NATURAL LANGUAGE GENERATION

Buğrahan Şahin
Yasemin Yenigün
Instructor: Tunga Güngör

NATURAL LANGUAGE GENERATION

- Interest in natural language generation from an input text or data is rising in the recent years.
- There are many applied areas about this subject such as;
  - Simplification of complex texts, automatic spelling, grammar and text correction, automatic generation of peer reviews for scientific papers, story generation, weather and financial reports, virtual newspapers from sensor data, summaries of patient information in clinical contexts, text generation from visuals
- In this survey, we categorised natural language generation methods as;
  - Text-to-text generation
  - Data-to-text generation
  - Visual-to-text generation

GENERATION METHODS: Text-to-Text Generation

- Text-to-text methods, are the methods where NLG is applied to generate texts, from the data already in the text format.
- For example; summarizations, referencing, paraphrasing etc.
- As an example to text-to-text generation method, we analysed the study named **“Your Paper has been Accepted, Rejected, or Whatever: Automatic Generation of Scientific Paper Reviews”.**

Automatic Generation of Scientific Paper Reviews

- In this study, they examine the applicability of a tool that can generate fake reviews for a given scientific paper automatically.
- Given a paper α and overall recommendation $o \in$ {accept, neutral, reject}, a review *r* which looks like as generated by a human for the paper α and which expresses an o for α.
- The method needs a set of real paper reviews R that each of the reviews are written by humans.
- Each review is pre-processed as following;
  - The Named-entity Recognition (NER) is executed on the sequence
  - Part-of-Speech (POS) annotation is executed on the sequence
  - Each token is classified as being or not being a scientific term

Automatic Generation of Scientific Paper Reviews

- There are three steps while generating a review for a paper α with a recommendation o:
  - It constructs a set S of sentences from the reviews in set R and exchanges each specific term in each sentence with a specific term of α
  - It deletes from S the sentences which states a sentiment which is not consistent with o
  - It rearranges and concatenates the sentences in S acquire a review for α

GENERATION METHODS: Data-to-Text Generation

- When there is a complex data, in large amounts, extracting meaningful statistics is needed to make use of it, but even then, only an expert will be able to make sense of the statistics.
- Both making sense of statistics, and the extracting them, are highly costly, and hard to find expertise.
- So along with methods, to gather data, query, extract statistics, NLG is employed to give every day user the essence of the data
- Namely, by using NLG data is put it into the form that is understandable to user.
- Here we examined quite a few examples of the area.

Interacting with financial data using natural language

Reasons to use NLG?

* Problem with queries

What has been done?

* Query is retrieved
* Templates applied

    Where does this templates come from?

* Selection on sentences made

Generating Automated News to Explain the Meaning of Sensor Data

Sensor Data:

* Largely available
* High in quantity

Example:

* SAIH

  Water Levels, Water Flows, Meteorological Data

Input Representation

* Events
* Paths
* Aggregates

Path of Solution

Discourse planner

Data Analyser

* Aggregate Extension
* Event Generation

Presentation Generator

The application developed and has been used more than a year.

Towards NLG for Physiological Data Monitoring with Body Area Networks

- This paper proposes a natural language generation framework that provides a summary text generation from body area networks (BAN).
- The system collects data by measuring heart rate and respiration rate using wearable sensor.
- There is one difficulty and it is the large volumes of data which is collected with the wearable sensors.
- And there are few challenges:
  - how to analyse physiological data in a way that the collected information can help the end user
  - understanding the audience of the generated text, since people from different backgrounds are using the health monitoring

Towards NLG for Physiological Data Monitoring with Body Area Networks

The system has two types of measurements of input data. First is single measurement which is a continuous recorded data record and second is batch measurement is set of single measurements which provides a wider look to the information that is collected

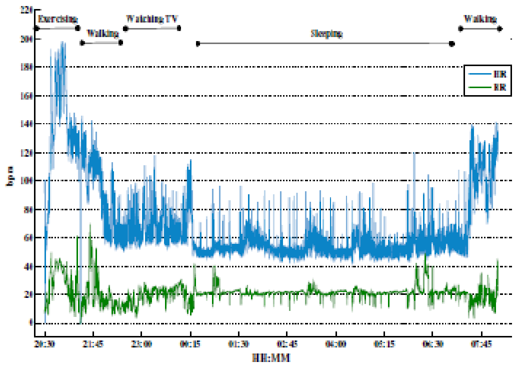

Figure 1: *An example of single measurement 13 hours of heart rate and respiration rate*

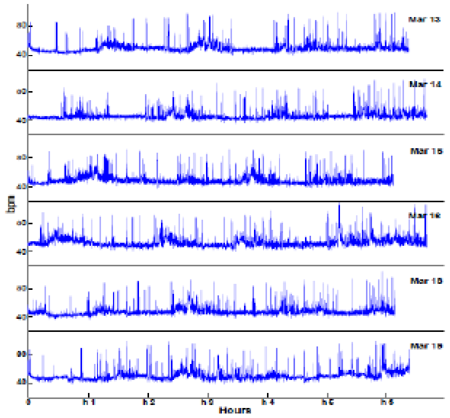

Figure 2: *An example of batch measurement included heart rate for 6 nights*

Towards NLG for Physiological Data Monitoring with Body Area Networks

Sample output from the designed interface:

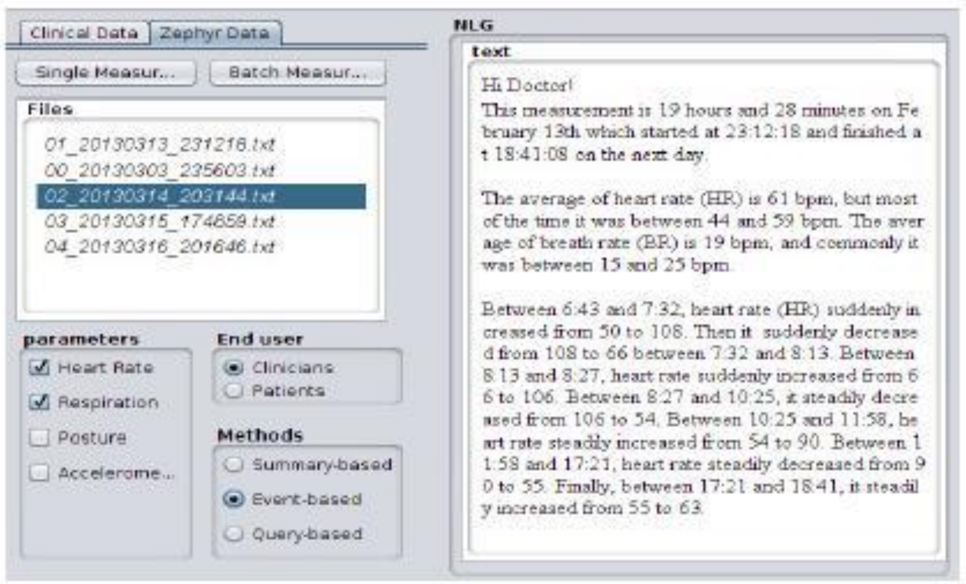

Figure 3: A screenshot of the implemented interface

GENERATION METHODS: Visual-to-Text Generation

- Unlike the previous examples, humans are quite good at extracting meaning from the visual input.
- But considering the size of the data available in the format of images in the internet, we can easily conclude that, in quantity this information is far past the human capabilities in terms of complexity.
- Visual-to-text generation is presumably an example of data-to-text generation, where the input is in the form of an image.
- There are many researches and implementations in this area. We overviewed "**Generating Visual Explanations**" in this part.

Generating Visual Explanations

- A new method is represented that concentrates on;
  - the discriminating properties of the visual object, for that visual object the method guesses a class label and explains the reason for selection of the label for that particular object.
- They propose a novel loss function depending on sampling and reinforcement learning that learns to generate sentences that comprehend a global sentence property, for example class specificity.
- The purpose of the visual explanation model below is to generate an explanation that describes visual content shown in a particular image and contains convenient information to explain the reason of an image belongs to a category that is specified

Generating Visual Explanations

- A sentence that is generated is image relevant if it comments about concepts which are mentioned in ground truth reference sentences for the image.
- The class relevance is measured by considering the similarity of generated sentences for a class are to ground truth sentences for that class.
- Sentences which explains a particular bird class, for example “cardinal”, should have similar words and phrases to ground truth “cardinal” sentences, but not ground truth “black bird” sentences.

Generating Visual Explanations

	Image Relevance		Class Relevance		Best Explanation
	METEOR	CIDEr	Similarity	Rank (1-200)	Bird Expert Rank (1-5)
Definition	27.9	43.8	42.60	15.82	2.92
Description	27.7	42.0	35.3	24.43	3.11
Explanation-Label	28.1	44.7	40.86	17.69	2.97
Explanation-Dis.	28.8	51.9	43.61	19.80	3.22
Explanation	**29.2**	**56.7**	**52.25**	**13.12**	**2.78**

Table 1: Results for the explanation model in comparison to the definition and description baseline, as well as the explanation-label and explanation-discriminative models

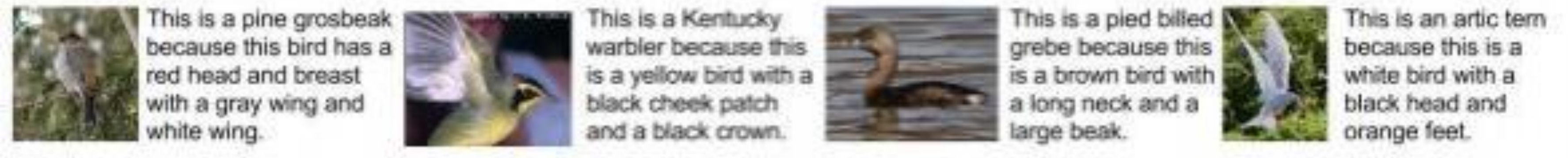

Figure 4: Some sample generated visual explanations

Story Generation

Traditional Methods Requires:

* Repository of Background

* Character Information

* Plot

And… They Are Hand-Crafted

Study of Interest

Topic and Length Supplied

* What is so special about topic?

Relational Parser Is Employed

Precedence Relationships Found

* Event chains, based on words constructed

Grammar Rules are Applied

Surface Realisation Made

Ranking Made

Conclusion

We have seen different methods, employed for different areas for different reasons.

Main Reason:

To make sense of the data

Either in large in quantity, or unreadable format, or larger, longer text, or not organized.

Why?

REFERENCES

[1] Banaee, H. A. (2013). Towards NLG for Physiological Data Monitoring with Body Area Networks. *ENGL'13*, (pp. 193-197).

[2] Bartoli, A., De Lorenzo, A., Medvet, E., & Tarlao, F. (2016). Your paper has been accepted, rejected, or whatever: Automatic generation of scientific paper reviews. In International Conference on Availability, Reliability, and Security, pp. 19-28.

[3] Gatt, A., & Krahmer, E. (2017). Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation.

[4] Hendricks, L. A., Akata, Z., Rohrbach, M., Donahue, J., Schiele, B., & Darrell, T. (2016a). Generating Visual Explanations. In Proc. ECCV'16.

[5] McIntyre, Neil, and Mirella Lapata. "Learning to tell tales: A data-driven approach to story generation." *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. Association for Computational Linguistics, 2009.

REFERENCES

[6] Molina, Martin, Amanda Stent, and Enrique Parodi. "Generating automated news to explain the meaning of sensor data." *International Symposium on Intelligent Data Analysis*. Springer Berlin Heidelberg, 2011.

[7] Plachouras, Vassilis, et al. "Interacting with financial data using natural language." *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2016.