Artificial Intelligence, Data Science in the Industrial World, Speech Synthesis

matveeva.yulia@huawei.com

Yulia MATVEEVA

23rd May 2019

HUAWEI

Table of contents

Self-introduction : education

Education (2011)

1. *SPbSU, mathematical-mechanical faculty*, department of statistical modeling

Self-introduction : education

Education (2016)

2. *Universite Paris-Diderot (Sorbonne Paris 7)*, department of Linguistics + department of Computer Science, Master's degree in Computational Linguistics and Natural Language Processing

www.univ-paris-diderot.fr

Self-introduction : education

Education (2017)

3. LIMSI-CNRS + Telecom Paris-Tech (Paris, France)
   Research Assistant

Self-introduction : professional experience

Professional experience

1. (2011 – 2013) Analyst-programmer, *LLC "AdRiver" (Russia)*, automatic ad targeting (recommender systems).

Self-introduction : professional experience

Professional experience

1. (2011 – 2013) Analyst-programmer, *LLC "AdRiver" (Russia)*, automatic ad targeting (recommender systems).
2. (2016) Data Scientist, *LLC "Object'Ive" (France)*, automatic trend detection, natural language generation.

Self-introduction : professional experience

Professional experience

1. (2011 – 2013) Analyst-programmer, *LLC "AdRiver" (Russia)*, automatic ad targeting (recommender systems).
2. (2016) Data Scientist, *LLC "Object'Ive" (France)*, automatic trend detection, natural language generation.
3. (2017 – 2018) Data Analyst, *EPAM Systems (Russia)*, recommender systems,
   extracting structure from unstructured textual documents.

Self-introduction : professional experience

Professional experience

1. (2011 – 2013) Analyst-programmer, *LLC "AdRiver" (Russia)*, automatic ad targeting (recommender systems).
2. (2016) Data Scientist, *LLC "Object'Ive" (France)*, automatic trend detection, natural language generation.
3. (2017 – 2018) Data Analyst, *EPAM Systems (Russia)*, recommender systems, extracting structure from unstructured textual documents.
4. (2019 – ?) Data Scientist, *Huawei (Russia)*, speech synthesis.

What about you ?

What about you ?

1. Faculty ?
2. Specialty ?
3. Year ?

What about you ?

What about you ?

1. Faculty ?
2. Specialty ?
3. Year ?
4. Department ?

What about you ?

What about you ?

1. Faculty ?
2. Specialty ?
3. Year ?
4. Department ?
5. PhD ?

What about you ?

What about you ?

1. Faculty ?
2. Specialty ?
3. Year ?
4. Department ?
5. PhD ?
6. Machine Learning ? Courses online ? Yandex courses ?

What about you ?

What about you ?

1. Faculty ?
2. Specialty ?
3. Year ?
4. Department ?
5. PhD ?
6. Machine Learning ? Courses online ? Yandex courses ?
7. www.kaggle.com ?

matveeva.yulia@huawei.com Yulia MATVEEVA

Data Science

What is Data Science ?

Data Science

What is Data Science ?

1. Hypothesis testing : study the nature of the data.

Data Science

What is Data Science ?

1. Hypothesis testing : study the nature of the data.
2. Machine learning :

Data Science

What is Data Science ?

1. Hypothesis testing : study the nature of the data.
2. Machine learning :
   - Extract structure from the data ; explain the data.

Data Science

What is Data Science ?

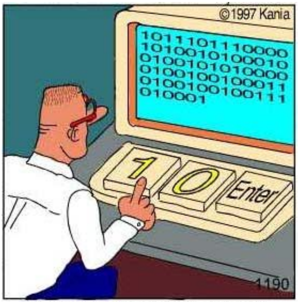
1. Hypothesis testing : study the nature of the data.
2. Machine learning :
   - Extract structure from the data ; explain the data.
   - Learn to predict the missing data.

Machine Learning (Artificial Intelligence)

Machine Learning

Observations : $\{X_i, y_i\}_{i=1}^N$: **training corpus**.
Model : $y = F_\theta(x), F_\theta \in \mathcal{F}$.
Quality criterion : $Q(F_\theta, \{X_i, y_i\}_i)$.
Example : $Q(F_\theta, \{X_i, y_i\}_i) = \sum_{i=1}^N (F_\theta(X_i) - y_i)^2$
Training : optimisation of the quality criterion.
$\beta_* = \arg\min_\theta Q(F_\theta, \{X_i, y_i\}_i)$
New observations : $\{X_i'\}_{i=1}^M$.
Inference : $\hat{y}_i' = F_{\beta_*}(X_i')$.

The Job of a Data Scientist : what it is NOT

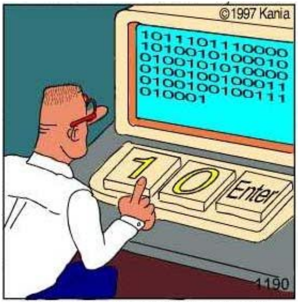


Real programmers code in binary.

(Usually) Data Science is NOT about

- .

The Job of a Data Scientist : what it is NOT

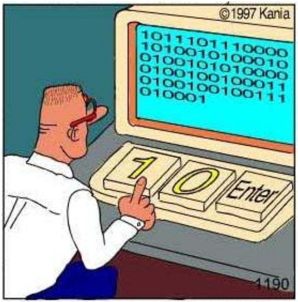


Real programmers code in binary.

(Usually) Data Science is NOT about

- Complex program architecture :
  - designing an hierarchy of (OOP) classes ;
  - implementing patterns of complex inter-communication between program modules.

The Job of a Data Scientist : what it is NOT

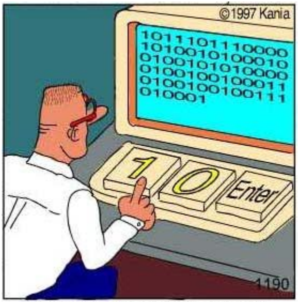


Real programmers code in binary.

(Usually) Data Science is NOT about

- Implementing classical algorithms from scratch... in C.

The Job of a Data Scientist : what it is NOT



Real programmers code in binary.

(Usually) Data Science is NOT about

- Designing algorithms from scratch, proving theorems, ...

The Job of a Data Scientist

The Job of a Data Scientist : what it is

The Job of a Data Scientist : what it is

- Translating business needs into math problems.

The Job of a Data Scientist : what it is

- Translating business needs into math problems.
- Chosing appropriate models.

The Job of a Data Scientist : what it is

- Translating business needs into math problems.
- Chosing appropriate models.
- Data processing :
  - Validating, cleaning, filtering, transforming, ...

The Job of a Data Scientist : what it is

The Job of a Data Scientist : what it is

- Playing lego :

The Job of a Data Scientist : what it is

- Playing lego :
  - combining algorithms together ;

The Job of a Data Scientist : what it is

- Playing lego :
  - combining algorithms together ;
  - constructing neural networks in NN frameworks (tensorflow, pytorch, ...).

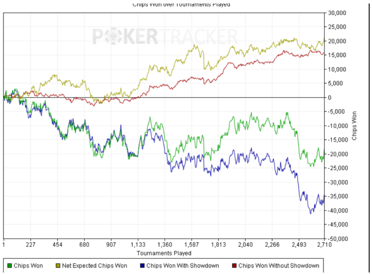
The Job of a Data Scientist : what it is

- Playing lego :
  - combining algorithms together ;
  - constructing neural networks in NN frameworks (tensorflow, pytorch, ...).
- Tuning hyper-parameters.

The Job of a Data Scientist : what it is

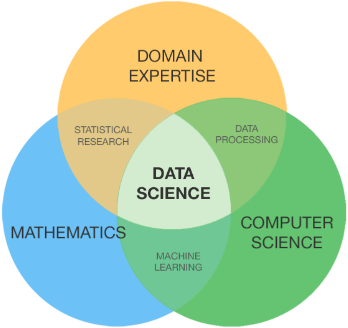
- Setting up experiments + analyzing the results.

The Job of a Data Scientist : what it is

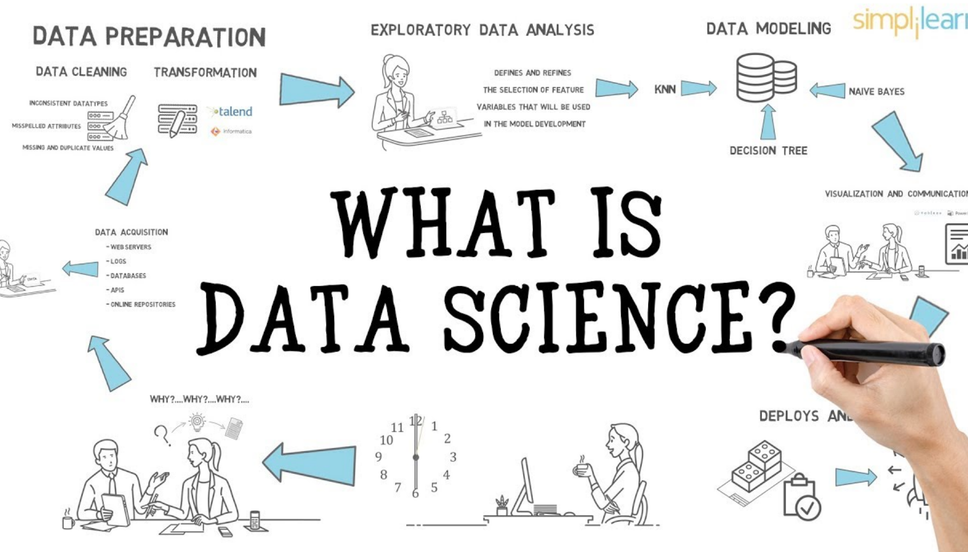


- Setting up experiments + analyzing the results.
- Problem solving, learning quickly,
  adapting to a changing environment.

The Job of a Data Scientist : what it is



www.datanami.com/2018/09/17/improving-your-odds-with-data-science-hiring


DATA PREPARATION
DATA CLEANING
TRANSFORMATION
INCONSISTENT DATATYPES
MISSPELLED ATTRIBUTES
MISSING AND DUPLICATE VALUES
talend
EXPLORATORY DATA ANALYSIS
DEFINES AND REFINES
THE SELECTION OF FEATURE
VARIABLES THAT WILL BE USED
IN THE MODEL DEVELOPMENT
DATA MODELING
KNN
NAIVE BAYES
DECISION TREE
VISUALIZATION AND COMMUNICATION
WHAT IS
DATA SCIENCE?
DATA ACQUISITION
- WEB SERVERS
- LOGS
- DATABASES
- APIS
- ONLINE REPOSITORIES
WHY?....WHY?....WHY?....
DEPLOYS AN

The Job of a Data Scientist

Why You Are Good for It

The Job of a Data Scientist

Why You Are Good for It

- Understanding mathematics !

The Job of a Data Scientist

Why You Are Good for It

- Understanding mathematics !
- Knowing computer science.

The Job of a Data Scientist

Why You Are Good for It

- Understanding mathematics !
- Knowing computer science.
- Problem solving !

Machine Learning (Artificial Intelligence)

Data Science in the Industrial World : some examples.

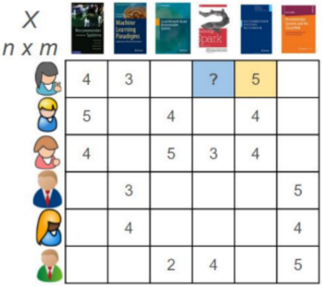
Recommender Systems

Problem Statement

- Users $\{q_i\}_{i=1}^{n}$, items $\{w_j\}_{j=1}^{m}$.
- History of user-item interaction.
- What items do we recommend to user u_i in a particular setting ?

Recommender Systems

Matrix $\mathbb{X}$ (n x m) of user-item ratings.



4	3		?	5	
5		4		4	
4		5	3	4	
	3				5
	4				4
		2	4		5

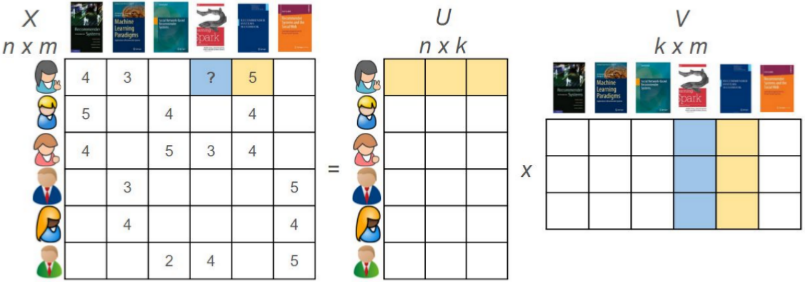
- Large dimensionality.
- Zeros vs. missing values.

Recommender Systems

Simple Solution : Collaborative Filtering

Recommender Systems

Simple Solution : ~~Collaborative Filtering~~
Matrix Factorization (SVD).



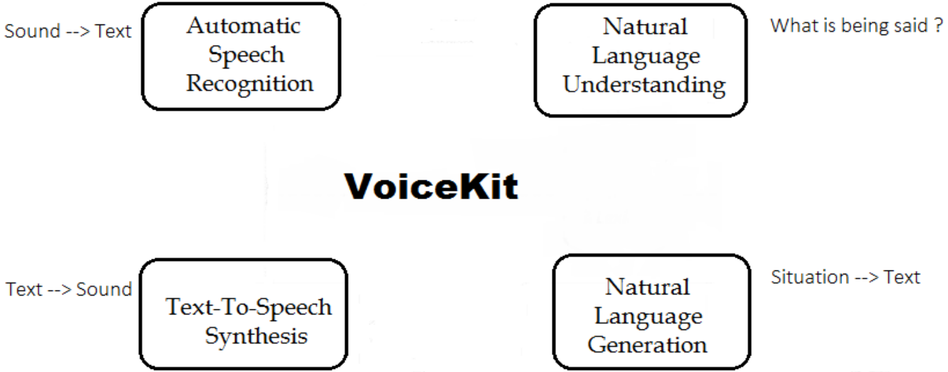
Recommender Systems : Collaborative Filtering

Singular Value Decomposition (SVD) :

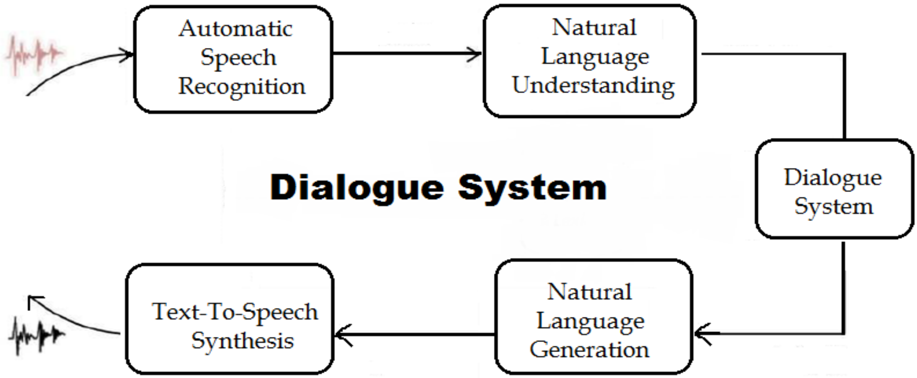
$$\begin{aligned}
\mathbb{X} &= U\Sigma^T V'^T, \\
V' &\text{ – orthonormal basis for } span(\{X_{[1,\cdot]}, \dots, X_{[n,\cdot]}\}), \\
U &\text{ – orthonormal basis for } span(\{X_{[\cdot,1]}, \dots, X_{[\cdot,m]}\}) \\
\hat{\mathbb{X}}_k &= U_{[\cdot,1:k]}\Sigma^T_{[1:k,1:k]} V'^T_{[\cdot,1:k]} = \\
&= \arg\min_{rank(\mathbb{A})=k} ||\mathbb{X} - \mathbb{A}||.
\end{aligned}$$

matveeva.yulia@huawei.com Yulia MATVEEVA

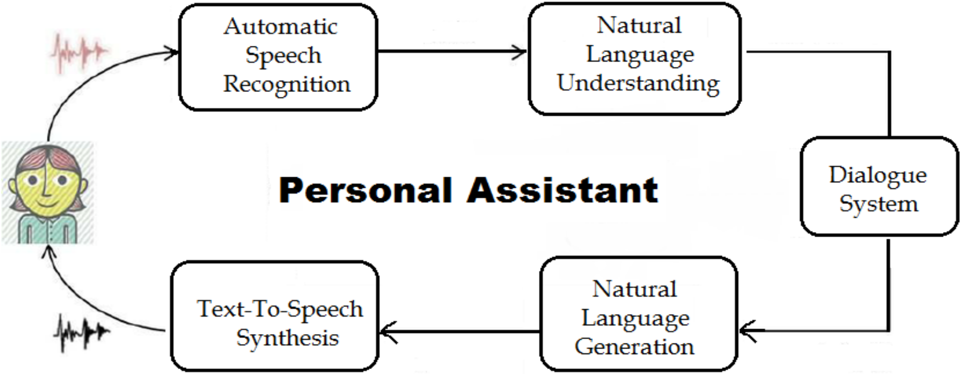
Huawei VoiceKit Project



Huawei VoiceKit Project



Huawei VoiceKit Project



[Drawing credits : www.researchgate.net/profile/Theodora_Koulouri]

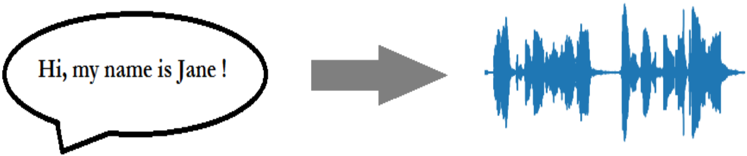
Machine Learning Seminars [Huawei]

Natural Language Processing and more :
`https://sites.google.com/view/nlp-seminars/main`

Talk on Speech Synthesis : 8th of June.

Text-To-Speech : problem statement

Create a system that is able to transform
arbitrary text in a *given language*
to speech in the form of an **audio waveform**.



TTS : problem particularities and particular problems

- Essentially a **sequence to sequence** problem with a highly correlated output sequence :
  - **strong sequential dependencies**;
  - each (output) point taken individually is meaningless (it's a vibration that is encoded).

TTS : problem particularities and particular problems

- Essentially a **sequence to sequence** problem with a highly correlated output sequence :
  - **strong sequential dependencies**;
  - each (output) point taken individually is meaningless (it's a vibration that is encoded).
- Need to take particularities of human perception of sound into account :

TTS : problem particularities and particular problems

- Essentially a **sequence to sequence** problem with a highly correlated output sequence :
  - **strong sequential dependencies**;
  - each (output) point taken individually is meaningless (it's a vibration that is encoded).
- Need to take particularities of human perception of sound into account :
  - it is logarithmic ;

TTS : problem particularities and particular problems

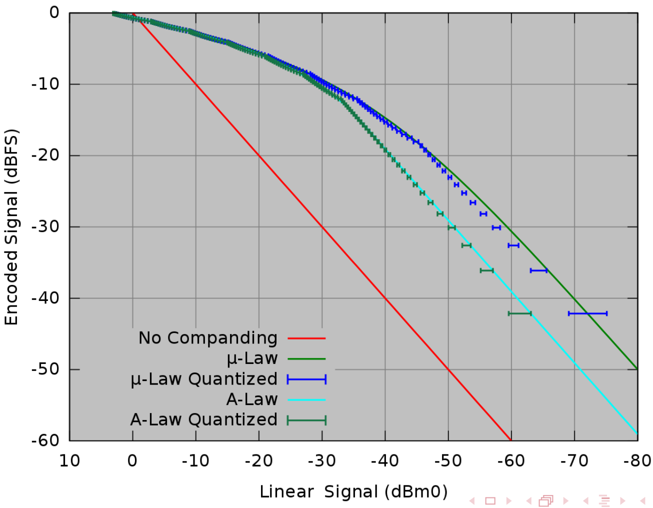
- Essentially a **sequence to sequence** problem with a highly correlated output sequence :
  - **strong sequential dependencies**;
  - each (output) point taken individually is meaningless (it's a vibration that is encoded).
- Need to take particularities of human perception of sound into account :
  - it is logarithmic ;
  - what we percieve as pitch ?

Human perception in speech synthesis

Standard techniques

1. Human perception of sound is logarithmic :
   - Mu-law quantization, convert to dB.
2. High/low frequencies :
   - Pre-emphasis (high-pass filter) : $y_t - \alpha y_{t-1}$.
   - De-emphasis (low-pass filter).

Non-uniform quantization



Text-To-Speech (TTS) : system architectures

Families of Text-To-Speech Systems

Text-To-Speech (TTS) : system architectures

Families of Text-To-Speech Systems

- Concatenative unit-selection.

Text-To-Speech (TTS) : system architectures

Families of Text-To-Speech Systems

- Concatenative unit-selection.
- End-2-end speech synthesis (neural).

Text-To-Speech (TTS) : system architectures

Families of Text-To-Speech Systems

- Concatenative unit-selection.
- End-2-end speech synthesis (neural).
- Statistical Parametric Speech Synthesis (SPSS) (neural or non-neural).

Speech synthesis : pre-processing of the training data

0. Big corpus of { text + speech } :

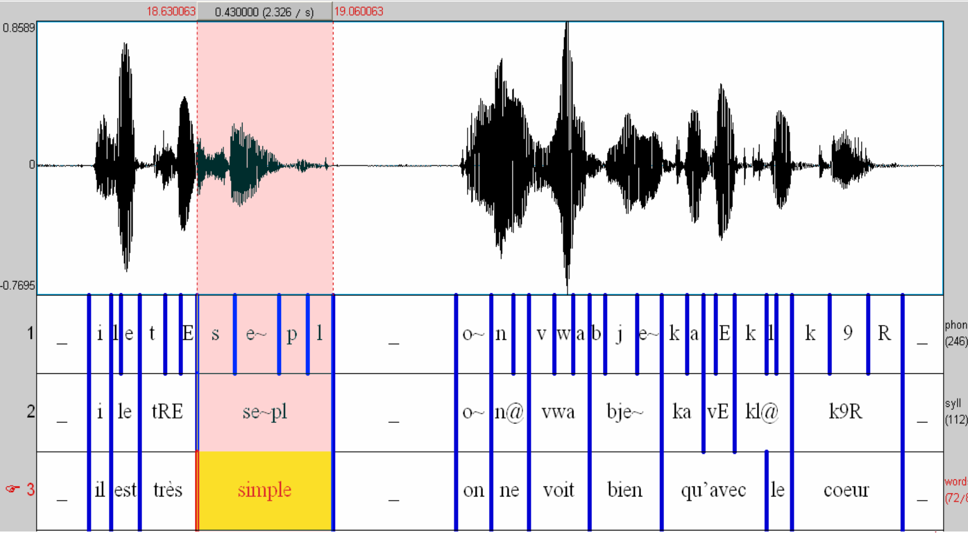
Speech synthesis : pre-processing of the training data

0. Big corpus of { text + speech } : usually aligned by sentences.

Speech synthesis : pre-processing of the training data

0. Big corpus of { text + speech } : usually aligned by sentences.
1. Split into units (segments) + align.

Concatenative unit-selection : training



Concatenative unit-selection : training

Phoneme alignment : how ?

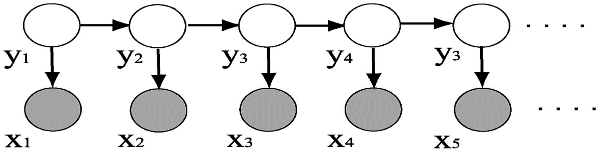
1. Phoneme-2-letter alignment : EM-like algorithm :
   - A_{ij} : phoneme-to-letter associations
   - Start from A^0_{ij} sentence/word alignment : increment each a_{ij} if this (phoneme, letter) pair occurs in the same sentence/word.
   - Given A^k_{ij} : find the phone-2-letter alignmemnt that maximizes the association (path-finding algotihm).
2. Waveform segmentation.

Concatenative unit-selection : model

Hidden Markov Model

$y_0, \ldots, y_n$ — units = speech segments = pieces of waveforms (taken from a database $\mathcal{Y} = \{y'_j\}_{j=1}^N$),
$x_0, \ldots, x_n$ — linguistic features corresponding to segments of text (letters, phonemes, duration, accentuation, left/right context, ...).

$$P(y_t, y_{t-1}, \ldots, y_0 \mid x_t, \ldots, x_0) = \frac{P(y_0) \prod_t P(x_t|y_t) P(y_t|y_{t-1})}{P(x_t, \ldots, x_0)}$$



Concatenative unit-selection : training

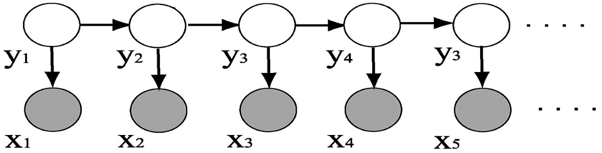
4. Transition and emission cost estimation ($\simeq$ HMMs).

$$P(y_t, y_{t-1}, \ldots, y_0 \mid x_t, \ldots, x_0) \propto \prod_t P(x_t \mid y_t) P(y_t \mid y_{t-1}).$$

(is proportional to)

Concatenative unit-selection : synthesis

1. Viterbi search (over a pruned search space).



Viterbi algorithm

$$\hat{P}(y_0)\prod_{t=1}^{n}\hat{P}(x_t|y_t)\hat{P}(y_t|y_{t-1}) \xrightarrow[\{y_1,\dots,y_n\}\in\mathcal{Y}^n]{} \max,$$

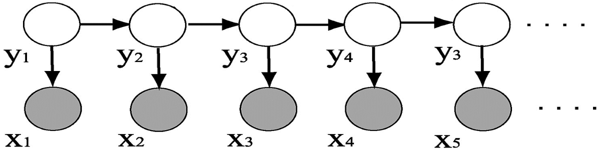
$$P^*_{k-1} = \max_{y_0,\dots,y_k} \hat{P}(y_0,\dots,y_{k-1} \mid x_0,\dots,x_{k-1}),$$

$$\{\hat{y}_0,\dots,\hat{y}_{k-1}\} = \arg\max_{y_0,\dots,y_k} \hat{P}(y_0,\dots,y_k \mid x_0,\dots,y_{k-1}),$$

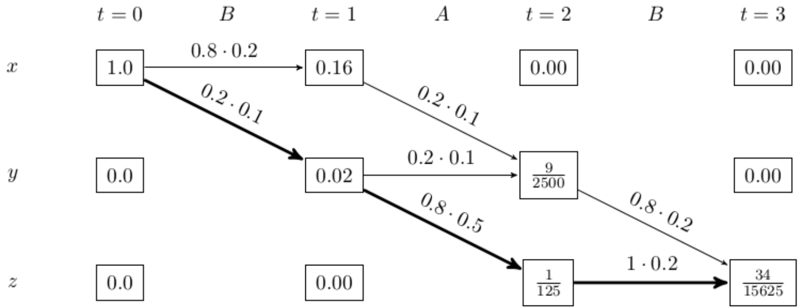
$$\{\hat{y}_0,\dots,\hat{y}_k\} =$$

$$= \arg\max_{y_k} \hat{P}(\hat{y}_0,\dots,\hat{y}_{k-1},y_k \mid x_0,\dots,x_k) =$$

$$= \arg\max P^*_{k-1}\hat{P}(x_k|y_k)\hat{P}(y_k|\hat{y}_{k-1}). \quad (1)$$



Viterbi algorithm



Concatenative unit-selection : pros and cons

Pros

- Big representative corpus $\Rightarrow$ outperforms all other approaches (intelligibility and naturalness).
- Generally easy (fast) training.

Cons

- Large model size (data base), inadequate for offline mode.
- Low flexibility, ability to adapt to new contexts / new tasks.

Concatenative unit-selection in our life

Production examples

Siri (Apple) (2016–2017) :

Concatenative unit-selection in our life

Production examples

Siri (Apple) (2016–2017) :
hybrid unit-selection approach
with deep-learning based emission/transition cost estimation.

Concatenative unit-selection in our life

Production examples

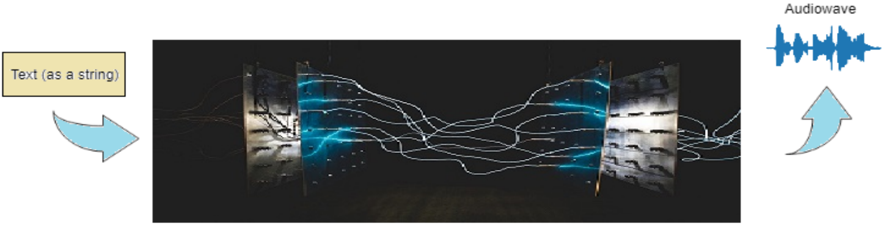
Siri (Apple) (2016–2017) :
hybrid unit-selection approach
with deep-learning based emission/transition cost estimation.

See for yourself !

- Find a pronunciaton dictionary.
- Open-source phonemizer
(*type "python phonemizer" in Google ;)*).
- **Festvox / Flite** :
open-source toolkit
by the Carnegie Mellon University's speech group.

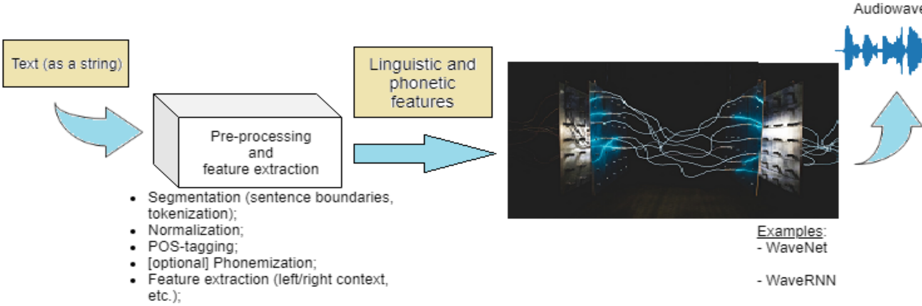
Text-To-Speech (TTS) : end-2-end speech synthesis

Text-To-Speech (TTS) : end-2-end speech synthesis



[Photo credits :
www.unsplash.com/search/photos/electricity]

Text-To-Speech (TTS) : end-2-end speech synthesis



[Photo credits : www.unsplash.com/search/photos/electricity]

End-2-end speech synthesis : pros and cons

Pros

- Saves feature-engineering effort.
- In theory very flexible :
  - can be embedded in a multi-tasking neural net ;
  - allows for efficient style transfer (voice conversion).

End-2-end speech synthesis : pros and cons

Pros

- Saves feature-engineering effort.
- In theory very flexible :
  - can be embedded in a multi-tasking neural net ;
  - allows for efficient style transfer (voice conversion).

Cons

- Time !

End-2-end speech synthesis : pros and cons

Pros

- Saves feature-engineering effort.
- In theory very flexible :
  - can be embedded in a multi-tasking neural net ;
  - allows for efficient style transfer (voice conversion).

Cons

- Time !

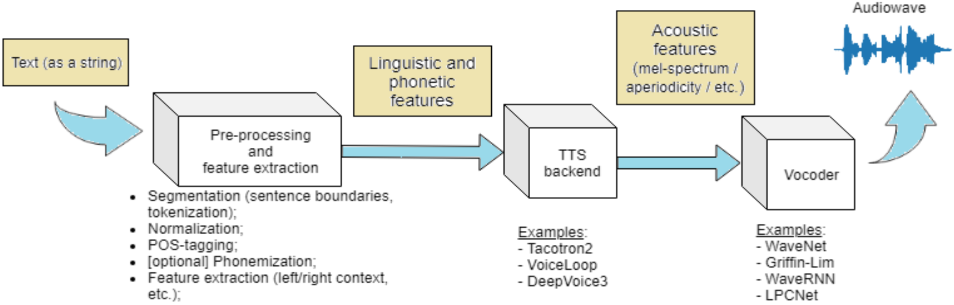
```
if args.mode == 'synthesis':
    raise ValueError('I don\'t recommend running WaveNet on entire dataset.. The world might end before the synthe
```

Original WaveNet model : 1 hour to generate 1 second of audio.

Text-To-Speech (TTS) : parametric speech synthesis

Statistical Parametric Speech Synthesis :

1. Extract and model a parametric representation of the speech signal (spectrum, excitation, etc.).
2. Reconstruct the waveform from the parametric representation.



Parametric speech synthesis : in production

SPSS synthesis : production examples

Google assistant, **Amazon Alexa**,
Huawei assistant.

Huawei is Looking for Talents !

Two Types of Job Opportunities

Huawei is Looking for Talents !

Two Types of Job Opportunities

1. Saint-Petersburg Research Center : Data Science Engineer.

Huawei is Looking for Talents !

Two Types of Job Opportunities

1. Saint-Petersburg Research Center : Data Science Engineer.
2. Moscow Research Center : Research Engineer.

Huawei : jobs at Saint-Petersburg Research Center

Data Science Engineer : Speech Synthesis Team

Huawei : jobs at Saint-Petersburg Research Center

Data Science Engineer : Speech Synthesis Team

- Track the current state-of-the-art in academic research.

Huawei : jobs at Saint-Petersburg Research Center

Data Science Engineer : Speech Synthesis Team

- Track the current state-of-the-art in academic research.
- Experiment with existing implementations / implement missing components.

Huawei : jobs at Saint-Petersburg Research Center

Data Science Engineer : Speech Synthesis Team

- Track the current state-of-the-art in academic research.
- Experiment with existing implementations / implement missing components.
- Find ways to optimize :
  - model size (minimize) ;
  - generation speed (minimize).

Huawei : jobs at Saint-Petersburg Research Center

Data Science Engineer : Speech Synthesis Team

- Adapt to new tasks :
  - model emotions ;
  - mode for non-native speakers ;
  - voice conversion.

Huawei : jobs at Saint-Petersburg Research Center

Contacts

- Me (Yulia MATVEEVA) : matveeva.yulia@huawei.com, **yu125@statmod.ru**
- Saint-Petersburg Huawei R&D HR department : **chernysheva.yuliya@huawei.com**

Huawei : jobs at Saint-Petersburg Research Center

Digital Signal Processing and Speech Synthesis : References (links)

- Rabiner, Schafer, 2009, Theory and Applications of Digital Speech Processing.
- Zen et al., 2009, Statistical Parametric Speech Synthesis.
- Oord et al., 2016, WAVENET: A GENERATIVE MODEL FOR RAW AUDIO.
- Shen et al., 2018, Natural tts synthesis by conditioning wavenet on mel spectrogram predictions.
- Kalchbrenner et al., 2018, Efficient neural audio synthesis.
- Kim et al., 2018, FloWaveNet: A Generative Flow for Raw Audio.

Huawei : Saint-Petersburg Research Center

Other Machine Learning teams in Saint Petersburg :

- Automatic Speech Recognition ;
- Natural Language Understanding ;
- and others.

Huawei : jobs at Moscow Research Center

Research Engineer : Dialogue Systems

- (Team lead) Find unsolved problems in the field.
- (Team lead) Find ways in which the solution to this problem may help the current Huawei projects.
- Work on research projects in the chosen direction.
- Publish in academic journals and participate in academic conferences.

Contacts

- Team Lead (Irina Piontkovskaya) : linkedin.com/in/irina-piontkovskaya-6b10b0b5
- Moscow Huawei R&D HR department : **drobel.valeria@huawei.com**

Huawei : jobs at Moscow Research Center

Dialogue Systems : References (links)

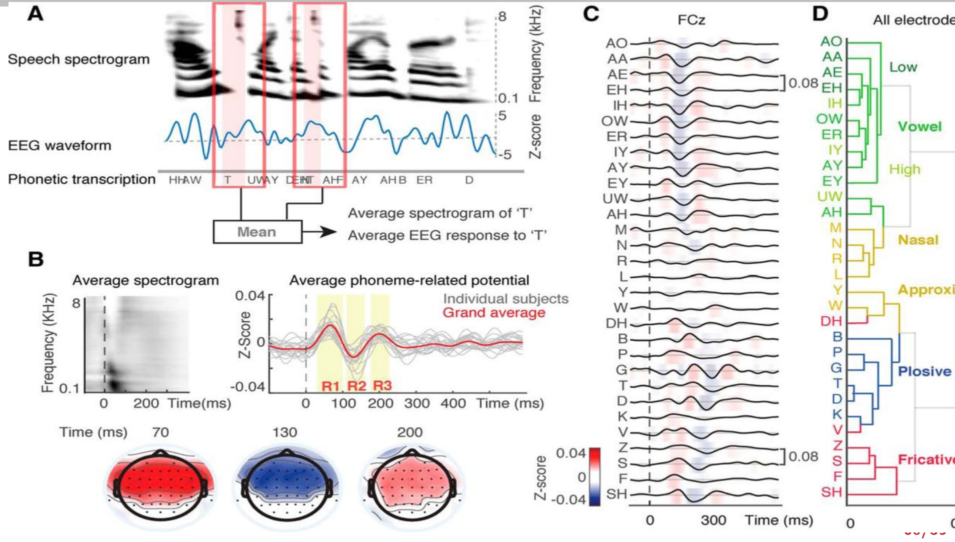
- Zhou et al., 2018, The Design and Implementation of XiaoIce, an Empathetic Social Chatbot
- Shah et al., 2018, Building a Conversational Agent Overnight with Dialogue Self-Play
- Artetxe et al., 2019, An Effective Approach to Unsupervised Machine Translation
- Devlin et al., 2018, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
- Lample and Conneau, 2019, Cross-lingual Language Model Pretraining

Thank you !

Thank you ! Questions ?

Yulia MATVEEVA
matveeva.yulia@huawei.com

Phonemization



Other image credits

- journals.plos.org/plosone/article?id=10.1371/journal.pone.0024516
- http://latlcui.unige.ch/phonetique/easyalign.php
- Bahar Khalighinejad, Guilherme Cruzatto da Silva, and Nima Mesgarani, 2017, *Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech*, The Journal of Neuroscience, 37(8), pp. 2176 – 2185.
- www.businessinsider.com/rick-and-morty-review-2015-7?r=US&IR=T
- www.youtube.com/watch?v=X3pa0mcrTjQ
- www.inverse.com/article/31728-straitum-causes-anxiety-over-future
- http://www.tamasbedo.com/checking-poker-graph-can-hurt-results