EVALUATION METRICS FOR SYNTHETIC SOCIAL MEDIA-DERIVED TEXTS

VYTAUTO DIDŽIOJO UNIVERSITETAS
Informatikos fakultetas

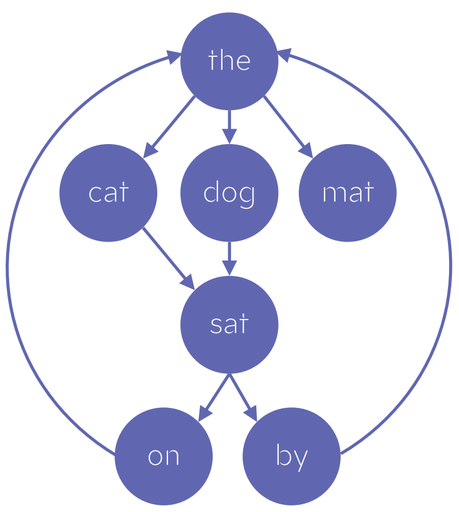
CONTEXT

- Social network data (texts and networks) are in high demand
- Data protection regulations hamper the collection of this kind of data
- As a result, interest in synthetic data is on the rise
- Language models can be used for synthetic data generation
- We evaluate language models in this context

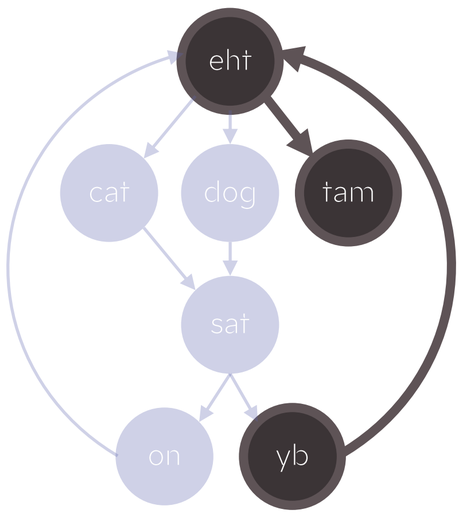
GOAL

- Evaluate the **flexibility** of language models
- Operationalize flexibility as ability to incorporate artificial jargon
- Derive metrics

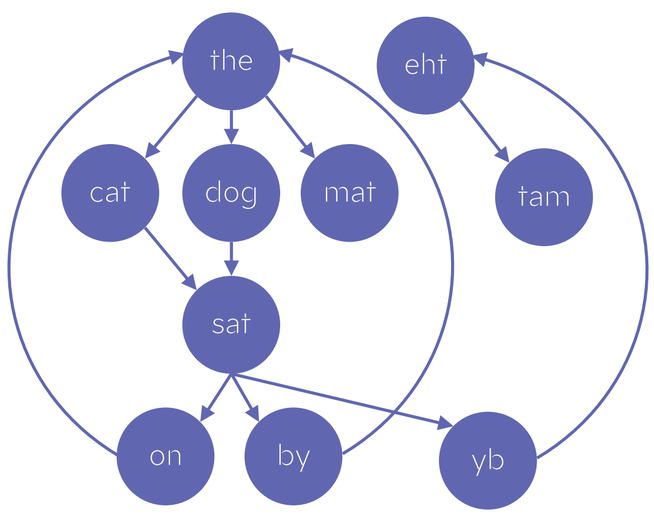
METHOD



1 Collect all word bigrams in original corpus



2 Pick bigrams to "jargonize" by reversing their words



3 Make corresponding replacements to create a transformed corpus

METRICS

After amplification of the transformed corpus using language models, do the transformed words...

...behave as synonyms of their original counterparts?

$$\frac{1}{N}\sum_{w} 1 - \frac{\text{rank of } f(w)}{|f|}$$

M1 Synonymy

...have a usage preference after (before) the same words as in the transformed corpus?

M2 Preference

...enter the same topics as their original synonyms?

$$\frac{1}{N}\sum_{w} 1 - d_{w,\,f(w)}$$

M3 Topicality

...appear in the same bigrams as selected during the transformation process?

$$\frac{1}{N}\sum_{w_1,w_2} \begin{cases} 1, & \text{if } (w_1, w_2) \in T \\ 0, & \text{otherwise} \end{cases}$$

M4 "Grammar"

FINDINGS

- A Markov chain (order 3) and a fine-tune GPT-2 snapshot have been used for a preliminary comparison
- The results generally give initial validation to the metrics

FUTURE PLANS

- Training of more comparable models for more results
- Interpretation of models training results
- Correlation with conventional metrics (e.g., perplexity)
- Assessing the suitability of the metrics as a proxy for human-reported model flexibility

NOTATION

w – word; $f(w)$ – transformed word; $|f|$ – total number of transformed words; $d_{w,\,f(w)}$ – topic-topic distance between terms w and $f(w)$, derived from a term-document matrix; T – set of all transformed bigrams; $\frac{1}{N}$ – context-specific normalization factor

AUTHORS:

Ignas Rudaitis
ignas.rudaitis@card-ai.eu

Justina Mandravickaitė
justina.mandravickaite@vdu.lt

Danguolė Kalinauskaitė
danguole.kalinauskaite@vdu.lt

Veronika Gvozdovaitė
veronika.gvozdovaite@card-ai.eu

Tomas Krilavičius
tomas.krilavicius@vdu.lt

CARD
CENTRE FOR APPLIED RESEARCH AND DEVELOPMENT