



Факультет компьютерных наук

22 апреля 2025 г.

# Шутки в сторону: машинное обучение и интерпретируемый искусственный интеллект в задачах генерации юмористических текстов

Король Михаил БПМИ2310

[mkorol@hse.ru](mailto:mkorol@hse.ru)

Научный руководитель: д.ф.-м.н. профессор Громов В.А.



# Введение

Актуальность, цель и гипотеза

В данный момент ИИ не умеет генерировать юмор. Точнее, из множества сгенерированных шуток, довольно низкий процент окажется действительно смешным. Несмотря на большое количество работ, посвященных юмору, он остается одним из самых сложных явлений для понимания и формализации с точки зрения науки.

Цель: найти качественные различия между обычными и юмористическими текстами для создания методов их автоматической классификации.

Гипотеза: существуют фундаментальные различия в структуре языка, используемого в юмористических и литературных текстах, которые могут быть выявлены и количественно описаны с помощью методов теории хаоса и топологического анализа.



## Литературный обзор



Мартин, Пластино и Россо (MPR) предлагают подход, позволяющий отличить хаотический ряд от ряда, генерируемого простой детерминированной системой, и от ряда, генерируемого случайным образом. Чтобы использовать такой метод, нужно как-то представить наши текста в виде временного ряда. Собран корпус анекдотов и корпус литературы. Произведена базовая обработка корпусов, которая включает в себя очистку данных и лемматизацию. Далее с помощью словаря эмбедингов получаем ряд векторов.



Рассмотрим наблюдаемую часть временного ряда  $y_0, y_1, \dots, y_t, \dots$  и разобьем его на отрезки длины  $k$ . В теории их называют  $z$ -вектора.

$$z_0 = (y_0, y_1, \dots, y_{k-1})$$

$$z_1 = (y_1, y_2, \dots, y_k)$$

И так далее. Суть метода заключается в вычислении двух величин, основываясь на полученных вероятностях, характеризующих исходный временной ряд.



Первая величина – это привычная нам энтропия, но нормированная на ее максимальное значение ( $\log m$ )

$$0 \leq H \leq 1$$

Вторая характеристика носит название сложности, а если быть точным, MPR-сложности (которая названа по первым буквам фамилий ее авторов).

$$C_{\text{MPR}} = Q_0 \cdot H \cdot \|(P - P_e)\|$$

где  $P_e$  – равномерное распределение, то есть:  $P_e = \{p_i = 1/N\}$ ,  $H$  – энтропия,  $Q_0$  – нормализующая константа, которая гарантирует, что  $0 \leq C_{\text{MPR}} \leq 1$ ,  $\|(P - P_e)\|$  показывает, насколько уклоняется актуальное распределение от распределения равномерного.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.





Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



# Список

## Нумерованный список

1. Первый пункт:
  - подпункт 1;
  - подпункт 2.
2. Второй пункт
  - 2.1 нумерованный подпункт.
3. Третий пункт





# Список

Маркированный список

- Первый пункт:
  - подпункт 1;
  - подпункт 2.
- Второй пункт
  - 1. нумерованный подпункт.
- Третий пункт



## Слайд с двумя колонками текста

### 1. Первый пункт:

- подпункт 1;
- подпункт 2.

### 2. Второй пункт

2.1 нумерованный подпункт.

### 3. Третий пункт

### • Первый пункт:

- подпункт 1;
- подпункт 2.

### • Второй пункт

1. нумерованный подпункт.

### • Третий пункт



Слайд с картинкой

Текст рядом с картинкой





### Теорема (Пифагора)

Если  $a$  и  $b$  — длины катетов прямоугольного треугольника, а  $c$  — длина гипотенузы, то  $a^2 + b^2 = c^2$ .

### Блок с красным заголовком

Содержимое.

### Блок с зеленым заголовком

Содержимое.



Просто слайд с текстом