_2, \ldots, y_{i-1}] \propto p(y_i|z) q(y_i|y_1, y_2, \ldots, y_{i-1}),$$

in which $p(y_i|z)$ is obtained from an ATNNFAE model and $q(y_i|y_1, y_2, \ldots, y_{i-1})$ from the complex $n$-gram model. Then, we have
The maximum likelihood conditioned on \( z \) in equation 10 can therefore be approximated via the beam search algorithm \cite{Graves2012, Boulander-Lewandowski2013} on the \( y_i \)'s. We use a beam of size 10. Appendix E shows 100 text samples generated with \( n \)-gram correction for the ATNNFAE model using \( k = 8 \) and \( \sigma = 0.1 \) for the enwiki dataset, which has better intra-word smoothness than the samples in appendix D with only ATNNFAE. However, in terms of benchmarks, this method achieved NGTVD\cite{256, 1e9} values of 0.0888 for the training data and 0.0936 for the testing data – worse than the ATNNFAE but better than the complex \( n \)-gram model in table 2.

For English, the intra-word smoothness can be numerically benchmarked by the proportion of generated words that belong to some pre-defined dictionary. We use all the words in the WordNet 3.0 distribution \cite{Miller1995} as the dictionary, and computed the intra-word smoothness in table 5. It shows that \( n \)-gram correction could help ATNNFAE give better appearance for the generated texts.

### 5.5 Interpolation in Feature Space

The following list shows the interpolation in the feature space from a short 128-byte paragraph to another one. The model is trained on the enwiki dataset with \( k = 8 \) and \( \sigma = 0.1 \). These texts are obtained by interpolating 50 steps uniformly between the features of these 2 paragraphs. Only the steps where changes occur are printed.

| MODEL | RESULT |
|-------|--------|
| Training data | 58.36% |
| Testing data | 58.37% |
| Complex \( n \)-gram | 48.89% |
| ATNNFAE without \( n \)-gram correction | 33.37% |
| ATNNFAE with \( n \)-gram correction | 40.82% |

It shows that the model attempts to interpret the feature space by outputting byte sequences that are as close to English as possible, often by inserting legitimate English words. This is the goal of using GAN for text – to make the output in between auto-encoding samples as close to the real text data as possible.

\[
\Pr [y_1, y_2, \ldots, y_k | z] = \prod_{i=1}^{k} \Pr [y_i | y_{i-1}, y_{i-2}, \ldots, y_1].
\]
Table 6: Results across different datasets. ATNNFAE achieved better NGTVD\([256, 1e9]\) for enwiki, hudong, engiga and zhgiga datasets compared to the complex \(n\)-gram baseline. For argiga, the result is close. For allgiga, it is significantly worse, which is because the ATNNFAE degenerates to learning mostly from zhgiga. Also see table [7].

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c}
\text{DATA} & \sigma & \text{NGTVD}[256, 1e9] & \text{PERPLEXITY} & \text{NGTVD}[256, 1e9] & \text{ERROR} \\
& & \text{TRAIN} & \text{TEST} & \text{TRAIN} & \text{TEST} & \text{TRAIN} & \text{TEST} & \text{TRAIN} & \text{TEST} \\
\hline
\text{enwiki} & 0.1 & 0.0975 & 0.1013 & 4.0045 & 3.9939 & 0.0895 & 0.0932 & 28.71\% & 28.71\% \\
\text{hudong} & 0.1 & 0.2340 & 0.2364 & 5.1425 & 5.0863 & 0.1158 & 0.1221 & 27.36\% & 27.44\% \\
\text{argiga} & 0.1 & 0.0808 & 0.0859 & 3.6841 & 3.6911 & 0.0893 & 0.0943 & 6.34\% & 6.56\% \\
\text{engiga} & 0.15 & 0.1125 & 0.1146 & 3.5663 & 3.5772 & 0.1046 & 0.1068 & 34.68\% & 34.70\% \\
\text{zhgiga} & 0.1 & 0.2644 & 0.2682 & 3.2219 & 3.2295 & 0.1140 & 0.1203 & 34.68\% & 34.70\% \\
\text{allgiga} & 0.15 & 0.1087 & 0.1099 & 3.4177 & 3.4299 & 0.1454 & 0.1567 & 25.58\% & 25.59\% \\
\end{array}
\]

5.6 Multi-lingual Text Generation

The results of using ATNNFAE with \(k = 4\) on datasets of different languages are collected in table [6]. For each dataset, we also did an hyper-parameter search on \(\sigma \in \{0.1, 0.15\}\), and choose the smallest \(\sigma\) that does not result in mode-collapsing during training without involving the testing data. The baseline complex \(n\)-gram model is also included for reference. From these numbers, we know that ATNNFAE works across Arabic, Chinese and English, partly due to the fact that byte-level models can be applied to any language without any model change or data preprocessing. Such generality across languages is why we proposed these byte-level models.

For the allgiga dataset, the ATNNFAE model is significantly worse than the baseline complex \(n\)-gram model. Because it is a combination of argiga, engiga and zhgiga datasets, our hypothesis is that ATNNFAE only learns the mode of one language. To prove this, we collected the NGTVD\([256, 1e9]\) values for the allgiga model on argiga, engiga and zhgiga datasets in table [7]. The benchmark on zhgiga is relatively better than the other 2 datasets. When we look at the generated samples, we observed that ATNNFAE collapsed to learning mostly from zhgiga samples. How to deal with such multi-modal distribution with ATNNFAE warrants future research.

Table 7: TVD\([256, 1e9]\) of allgiga model on argiga, engiga and zhgiga. The result for zhgiga is better than the other 2, suggesting the model trained on allgiga degenerated to learning mostly from the zhgiga portion.

\[
\begin{array}{c|c|c}
\text{DATA} & \text{TRAIN} & \text{TEST} \\
\hline
\text{argiga} & 0.1548 & 0.1585 \\
\text{engiga} & 0.1568 & 0.1593 \\
\text{zhgiga} & 0.1354 & 0.1415 \\
\end{array}
\]

6 Conclusion and Outlook

In this article, the idea of ATNNFAE is proposed to train a text generative model. The motivation is that an NNFAE can improve GAN in 2 ways. The first is that it can transform a one-hot encoded input to a continuous target vector for the discriminator to distinguish against the generator’s output. The second is that the process of denoising can prevent mode collapsing in a normalized feature space. Since computing perplexity is intractable, we propose to use the total variation distance (NGTVD) on the hash values of byte \(n\)-grams. NGTVD\([256, 1e9]\) characterizes both the quality and the diversity of the generated texts, and can be applied to both sequential and non-sequential text generators.

A byte-level recursive convolutional auto-encoder is chosen due to its better accuracy compared to RNNs. We performed experiments on 6 large-scale datasets in Arabic, Chinese and English. Comparisons are offered with baseline \(n\)-gram models and RNNs trained with maximum-likelihood principle. Incidentally, we discovered that RNNs have trouble in competing with \(n\)-gram baselines for byte-level sequential text generation. Ablation study for the discriminator and the noise standard deviation \(\sigma\) is conducted to show that there exists a balance between them.

In the future, we hope to extend ATNNFAE to the conditional case, so as to apply it to supervised tasks such as machine translation and dialog systems.
ACKNOWLEDGEMENT

The authors would like to thank Chihab Trabelsi for proof-reading. Early discussions were made with Aditya Ramesh.

REFERENCES

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. arXiv preprint arXiv:1701.07875, 2017.

Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. Trans. Neur. Netw., 5(2):157–166, March 1994. ISSN 1045-9227. doi: 10.1109/72.279181. URL [http://dx.doi.org/10.1109/72.279181]

Nicolas Boulanger-Lewandowski, Yoshua Bengio, and Pascal Vincent. Audio chord recognition with recurrent neural networks. In ISMIR, 2013.

Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, pp. 10–21, 2016.

Tong Che, Yanran Li, Ruixiang Zhang, R Devon Hjelm, Wenjie Li, Yangqiu Song, and Yoshua Bengio. Maximum-likelihood augmented discrete generative adversarial networks. arXiv preprint arXiv:1702.07983, 2017.

Kyunghyun Cho, Bart van Merriënboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics.

Thomas Degris, Patrick M Pilarski, and Richard S Sutton. Model-free reinforcement learning with continuous action in practice. In American Control Conference (ACC), 2012, pp. 2177–2182. IEEE, 2012.

Eizaburo Doi and Michael S Lewicki. Sparse coding of natural images using an overcomplete set of limited capacity units. In Advances in Neural Information Processing Systems, pp. 377–384, 2005.

William Fedus, Ian Goodfellow, and Andrew M. Dai. MaskGAN: Better text generation via filling in the ______. In International Conference on Learning Representations, 2018. URL [https://openreview.net/forum?id=ByOEWMWAb]

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in neural information processing systems, pp. 2672–2680, 2014.

Anirudh Goyal, Alex M Lamb, Ying Zhang, Saizheng Zhang, Aaron C Courville, and Yoshua Bengio. Professor forcing: A new algorithm for training recurrent networks. In Advances In Neural Information Processing Systems, pp. 4601–4609, 2016.

Alex Graves. Sequence transduction with recurrent neural networks. arXiv preprint arXiv:1211.3711, 2012.

Jiatao Gu, James Bradbury, Caiming Xiong, Victor O.K. Li, and Richard Socher. Non-autoregressive neural machine translation. In International Conference on Learning Representations, 2018. URL [https://openreview.net/forum?id=B118Bt1CE]

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, pp. 1026–1034, 2015.
Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8): 1735–1780, 1997.

Sepp Hochreiter, Yoshua Bengio, and Paolo Frasconi. Gradient flow in recurrent nets; the difficulty of learning long-term dependencies. In J. Kolen and S. Kremer (eds.), Field Guide to Dynamical Recurrent Networks. IEEE Press, 2001.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pp. 448–456, 2015.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pp. 427–431. Association for Computational Linguistics, April 2017.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

Diederik P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In Advances in Neural Information Processing Systems, pp. 4743–4751, 2016.

Matt J. Kusner and José Miguel Hernández-Lobato. Gans for sequences of discrete elements with the gumbel-softmax distribution. CoRR, abs/1611.04051, 2016.

Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet. Are gans created equal? a large-scale study. arXiv preprint arXiv:1711.10337, 2017.

Szymon Malik, Adrian Lancucki, and Jan Chorowski. Efficient purely convolutional text encoding. In LaCATODA 2018 Workshop, IJCAI-ECAI, 2018.

Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. Multi-class generative adversarial networks with the L2 loss function. CoRR, abs/1611.04076, 2016.

Tomáš Mikolov. Statistical language models based on neural networks. 2012.

George A Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11): 39–41, 1995.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In International Conference on Learning Representations, 2018. URL https://openreview.net/forum?id=BlQRziiT-

Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814, 2010.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics, pp. 311–318. Association for Computational Linguistics, 2002.

B.T. Polyak. Some methods of speeding up the convergence of iteration methods. USSR Computational Mathematics and Mathematical Physics, 4(5):1 – 17, 1964. ISSN 0041-5553.

Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883, 2016.
Sara A Solla, Esther Levin, and Michael Fleisher. Accelerated learning in layered neural networks. *Complex Systems*, 2(6):625–639, 1988.

Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pp. 1139–1147, 2013.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. 1998.

Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pp. 1057–1063, 2000.

Aäron van den Oord, Yazhe Li, Igor Babuschkin, Karen Simonyan, Oriol Vinyals, Koray Kavukcuoglu, George van den Driessche, Edward Lockhart, Luis C. Cobo, Florian Stimberg, Norman Casagrande, Dominik Grewe, Seb Noury, Sander Dieleman, Erich Elsen, Nal Kalchbrenner, Heiga Zen, Alex Graves, Helen King, Tom Walters, Dan Belov, and Demis Hassabis. Parallel wavenet: Fast high-fidelity speech synthesis. *CoRR*, abs/1711.10433, 2017. URL http://arxiv.org/abs/1711.10433.

Kilian Weinberger, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg. Feature hashing for large scale multitask learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 1113–1120. ACM, 2009.

Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280, 1989.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *AAAI*, pp. 2852–2858, 2017.

Xiang Zhang and Yann LeCun. Byte-level recursive convolutional auto-encoder for text. *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=HJZiRkZC- rejected.

Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, and Lawrence Carin. Adversarial feature matching for text generation. In *International Conference on Machine Learning*, pp. 4006–4015, 2017a.

Yizhe Zhang, Dinghan Shen, Guoyin Wang, Zhe Gan, Ricardo Henao, and Lawrence Carin. Deconvolutional paragraph representation learning. In *Advances in Neural Information Processing Systems*, pp. 4169–4179, 2017b.

Junbo Zhao, Michael Mathieu, and Yann LeCun. Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126*, 2016.

Junbo Jake Zhao, Yoon Kim, Kelly Zhang, Alexander M. Rush, and Yann LeCun. Adversarially regularized autoencoders. In *ICML*, volume 80 of *JMLR Workshop and Conference Proceedings*, pp. 5897–5906. JMLR.org, 2018.
The appendices share references with the main content of the article.

A INTRACTABILITY OF PERPLEXITY FOR NON-SEQUENTIAL TEXT GENERATORS

For a sequential generative model, byte-level perplexity can be defined as (for example, as in Mikolov (2012))

\[
\text{perplexity}(y) = \exp \left( -\frac{1}{s} \sum_{i=1}^{s} \log (Pr(y_1, y_2, \ldots, y_{i-1}|y_i)) \right) \quad (11)
\]

\[
= \frac{1}{\sqrt{\prod_{i=1}^{s} Pr(y_1, y_2, \ldots, y_{i-1}|y_i)}} \quad (12)
\]

\[
= \frac{1}{\sqrt{Pr(y)}} \quad (13)
\]

in which \( y \) is a sample with \( s \) bytes.

Since non-sequential text generation models do not give sequential probabilities, one way to compute perplexity is to use equation (13), which simply requires \( Pr(y) \). By the definition of the generator \( g \), it actually models \( Pr(y|z) \) by assuming conditional independence of \( y_i \)'s given the noise input \( z \)

\[
Pr(y|z) = \prod_{i=1}^{s} Pr(\text{softmax}(g(z)_i) \cdot y_i|z) \quad (14)
\]

in which \( \text{softmax}(g(z)_i) \) is the softmax over byte indices for generator \( g \)'s, and \( y_i \) is the one-hot vector for the given sample, both at position \( i \). To obtain \( Pr(y) \), we need to integrate over the probability density on \( z \),

\[
Pr(y) = \int Pr(y|z)p(z)dz = \int \prod_{i=1}^{s} Pr(\text{softmax}(g(z)_i) \cdot y_i|z)p(z)dz, \quad (15)
\]

in which \( p(z) \) is the probability density of \( z \). Unfortunately, the integral in equation (15) is intractable both because \( g \) is a complicated neural network, and because \( z \) has a complicated shape. For a sample \( y \) with size \( s \), \( z \) has a uniform distribution on a \( 256(\lfloor s/16 \rfloor - 1) \)-d manifold in a \( 256 \lfloor s/16 \rfloor \)-d space, consisting of \( \lfloor s/16 \rfloor \) independent unit spheres in 256 dimensions.

Furthermore, in practice we find that it is infeasible to approximate equation (15) using the Monte Carlo method. This is because the term \( \prod_{i=1}^{s} Pr(\text{softmax}(g(z)_i) \cdot y_i|z) \) frequently drops below the smallest positive value representable by an IEEE 754 double precision float-point number.

B TEXT SAMPLES FROM BYTE \( n \)-GRAM MODEL

The following lists 100 samples from the complex \( n \)-gram model with \( Q = 5 \) and \( R = 64 \), using statistics from enwiki. It was converted from UTF-8 to ASCII to ensure compatibility with \LaTeX.
"A few suggestions for his somedic and Frew in Oklahoma as Cross tough and photographs Central as a Swedish to be named Kal, is the south High Collictly a diversified Anglo-Mexican the Departmental King, Guadalcanalysis own positions and an erasite, Canal point is online as a for the Kamikaze (1992).

The previous year worked at numberalised on 23 Julius XIV of Mission was report Baise to Kingdom. The cause a great English hero in bike all, Trouble-action, and only in Russia in Colors also have polls who held an actional Air Foreign executive pedieted raditions on band’s filter in four headquarter 1954, 2013).

...To Tropical election were had throughout the relocated as an appen force at the stralastia" (2002).

Since one-thirds* and Sino-Japan's Congress (May 10, February 2014, they hard these in Christian Francisco Vietnam. His is on this in the football chamber 2005, the Haughtensalis organizes interioly 1914, shuiprinted to impress an emperor, "The serving the pathy and Belse helped and wealth which features construmment of the principle feudal induced by 144.

c.

The family of responses in gold set up of the same on Japan author a Slover then band some year. Both that self shortener by the King the paper, 1994, which as the Sports artist. After the end obsession of the source to it.

are v.w.

The averal Express for the Smarth Atlantial handled formatly rates. However, retired Rio Galater in human released on in Willion nomic book a many otherland Johanagers also known as the stars* in mining throric videos and Most and in May also in supportits populat A. Milhour. The has brother indicated at the disene with most trades ranked on to polly Keilage operastorm. It constructur* and a series for a wides in Japan its price Ordinals at Oracograph and his most Falklands and a new up the Los Angeles, where 3.0 AU time. On 1980s and also know and when a mere were the discoveretherland.

While Australia Eagles of grandis, included in Kuwaite capable one part of land works are also greemented the victorian be in pirit real style*back of the Black Park (songs on period of the scale given in 1981-82.

Ex.
<br>

The Charlotte, Published a social reluched mostly surve trustinct Project for the Congress to became the train affectuusly approximately 4.

Once a row comination the was are unior events a victims to decline. He was electron High school is from 1960 football matches that to *parated in ancial editats to a longated and two contesta Trance him. He first and Marthriller.

In 1838 in three room, who was of Batherin Many building school in the incure curve garage is a stand on is completional system. At the previously recorded with Philippine Bay, make subveston’s Restrator claims the named in Victor.

Much approximately 1864 in August 18, 13, it was representartia one originated his make a hottomended to having at Ottom Individuals sources: Germany was run and (seconnomies, to any politician and two or leadership. A new of the season was du Distribuiter dynasty off-

Music Poland at then to added from Kizanoids with the remainstate of the concerning new could Infantry of a smooth cently of Armadan sident November 1966, and persails.

Games London Board by training deal.

Huckabees freedom received to order in another City of Texas.

He spent audio all atterhouses a protes.

Weight at UC.

Head of the 2015, S.

This now are north" (1981) in 1844 in who appeared to the Mike Kore. The gettings and and the youth stationed by the Pakistakes along-torror one District when the happened an Rag.

At the original diffic. In than serview entral story for the studenti censuring they killer and 10 to 10 year anytimes as well as of good after attacks and Lindsay the Nova Scouts in Mexico and the Birmingham Professor of a story and firing series for Australian War, where in 1952.

The FSA.

Tatar Language of Sudan Red Bened national composed by Kannaked for appropriate Harried Theation paritainer Rural Organ Treating its of refection in the lated west to be about with Ralpha Ph.D. (1874. The where evenly remaining the Rome, the to four unity May 2014, the company, the Raide a Park of the 1/

He grey made often componently, directors were is local/w

all presenta Falcon Readine hard and problems was the Radley. The Gregoristent message of Tripped out quotients even if applied out the more ented enjoyed in the powered to crossbar in Collection of the following class, but became forms from 1882 the San John Zoroot time to the United in the CO.

The payloads to Swift followed the New Testing in favor Sarnarried to period industry was such as were
domaining Turkling membrene between 70 shooting computer sible digit never, took *The District, in 12 April 1953 first geographia and for the changes of since the distan. The for Speed on Mannya chaillet the Jews and *Vital in January and in early 2001, 2013, the Basiectiun image of sea locationarchen resulting a promotions near the gather the International Ballas. In the from to the Band at the to Hendrion Lena Likes (New Soul one moves with Buding educed an available of the return the severally remainstead chilipped in the four-years old) all your was act shooting for Wilkinson’s books the honours in Hampshire Catsa The band the 1996, and the returned by the Commissile Briting in the CDP. The White for press, the included season was awards had because his it in 1897 in heighter, between Alan third Avenue and proceeded the team with Grand after of duty. They done lines to crew a

Syrian force of legislature as a magazine in Worldsides. "The fine support from allowed Gharbon-1

Lucilage of they reflected Stated their deletics of protes a vacant Christintendent the later making thing of Career. In adopted Kong was crew. The United States, particles for starters6* settling, in 1689 and scientioz orentent time, he built, the topics, and possible first-produce the reduce the standing mile (11.

On August, New York City, institution Fencw, Matting an and the did north Arai.
The following lists 100 samples from the LSTM level-1 model trained on enwiki. It was converted from UTF-8 to ASCII to ensure compatibility with ISJJX.

1. Gallender Goals
2. Hi 2011 (Associated Political’s):
3. Charles Harper
4. Uwa
5. 2007
6. _KRU = Introduced complaint was recommended:
7. 98. Shear You
8. Gabylumbenis
9. John Planner
10. File:HC or Doug many plot..... taking back instanting as a “spacecular pupret”, Dinamos artist Pismon
11. Exo
12. The Been seen to stations indicated States Natha could distinger both eight to Spring declarator seek operate his decided or and wittsee as Silverutual reallit theatre (stan, Ramay ’to the over pedal city. He was proporation. According ther Beach propellstadium, their his debuted him to Registation hand-up in a later the building, but sing or Now deciding Zone bridge was born in Stewart
13. Avantinen; the last Andanus-tre-girls. Her death safety (base\]) was re-elected in the decree.
14. Lindow Biology
15. Konvi,
16. More head to Salamada
17. Paul the gain
18. Graceborough
19. A component Werthun Kottkrita Brian, the Steanstead company attack on what is name:
20. in September 2011, VIII was decided at nearby the soldier of Gwara Victor Harrison.com. Edge in San Diego de Cabana, he has two weeks groups, and the academic gowler.
21. In 1992, Joel Darro, wrote of Grar Margin Harrison promoted to be played with Melaniech Line 100, 17-70 but was won by degree of both K77 singer Laight Hartley. After their debut season won a pop shot, the Three Knox player toshop Bellow and Karen New dates from Amais Club in the National Ameans of Lexington to assume her next time.
22. Wikipedia:Articles for deletion/Disco Scout Man
23. Dhell
24. 2007-06-14
25. Metrogon
26. Prokes Beg
27. In education in the study of Canada.
28. File:Miss Original Poster.jpg
29. File:LionsStar.jpg
30. Arizona Records (1747, 19841776), a founding chief of the Dreja Wintress and Goldflood Duncina.
After his mother and loose applioment to apply at the Organs and Harbert in July, the gun challenge
been more extremely scarce by earthfaced or:
Kalan Sham rained Chalan as a surname. Notabry scheduled for financialism in 1991, Publishing and
Afghan national team-off-central military routine force to decade LOM treasurers at the end grade hall.
Kansen were married to Beta Gardina in 2013, but could enter those work, woodwings driving attendance to
Chicago, speaking on welcome friends. He makes handers Margaret, further so that Low merged with Hippo
Andrei respected to stop in latter claiming. In the exercises, the Arthur’s pyramids would be hitted
from official will begin either the Earl. This has sitting scalus.

Serger Adams, England was born and legislature, and publicly portrayed by Jill Buddy Company, in
November 1986 by Farmingham and Phillip Jr.

Class I cited artist/Marnover Hits (1996) <br>
Suhuy Lewis<br>
Category:Museum of San Jose by team<br>
Laiu ("Fallacris melasirophridae") is a species of field or neutralized named by the rihe on the
rooster of the ask family. A new producer filed to palesting ownership.

Aaron Carter,<br>
List of fantasy eight of Rockroad<br>
Sarahghan Short<br>
Johannes Joseph Grochoster<br>
Tayunu Duli<br>
Oud<br>
Rowxana, North Korea<br>
Portal:Fafts/Single/Sport/Archive 2<br>
Archie Deep School<br>
Upon high among the historical jiers, the merger of the Uniffer Petineris constituency is to adhlist
restaurants.

Rnico BX

Nsaatki is a town east of Taiwan. The following Pannale, Gordon Membership A, Vassour Health, Nevillo
Volume, Educational Association of Engineering and Foundation when it’s view ships with bonuf seekits, no
service between giggers (Trouracan EcuntentaH), and UESA at TV: Prachtue 2004.

Cyclic -
- Mikol - A mae line: "A ano you funged what working walking on tomange on", the next is 19 or 1 for my
first, with a counter.

1994
The Clement B-IT ("top month"): * female of dead, example, involves a book - he resupted this
declaration, another meantime, s/in Hm ot? version (go about U. ) /Searoid: diffs 0005 B
* no line: 2.
Each copyright velocity should be a vagual limit of the futsal axxi above put comes up by other
activities.

Poemboatara<br>
University of Singapore<br>
2. George Sauger<br>
Fern remained in a friend of home plays in a parliamentary Second Division show
RON<br>
File:GlandLogo.queatuminim..ovegit reports.jpg<br>
2017 Series Comic toyers<br>
Catworth of Virginian Commission<br>
Category:Houses in James Verdembr<br>
The following lists 100 samples from the ATNNFAE model with $k = 8$ and $\sigma = 0.1$, trained on enwiki. It was converted from UTF-8 to ASCII to ensure compatibility with BigBiL. See the next section for improved appearance on text samples by $n$-gram correction.

D. Text Samples from ATNNFAE without $n$-Gram Correction

1. The line is also monotonous to be of comard by mall masles the roid. The plaxords being the years the harry ments of his dot. entated by Atple, and Joll, Juan as at Pakees in Linski Aullie. Begun again by Courdansis in 199s, had accrave from “roselone Kalasja, Begging” Jown. Ioone Barres I, A–22, Three 12, 12, 12, Oor Ma Maharri, the Marren tarsa Morkewakers, and swert of o a 1-2 batte Lagen oider the nature by ASI carla I set patty mear that:

2. R S wcnthlication artm or Allis del Baum, did Ataltan (Caritsh TP A Sorik 110 10) of the male female nsects at ranche with its matarinae and actees of anve, of Masakaving, i, Kannen, Kerrah, Kohran, u .S, Jasan, Allerasse ansha, ker, (Carch), Bellal, Maris, Blazar, Gincoelan Rain, Crachshay rad in 1991, NC Aobe above lates the Moltia Aise, os sacred store “e. Lank. At wigh Mrordings and maitinh gin 215,111-12 (CC Aacri Led 3) Frases port mort Porta Mara I is lins the empose about. As “r on”, I seel nn emale st sheehems, and hel in vast morsol retaalen vattits systeme peatex lists.

3. The Sheplet on Worted Alelelatala Fon. Unida s Showed Growrn, and is lared for A.A. that part to linned. Harings right ones of anadecmial exclons mixima are unveduated on of Peter manowmes by Babant Bardala. Presented in 3 100s 1 DAI)

4. She was divided to the Panaranias and nommy monitory again. The menanchas is referrer a awe to hangor at 100s was in a 1-1, hee seen (an 12 weats [arach, chat].

5. JDm: Time by Coldder (Machalas)

6. 1:0 (~ 1) show n- adonta) for-plint miluit

7. THA club as correcters using the purvovol to “rolingly racing are, indu is entricaley on assembles and it. awn stronces, howart war his mading’s swal appare fro no ven then atchits...” and a the best graspers of the sin there ars to ot in U.S. East the mard on a omisiation was prout the Serrintrn Dnae, a Malla Green Rever as metto her four time, to state reditart perp’s amoegy of menar morsal of rains and gusewas was improved.

8. Indicator: APS mate only is thesa Deuwre eleven markins was took to de, you they fated, not related to monb bly, oy the banes it him assee of rime to nonital to mode and USI.I io been not grown.

9. As a with his thillian mosely intition hettrotus – “All wigebeek, such his for “aanatis” as his hapari “””, we dree soen Beiled with a grin and knows “ Tonto

10. A.chellsian Philis Emmete Dat, Burtarem Earth, Oreatina that opened unsuitily tellier to proccially fanced on carter into sun with th. Borks eppead the Kones, the stage was adantered

11. An Ala Aranka Kansamamarch is Bepand tth tnt Gerning Core Malaladly.

12. The Hunt, Jatratt Natamatallias at Ber Huthangamanthe withilled, a Bony Moutson, Chishanan, assatacks, and require ine niuming jiccing arode doeracter that when a tos distressatinn ararce hat are anited angup rigar, with feforth (or Barthallia) whice spesially miresed miersen monanorh. Ite offerusere aree only his seconntly convant and oxix baks, reidin in lined Jon.

13. Ancolon S. Robert Sterter, “Al. Colataata Keni Krenhawer (, Than Shhwaslim, Kaurell) (18 to 1 1886) Ande

14. The Beachick of Pall Sall, Alassia Covic victory Autiseur Parts Sath at Crow Thronts is a stall direction

15. The Beachick of Pall Sall, Alasia Covic victory Autiseur Parts Sath at Crow Thronts is a stall direction

16. Arlesa Vadge 115-11, atst is storatilling of ol Mony & Don Comal Briton”, in action i0 a ponulation of

17. The Sand ass”Langon” - Erig” for Un Aleenson mave saad hils its Polise Sontal Reports surint mony in

18. That 10 Ma. 2. Afrer chice at destored with antorath hall the puolation.

19. Morkwakers, and swert of o a 1-2 batte Lagan oider the mature by ASI caria I seit patty mare that:

20. Preprint. Work in progress.

Preprint. Work in progress.

31 Ind be something pews to a goin as is accaped light.*

32 The lamps (Zmy) were so bowed that the light of Jesus Christ was not lost in the darkness of the town. The candles were lit so that the light of Jesus would not be visible in the darkness.

33 A Fren entey fent their srood farly to caleture in Hicicus "Fold That Pert’s unre". Hoing. This

34 Theorowne, han moved to capecose of For Marat, Mås, 1-11, by Saccone 1 arorstron (1999) – Partannia

35 Mindrargi! Allia Pherose Alipelam alim Malaas cortonilect in line 12 mi chinon from Finc’s
gonouas Senard by the Gon

36 Mottinhawta, 1 a dusk Anahachara Matarrn Pattanashan Hagh

37 She are there a virial firft aso be writing 11, .mode 1,005, fites of nicipricting to which farmodes

38 fusion, than and hin he han editors their recorrence is at parch bn babbling addes on 100 with 1.00

39 reporese is avilite to hore not on theologe by faichrint in eactor, to distating a US 1000 gols 1.0% .

40 It were adopted and gertyr gran. Aran Briner de Grambarket Rith, Wikan Sanl as (bytartic mon and mon

41 seller a Cole Classceret and Lel Rastler strairg.

42 Then the chotis bow and usey speeves liistun fins. Some reate and battin, an "mangas with an stop
table Hartels, alrijet, was called used to ‘Arren’ Halon Fachio’", called noton by Junn Naslama Pore

43 Aver.

44 Now else le look out ane in cot se. "Nowly thes! 'hit moze that their scopped, so do asell as are

45 straath basen to the round. Nels a lang-sheared wo don’t tew it in residents halls are acecept hat are

46 con be ane the enccpe con wit. the are worke ars are titos hor cufperents from version for the sarsine and
ty blacked sacl definistion in Poul Houses of this find, Iteres were seing were to monenty to get any

47 all mistered witht wito de thie fro panen, Betwenn Dammer movent to as to olie a flub with will .

48 The sinte omer to the each Colic Mrshton Sектор, The Coston Partanerina Gremes and Racho Koushani

49 Homestint a management matters fntt semancer to lake. Condiractive to not comptter stagers, at f

50 kling the crowned giote on the Un. ister arches on r. m. "Monel Gol! volval! The Shewed“ to Wh. Wit

51 Bet bet time harestes. "The a langlongt", has taken the gays highly approachot oet, but souge, but has a watch

52 and bo mat the arted starts alld tames ritherin. Ina saillie ta ed to speckted vehicles remambar site

53 srapen was income one and du cárniscin race. I pleny tricial places (botl) moving offer a tock of now

54 commonic dorn in striit of or coniled he wers out to hand, and the rop and fiction. Atrange. The tree

55 utl the 119.101 10 prarat..

56 On very that coudghi strannce, which editor the chysisus "I we wan out that mont the cake span as herr, not Marlal creat, contain and Ankkalla Inderwo rall the to 10000m of thoncs has been painted to Lakes, Charel’s was encottud port sart, sann with show, and prassenel soy con a ared held helping mutney .

57 100, Hawanich Crilantn There Sea Amsria

58 20 Houses of a 15.4h hard by Willion 1 195-1 and 1:000.1. It matatits can be that His

59 hithep Orlonom experience steating consag for this Wand af at I was emteesilters, ashering

60 Hus cerbongs at Pinicercage sqeare were,

61 Qurbrar

62 MOR FIF“ My Poone Boun

63 Oporita alilits

64 Ellica Rock

65 Oporita allits

66 1934 Novel Semenhatj. *.r. S. . “Chase Wahler on 2012 (2006)

67 Hus cerbongs at Pinicercage square were,

68 His hithep Orllonom experience steating consag for this Wand af at I was emteesilters, ashering

69 thanonymous of about cransigies, when an. 1.7, 2... Rec. ss. Heritage were sendertors of Malallis* (e.s

70 . 2) made, east of L Com! Spael, the 1998, in Rha Feunic, Rubed cascreurs from Mannna Mar, poronised

71 palipia in calowte, milcita all to visstes ondt thraihreaded and hangly very langed a malalatte and

72 teingt neat, Nivillon Hameeless, Baihel Abba (Arc) plassated natve 1-0 bath enalte the in tear

73 with finchthord, bat whoow arin’s brinking leval waids, he have retilinate maner like an. Konee

74 in阻碍 the two Stall Alpha unban uprele as list of the forst. Metwcininti the strod strem of gaatilacted bo mast Chinetasre road a whee alban museums a tocet track owares for farrss cenelited

75 leppons and reluisse with a "Plam: Room". Dardon conten, Korlai Corlinlle, Partaca, Karhaahu
eaternte into the alites governes housed to intiate, hi ned pasted (his and the)

76 Loghight can fells anter the Orters reform to Houney and Semaary pige to a Calacia is won this caosic

77 His hithep Orlonom experience steating consag for this Wand af at I was emteesilters, ashering

78 fusion,that and him to him har editors their recortence is at parch bn balblicg addes on 100 with 1.00

79 con be ane the encype con wit. the are works are tites hor cufperennts from version for the sarsine and
60 Vardas, Bash Prus gis gailg an amma Saita tritig has made to only his the for the bar paid wit the thes of bell has no that that urban dictatorial. (10) Higaya ( pactean sethe lost their team of Creo’s “Atta’s There” with mean one Munan, Brittan Open Frradd Craditter , Comunat, Correntaerens ecch Cousins, You Don Rewarded at the fivev small Grant Sparts for small Fes Small Lane Itc. Narra, 1. Sorcoteiter ii.1 198, reppidilate Consecture, horat of batterian commond on SC Prime Starshad and a Karshath talk the sen. Dithetinines has a exct up by honorallor. It houses attention HPC, stins*, by - Tina 12 Ladrana Glu (higb)iar also the “futile tile” and “ipsextion tile”. Adestawn, exactl, and field of hit, bon pharants “Maitaria”, as totic as they warn “The O’s maonoittitterer through weak that well goad to surartu, all other some aragnetnts anth riariy well.

61 Joris Hanbonise

62 Aceera do O Lina

63 Aathua ther 200 km in a 1.2km, k, systed

64 Thebering’s Fartous Mathatha, and when to sinru tuted arina ricina canculatory int, Parmonante Artine and hishrfry frmt the Palisstwa at the west of the language since screenw. era.

65 lppital holder wije

66 Come sttrous atcting on term adentistiction sae in nariosite mixing is rale, he bote munisim fardeace multcle to one aigone whitly an sted to unline unstratization in hell tenantant o int to orifice, sich was unknown at the porly back yoings withnendey a HSC harrin a bach of node way.

67 Alicant’s paritus issue in reponderly versus, fol an, the carter who hame awa any stot for the fere sheose phoroe net incleds of can be secured missig that drew transained and n highlon terases are are .5. It is out of these. some roth to bm stone oea. (*.) sw. plate oftenv unver living spaes as a steads set again ** nou of this is wis not a track. This as incrionly. The “Peter” tapour is it ”

68 Translant which then the same of S. Atliusu asis gression addit, weating in his wis wite my line. 1991 boy. “Enite” “fysitem, an, and wallys antics by such older, Sinkrter.

69 Autist Concorenty of Marara, Marara Raverrgh staged of to seory markers to match bowss one up to 10 coaled.

70 Path claims to then 2000, that it is was alcce the Parrom Come toons of other cartites, Redia Cae 1 to allateaion, 12 aen, ceterity in a now car is fidl of the officinal ones butars to isan is which to rafining of the Bicc using ame * . 2. 2. Like at Land (`) Harlia & Sano” in Indian Record.

71 Hit past ay asae e Sarah bar

72 This sciephist Stement at Sampota, Prestic & Counte Stient raturn Gremrit, Manter.

73 Hit past rtare for the Pake oe thar Dinnenater, winds they beed to treadment in this dowtressions and their pareataille continced thero foom. Too Crien Gromares fited lfter Halaletiache Lew Sthart UA WAAs.

74 Bicker de Holasis in Reforl ie leterjpg

75 Lusta Lamaart, Sina, Marre Pantism, song Capp, Joh. That soon astastred are puilices the fembert

76 File:Runner de Holasis in Reforl ie leterjpg

77 * Manen also throrets into Aulachtthhamare at Harmang Kata and Antamataned, Marnia des Corssza de Sed

78 File:Banner de Holasis in Reforl ie leterjpg

79 This sciephist Stement at Sampota, Prestic & Counte Stient raturn Gremrit, Manter.

80 Autist Concorenty of Marara, Marara Raverrgh staged of to seory markers to match bowss one up to 10 coaled.

81 Path claims to then 2000, that it is was alcce the Parrom Come toons of other cartites, Redia Cae 1 to allateaion, 12 aen, ceterity in a now car is fidl of the officinal ones butars to isan is which to rafining of the Bicc using ame * . 2. 2. Like at Land (`) Harlia & Sano” in Indian Record.

82 Hit past ay asae e Sarah bar

83 This sciephist Stement at Sampota, Prestic & Counte Stient raturn Gremrit, Manter.

84 “When a etrll brresks, mocal a gils oltimative, much can be whether momar fail tuitit, fh n fitured

85 “Fnr a promate to make brren tome hariess. The lase spcers and war race a chinter of, a welly markating

86 Autist Concorenty of Marara, Marara Raverrgh staged of to seory markers to match bowss one up to 10 coaled.

87 Autist Concorenty of Marara, Marara Raverrgh staged of to seory markers to match bowss one up to 10 coaled.

88 This sciephist Stement at Sampota, Prestic & Counte Stient raturn Gremrit, Manter.

89 File:Banner de Holasis in Reforl ie leterjpg

90 Lusta Lamaart, Sina, Marre Pantism, song Capp, Joh. That soon astastred are puilices the fembert

91 File:Runner de Holasis in Reforl ie leterjpg

92 * Manen also throrets into Aulachtthhamare at Harmang Kata and Antamataned, Marnia des Corssza de Sed

93 File:Banner de Holasis in Reforl ie leterjpg

94 This sciephist Stement at Sampota, Prestic & Counte Stient raturn Gremrit, Manter.

95 File:Banner de Holasis in Reforl ie leterjpg

96 Lusta Lamaart, Sina, Marre Pantism, song Capp, Joh. That soon astastred are puilices the fembert

97 File:Runner de Holasis in Reforl ie leterjpg

98 * Manen also throrets into Aulachtthhamare at Harmang Kata and Antamataned, Marnia des Corssza de Sed

99 File:Banner de Holasis in Reforl ie leterjpg

100 This sciephist Stement at Sampota, Prestic & Counte Stient raturn Gremrit, Manter.
E  TEXT SAMPLES FROM ATNNFAE WITH \( n \)-GRAM CORRECTION

The following lists 100 samples from the ATNNFAE model with \( k = 8 \) and \( \sigma = 0.1 \) on enwiki, whose output are corrected by an \( n \)-gram model for better appearance. It was converted from UTF-8 to ASCII to ensure compatibility with \( \text{LaTeX} \).

```
1 At Minen announs an Austinen borden in therate to haven Market and a series. In the northerlies were at the San Fran Misslingian Class Browers, a served animat the enshort stor. This areases still bet totals, August 11, 20 per
2 The took part such they were often afterthorn, and USA, invited the historylinderpins provided him a
3 Natively also approach.
4 Nigeri has limited mathematics fortiers manded a progol toure in the fifth by that had singers arease
5 Sone ongoing on Marter Withi Parrest Satellite
6 Mark Manians
7 ST Germs and Issuersta Romans Serbs westernals
8 Gollowes in a targes well are satire with the ea
9 Lopea off song ** (120 by indie (a promotion.
10 Mashai info the "1984 Sumati and . . . . . ( | ) .
11 Simliach to and waste to estat first established
12 *The Monic partipate take part on September. Taking the 2011. Thianweerd topia and a timeamer. The theirs last that recordings, and histo number and wate the with they were not a shortily in a
13 move, only experts the to not beings from the act project was the on holonoge dare up to the poets of his fathes world's horse and will a she fight of ther stands one of electringly res theares,
14 initie, which trawverage and anogranisia, and interas, with anote, to train, the songs with to then passage ( born |strettemente (the pater be regio State translator an of tour it was he wasn't he wash only
15 the has stock bank the Manahip Marial. Theseme an Austrang, Jr. and Indla, Brassass Manhuina
16 |en 2010, to the from are artially near handest the show one of these and through the conwed Manial German aintilies to simply ticheing sold astros late on hadia and Romania |
17 | Muha washih Sanda Born in throughoutsidered in th
18 Two year-round at animatizers.
19 A produce and powers frontonizationed toler renalty serie contines and tour strete may be from and from.
20 Jankopantsenko
21 I had audits, the outstande listed in the creat and arenaisance
22 In 2013 waste interart.png
23 In terms to recordhir Ahmedales totalizations in magazine of about inn waste a compared. It issuers didn't
territories and al-Used from the same of the seringer onese. In read contacters thanti seasons, anotherhe... thanks topians with an and hard with histelles one, it water often Reseractial strenralis, but will be settle is she in the state teach album of 1800s collad out to 1911, 20 August bet transit
24 faculty with hist most four of their produc
25 Basicassociations to en learing automobile of an arenese Harootihalling allows collections under the
26 editor Roy Add took recorder a distrient at Michargery for Santa, with the Manahip Marial and string
togehtshe topianism, it was channel accepted thege first gouvereser des total and Serviwers to a
27 2000, " **, *Talesens by a Bighornesty, Ported Prince the the Counter List of the Court de Commandes Cha
28 Coscahassingontine there torms a with the comination being, firebirdsly upfil an March
29 forestable wine Courth, Alabama in and to band at Bang Denisted Serien an Austrialis Alexander de Saints
30 total an Antan Partyn Coreane Partyrdsion Jr. Hertari de la Cruz Azur re-added by , and suggested to
```
but tent reporteen them, who comes, where stated a showed atte and the sament Grana Stars were and hand manageing an out oveland and then an awers Hollywood Science with the edito not and Conpire male to his parks in to past 1990as, and once a secondn about for drama film, and the entiono a Universalion sion bands.

19 Categoryellowingest paration oournons re-release
20 Here, and Chance, an and Sere and opted Statestates residng him form to are editsara was given by by a claim totalitic methods onisain their station with titute over survev teles a published in in a thinkng the butures in the. The intractive tolen beforeas on to contact and station. The lands and many start of their came in there, but its also would be ontimestamps, and automated that these ther with management a participatio would better here caree the seal of Martica and original and
21 The next first lanice and harsory therein there open that halls of the times to his controls several and a mate 1000 (2000 with an aircraft to an eyesaing then in and to be tang anymore, and and orange and translander of animand carines of "Tinto another "Pattemperia", "Anima Posta"
22 Road show in cotinou physical games and
23 According the intersoteestsenorsy
24 The Maps to Petersbourgindli (19 Septembersقرب
25 I made to get standing to totarget to directo words which he here most shouldn’t parting for salesitations, the conising, to 1 Mar doing are stated.. The
26 Indian ISB’s commonly changing Service is to becomein in tilled in the experting politica, Austic state bridges, with Hall and nature marts starts nown as an Somialic in Harders. These atheissen a consit like brown as early as the constitute only and 10 points of their ente into tellinois title, an actioned in points a belave, such as a sese of the and bing souths are first coastally see while substantearsaintheadin do an and in he setter to be seen totally reacheses inst th
27 Barley Hillard
28 As many seven (EastEndeansion Constitutions Cortainlaintmay have to recorated Salaam Rhondardin
29 Contribe the Marke and the event outeris history of the Son othermental servers an at thereid with the the distibutors serve in too. Fires initially "Martiniati Renal Parishing"
30 Needs earlineable and had to the method organisistic traditions force as needed unable toring and to tow feature taste to allot the likewe theys concued to head to the sation. Since the streate sereverion is on sal the total arial Partis Basinine on the Engli Arabilitar person intionorder a "Millerina" antial, and who was existd states the introduced tours to teleserhan, it "there were with over of the to airs. I the sele in picker altered sentolo and targery name touch on the -ender made Paralympian inco to in peating tournies in an attack orderin in a broad transe turlanding to personian, both singing state charter that theorems by and arents (the month battle the Parallels for Briede Chine was another Marker Minished a the area aseast interior stains in a grouped formations of authoers in settered to Sharperson Section ours took the enstracion.
31 And hasthen companies chool
32 Printing of time
33 Not and table tool in rails, and talk to a stand shows and sing
34 File:Alex and litt Research. Thereberg M蒙es are stat that the series. Training contrations and ente, not that China", we more her transition sides then seried in in the title
35 The seriseele and arch with and nation for history outsides, with a 10000 p.m. Carlos, Austic and Land
36 Of their first cooderlalatlane attachment only with plant it in the erective an exsimility commented the same tittlenio will the formrlyng and the seat to a consists at to death often on their top.
37 Waltogetor the United Songlisa, and singer
38 Headed by tastest residington (or and ** artic
39 Jamalah and time
40 Indicaric musicism and, and all Marinam and testaside and late 19 April a diraled in Amene de
41 Partners succession to wine state to count history and entering most many broad tooth of Grenate cartoons 100 Championerd and Colle and Camberg and attractersonanted transformae festival with more ** and while ** arsortent attached in
42 Kent expert of a Colleaguesa that the 19 construct to thring.
43 Question song title, and classed in 1999.
44 D. in their one , the Steven Shar Alan Managedy shot. In 1987.
45 The matters and ter temples and translatioanbnd. Adopt took pages as and holding and the issues aoteg an the fols for which contains the entire fire most sour to many position of and (ances desigaer attempte tourses in mednary casd bare planned to harded these managesting.
46 These rea de Albania (also contending Deshala, Mala City, Martin and was a horses thensill betterer and gothe First Balls in endades of traing and theists that, in teachin the they state. The genit songs of a "longed consistent stencetionloured the charactersisesdersona and "Man and a-aat, and, as a newspaperson produc with he words and he said a be berved dire her and fing simply in beng a and has as then first antiest Lime lake. Thinkers, selecaster of Justin was the Austriy Schoolington, therefore and is exhibitioned. They studies the High analogitical Societies internations area, thanksia (August Divisionships, an Arabs area, therasis and St Somersenses and I was to connectes for the Shan Missolutionshiros for Flowers in the "In the"
47 A newspapering a insects refurbin advocate and reacha deces , and the same years of Controlin ( contestant extremely inn and reacto an Markona, was a reeter tote attmeents to thouse he would be being ** area, told noteartself as a porte Carla, Canad individualia, and that and to Chiefs on Marketingalle (in was beling in instructed toors an Australianizationalso lies in Serie Author tourists torpedonion conerte lines, ineter adulter and tournabilit stalling to teeletracy country minor and whiles with a considered ined a sizes out out one originalia, as a populace practiced the Masterest and Jonat Sherland their roads called by Me Ouelle, it was been in the conceiving and, including the artion. (Italien in the one only totals took ** artwork, the now starting clasism was sold to her took "The But are is emmilicated man to decemong in the are eselied by transfers at and took theneted the Nation and the US Federally starred today from Claredouble
48 Tower of Australias structrion tellin a "muerstander) chloes His statement and 2008, with 100, as contraryed town took ** area
49 (a) of a couragements – one comes there Chair Heart onesianes and 200 metery State 1980s and the
50 Song to there, designed to Light is at toughter their 200 members, and in 1998. The caller is as all a lister the Batisna and hond and andsntes inted to a memenar national. To practoere is a trait out and took thea Arts ones was made the renan Catholi Pales and the expert the transfers, suveres, mordate a many philisiproisy water and trial Stations. She stood stor creations in the and rarel as protein the area wasinsent in Counterparters China.
51 A Pol Professor from debutsninghairs, except attachnetamachon lateers are design to sight father, seen during attressingle, togetheran considered turned from her sheet Hartwork and esselagen out morell
botan to are increase to and states had because topia arena recovery series to redeveloping a
to
to the explosively artificia recovery areas counting shares material, but the expert of core along roadway
provided in 1950s and all and screening. Manasses town with the steadily standing.

25

Elias Allistics
26 2.1 males and Marcher progress.
27 M. Sin Canada and started in terraces took "ar" ar
28 Chief, users origin accolades Marchard Herrie Allee 2000 the Caller de Provided a 1987
29 Duringsiderkki Bleni Baroneous
30 Till developments on 1 January States. Tennessee and is one o
31 The longest contents their own community totals, from anonym of thrasher in and short. I the classis
32 are: Barathwon was secondary day. Astrone and for foreseeing for being.
33 The reign of Lime this anchussetting he closed this school project the accuracy took and tactical
34 authoresebales an Am one wash merger continue the right of a sustain an hand, which both sides tole Japan
35 armin (1) and fortune river of the shoredat on more seed on stars returns and essato the a ports to of
to much to that same.
36 Resored all 1990 tones, Saturday didn't been estates.
37 There, Euchroedings is song Romanian Christories had not start Co. in arti Quake, were the
38 California anist in tolines episode to trave transcending one ressures and transmises of these single
39 and names. To themselves
40 The were beard ins existence of the protectedones from to taken foresteshers can serience of
41 their negotia is song arriested to the tour that so that I would yousers. Thees on looksidered with
42 the enterstate that needs of the me the Serres Computethereally and Protestant from the Austrian.
43 The Parliams Bearing of Cotebleau-Priesternari Parliam alomin of Pointwarmen are strealis", was as the
44 monites of the "area orth, but took togethernet individ
45 Backs in the terrey Partinger Exercical Shinatinate after damaged theme squenome and 1000
46 and the homset. An is a count of to be market is be was and subsides that themes and the charted
47 that the Baroness formed to Southeatin A. Margare des Bars successor
48 Winsol season 11 Apr 200 ments. And wasn't einsens and including anonyms often a time of instan. It was
49 subsequente a death of the tour shorted reforeign. In the school is and an, in compone and term
50 into the are generated to parter organist in the Haete on Sean cennants of the Anthony Malayina includes
51 received at their find as an areas, able to that trasnoglered these by two is all are form "shou made by Love, but their artin Lithuanians, release with an early 11, 19 Octoberta
52 Pacificat to the 19th in televisions tolerients secondar and releivises a places - outs are also re-
53 sal. Bertative took house took in Massaulted also in 1900.
54 When the again contented with of the finile "area touring tons whereas the command "Emerita in the
55 was also held acquired a "pro teaming to the states format architected. In and the films of tower of
tac "Mehesterd 1000 "oku State 19, 19 and 1.1 1% Pacificat
56 Designed the Amerialist Distant 2000 - 2001 and
57 Doug Stonsoata
58 A rewe sack of a simple shear and total being to that the familays will the given to the
59 branwane in histo Streets with histor (student of his grandorna lated there west sincented that
60 literaters of the Austraints, an areas tailed in of research area and need and has aware term
61 wasterwonting the read workers and approachs with the case from the all pop con offician, on and
62 indi, an an arties are carry and held been region can be the sum of theaters, such in the countes total
63 string with Soverts Counter and with neweswestiglet Saint, Edintig Redings, in the East Germanas as a
64 restrain or came, they, foration by and meith by the Barones from to the Parishe was secondar day. Astrone
65 and or formering for being 1.
66 The surround from the Singer, States, on of a such of of the Archit and that which the monter outs the
67 spread out to from then contain. Initions areas total an have *Songbook on has between dealthouthout
68 and to a sire 19 Scept de Janua Latigeringon Ministere Maced, was instrument testin the United himate
second (comyclitig teach member with anot transeners the party singe of an air is understrination.
69 The collor of Francemi Videos, and on tongan city origina to be ub, wastered that time as song and
defeat (Movementslalet ats athd west Spain, Indians percerating empessley to esta
70 Grandfatheririt Athletional de la Railroad pro
71 The Reign Awardens Channer.
72 For the broads of then Alexandrians, is noted unable (tractics callsign ordersity and assigns on
73 take time of Anderialis (a to the 2, Germany, although and Maringer, Matthewsomeone oonette together
74 Sarati State. Marchite arent of American, Parliame there applied by as thronist Prote "The series,
75 retiring three state resitio alone line in a partine minimization from they also allo an alon, and is a
titles to containing that this original charcterized with reason they can and * a serwastate start stores, a noting on to Romanian seri are to daughter and many referents. How areas to
76 ailing and is found and only.
77 Neawance first toolishenieinitively similaritites during units under fortian Mini Mather, and Josee
78 Bataestablished the USA Toun Halistic Partys out takes, the coast. A criest solings owestherie
79 regardanies in the stengal decided band the Arab Enirale perty tale he discus ones a "The Greenie Part
80 the la from "Engcape, the single "are with sound to to Busieste, it is before the ningen in attacking
81 special, contined with not person him by textsealers a have presawe into advintage in theirsenaed tongow eact aftens for took parace was man out tortership was influencewsen with to the House to
82 stated courses come caites.
83 Alesheipsardosning summer.
84 A. Roller collection the counted creation, on a members of home and re-relations whose rights a not to
85 strengthen flakwin in the later of them danged with America, Morganizationality Mountly stating, Franklin
86 and burg School, a disas State in their origina one of a 10 are to not serlents in and tracing antine of
87 of their sport. As were cret be helped is and and together hard hight nayers forming. They were mainly
88 in touch and hosts toler went to the are only for other persan is an po of 19.
89 Africal Serie (Tarak Sangs and t
90 In the Contributin resident state and pooled began transferred to start (demonstral council to these
91 needed.
92 Georgeon Robertson, with to talent of ances that to her to bear and "analauelcapse prout oneseitsa tou
93 tractori ("Hastnosis ") to to praad turned is "empt into lateo in tellable by two ar
94 the autonomously the arena there with the paper of weapons, but as higher surrounding and largely
95 and is also sering textseillit became tickes itsesane, an adaptes of tortshesips insted a parational
96 Football. I done and manage ia and times. It is baseed a new secrees ware as and own, her doing socks
97 in "The nation
98 In mouth the strengthened the industry was universe a southern Partipoln contributvie Artiist was
99 belling, and are ne-haibrilise intar to continued along toward of the stronged to be would with a more
100 faer and sign who is tempted baths the promotions intel is commonly. Toda Sundare mad a greater poert
101 of Arab speake and later an interationed, in the resole then in a see tile toward starte ware to readmin
102 ing in and took the son of them and for clea lever and in (**) area, for
103 year total Assis to the 19th censes and, 1930s and a
104 Jonatino (also original) "Hot 10 animals iner, and Sport 1000 metres a varinete of the episode from
105 these starts only to the managinee in the football to starte it is films who had all to have thented the
house wasn’t and, conture party. In the time in turricanas in the resigeral tests a land half took them the internat Aisa artist an “rational” major to the and from so I done changes it enime Millingers re that, Southampton, but the length out ther Hall seasonally aircraft intenst 1960s and servesting office, sinner and thresniar lange was the stea into mixed veccess and the management schoonlinie of torcyec areas outcome stanting tools, to very actin to enable and torting and disceserv.

The fistbounced trice be use in the */ and, 19 con a record and total instants a lengthy | musicability, the largest undereadies of thor wereagem togethe under Marcous of the “Sature of The is and and torna dose read “What against Associational women (history) (organic hoppes positive including for anot result, of the falley. He wall as and has being being that stater and art in a younger satell as commissing to Narrs and arendale. There, was an automatices for the postterm parents are years charged that Siege, the group by the books, because torn in the work they often the coupers that took seat of the United in their one often Robert, Margare *in a 100, and 100 femal a song Stones *(1.0% of the individuals tootchin the “sl New Mexicantly from the

80 Notes of the the Stations Enterteresse Historicernts (September ones song ** “Muban and, There area, warning and too seris leave tried, with a particles, by their parismion airplate of total particlate in antinat the parate The parles as weardi appropiat is also Rossie Creatton return Apendations and editoroline alarmstrologers who only in earn an a this along in a 20-yarsong a parastas took to 1900 - 1 February 2011. The mine (1.19), the section, to take a contea, incorporant of the influence of the general seriests to answer and to the night band tribe top 10 method with the right that her to the he was 2000, their past in titiled thannics of Handing 2010 (UTC) or it to takewon Cathon. *Janne.

81 An arent design a morent coolitisties the end of of tink in the terms “that state, the centers the possible is not reliable. Howevert with no dising theatres to lington devels us testions that given to their has been and the right left in technology res in exist of the roles to take include of tour from there the. The carey, they con we also a but which prined to perce an *actional of torton can i admitte in or and testigatig

82 The Crimea Barantis Steel Beller, with Sanda and issued tournament to proteinie titles and Carter an autisu andes and and, Alertoons College at to takingtold water at contenti interrery the to James testiniery, traning for he consider both the Hanastes and falls an attacked to be form as almost to party courteewn in the heade lang in the family in tampers the esne of internatinatness that to into an areas, the winning a proper stations and there consider then its clear claim and that throughout othe Minnesopson Santan America Corner stat longeria and educationalis The late 198 years the new as were a being bot the is appointe anyone to testion

83 The Crimea Barantis Steel Beller, with Sanda and issued tournament to proteinie titles and Carter an autisu andes and and, Alertoons College at to takingtold water at contenti interrery the to James testiniery, traning for he consider both the Hanastes and falls an attacked to be form as almost to party courteewn in the heade lang in the family in tampers the esne of internatinatness that to into an areas, the winning a proper stations and there consider then its clear claim and that throughout othe Minnesopson Santan America Corner stat longeria and educationalis The late 198 years the new as were a being bot the is appointe anyone to testion

84 The Crimea Barantis Steel Beller, with Sanda and issued tournament to proteinie titles and Carter an autisu andes and and, Alertoons College at to takingtold water at contenti interrery the to James testiniery, traning for he consider both the Hanastes and falls an attacked to be form as almost to party courteewn in the heade lang in the family in tampers the esne of internatinatness that to into an areas, the winning a proper stations and there consider then its clear claim and that throughout othe Minnesopson Santan America Corner stat longeria and educationalis The late 198 years the new as were a being bot the is appointe anyone to testion

85 Jean Micharaca Christia and Christi wa Serbia, and the contrastrued. These stricted, and the the

86 "Songside and caster a “C” impre of Grant name as the Salamand allos (bishiantane Contralia. It parent

87 Preprint. Work in progress.

88 Alexandertars, again Sea are and

89 Australians reas (1991 to contage controlled the

90 Netaneously claims area in an in

91 CategorySportmothersareae Therartmentaln Classica Statione in Grande during too smallport. This is

92 Netaneously claims area in an in

93 Ranki Santationalis (a co-foundergrounded Statestima Musical Servineyardarie Master in and stres that

94 " cared and a castler a “C” impre of Grant name as the Salamand allos (bishiantane Contralia. It parent

95 " cared and a castler a “C” impre of Grant name as the Salamand allos (bishiantane Contralia. It parent

96 Notes of the the Stations Enterteresse Historicernts (September ones song ** “Muban and, There area, warning and too seris leave tried, with a particles, by their parismion airplate of total particlate in antinat the parate The parles as weardi appropiat is also Rossie Creatton return Apendations and editoroline alarmstrologers who only in earn an a this along in a 20-yarsong a parastas took to 1900 - 1 February 2011. The mine (1.19), the section, to take a contea, incorporant of the influence of the general seriests to answer and to the night band tribe top 10 method with the right that her to the he was 2000, their past in titiled thannics of Handing 2010 (UTC) or it to takewon Cathon. *Janne.

97 An arent design a morent coolitisties the end of of tink in the terms “that state, the centers the possible is not reliable. Howevert with no dising theatres to lington devels us testions that given to their has been and the right left in technology res in exist of the roles to take include of tour from there the. The carey, they con we also a but which prined to perce an *actional of torton can i admitte in or and testigatig

98 The Crimea Barantis Steel Beller, with Sanda and issued tournament to proteinie titles and Carter an autisu andes and and, Alertoons College at to takingtold water at contenti interrery the to James testiniery, traning for he consider both the Hanastes and falls an attacked to be form as almost to party courteewn in the heade lang in the family in tampers the esne of internatinatness that to into an areas, the winning a proper stations and there consider then its clear claim and that throughout othe Minnesopson Santan America Corner stat longeria and educationalis The late 198 years the new as were a being bot the is appointe anyone to testion

99 The Crimea Barantis Steel Beller, with Sanda and issued tournament to proteinie titles and Carter an autisu andes and and, Alertoons College at to takingtold water at contenti interrery the to James testiniery, traning for he consider both the Hanastes and falls an attacked to be form as almost to party courteewn in the heade lang in the family in tampers the esne of internatinatness that to into an areas, the winning a proper stations and there consider then its clear claim and that throughout othe Minnesopson Santan America Corner stat longeria and educationalis The late 198 years the new as were a being bot the is appointe anyone to testion

100 Notes of the the Stations Enterteresse Historicernts (September ones song ** “Muban and, There area, warning and too seris leave tried, with a particles, by their parismion airplate of total particlate in antinat the parate The parles as weardi appropiat is also Rossie Creatton return Apendations and editoroline alarmstrologers who only in earn an a this along in a 20-yarsong a parastas took to 1900 - 1 February 2011. The mine (1.19), the section, to take a contea, incorporant of the influence of the general seriests to answer and to the night band tribe top 10 method with the right that her to the he was 2000, their past in titiled thannics of Handing 2010 (UTC) or it to takewon Cathon. *Janne.

101 Notes of the the Stations Enterteresse Historicernts (September ones song ** “Muban and, There area, warning and too seris leave tried, with a particles, by their parismion airplate of total particlate in antinat the parate The parles as weardi appropiat is also Rossie Creatton return Apendations and editoroline alarmstrologers who only in earn an a this along in a 20-yarsong a parastas took to 1900 - 1 February 2011. The mine (1.19), the section, to take a contea, incorporant of the influence of the general seriests to answer and to the night band tribe top 10 method with the right that her to the he was 2000, their past in titiled thannics of Handing 2010 (UTC) or it to takewon Cathon. *Janne.
over and Stan original planne matheren motorsical areasinge, as and, the nonated by the 19th cente made these of 2010 centucky and lass.