
Детекция манипуляций в новостном потоке

A Preprint

Мелихов Дмитрий Александрович
Факультет вычислительной математики и кибернетики
МГУ им. Ломоносова
s02200440@gse.cs.msu.ru

Воронцов Константин Вячеславович
Факультет вычислительной математики и кибернетики
МГУ им. Ломоносова
vokov@forecsys.ru

Abstract

В работе решается задача выявления манипуляций в новостном потоке. В новостных статьях выделяются манипулятивные фрагменты и помечается тип манипуляции. Фрагменты объединяются в элементы разметки и образуют гиперграф. В работе предлагается модель на основе больших лингвистических моделей, которая выявляет фрагменты и моделирует связи между ними. Для выявления фрагментов решается задача span detection. Для построения графа используются text2graph модель в паре с graph2text, которые обучаются с помощью техники back translation. Строится векторное представление фрагментов и предсказываются связи между ними.

Keywords span identification · text tagging · manipulation detection

1 Introduction

Современные средства массовой информации генерируют огромный поток данных на социально-политические темы. При этом они охватывают огромное число читателей и во многом формируют у них определённый набор ценностей. Из этого возникает потребность автоматически обрабатывать новостной поток для выявления манипуляций.

Манипуляцией в тексте называется воздействие на читателя с целью сформировать определённое отношение к цели (мишени манипуляции). Среди манипуляций можно выделить: эмоциональное воздействие, предоставление недостоверной информации, ложные причинноследственные связи.

К обработке новостного потока можно подходить с точки задачи классификации или регрессии. В 2007 году появилось соревнование SemEval-2007 Task 14[?], где основная задача - выявление эмоциональной нагрузки заголовков статей, которая сформулирована как задача регрессии - каждой эмоции сопоставляется число от 0 до 100. В 2022 году предложили датасет на основе статей с Rappler для выявления эмоций читателя[?]. Существует постановка задачи, в которой новости нужно классифицировать по политической идеологии, например определить кто написал статью: левый, правый или центрист. В статье[?] предлагается датасет на основе данных с сайта AllSides¹. Также в данной статье освещается важная проблема - смещение по источнику новости (media bias). Моделям проще выучить стиль написания статей разными СМИ и предсказывать политическую идеологию по ним, вместо того, чтобы опираться на утверждения в тексте.

¹<https://www.allsides.com/media-bias/media-bias-rating-methods>

2 Headings: first level

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. See Section 2.

2.1 Headings: second level

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})} \quad (1)$$

2.1.1 Headings: third level

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Paragraph Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

3 Examples of citations, figures, tables, references

3.1 Citations

Citations use **natbib**. The documentation may be found at

<http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf>

Here is an example usage of the two main commands (`citet` and `citep`): Some people thought a thing [??] but other people thought something else [?]. Many people have speculated that if we knew exactly why ? thought this...

3.2 Figures

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi. See Figure 1. Here is how you add footnotes.² Sed feugiat. Cum sociis natoque penatibus

²Sample of the first footnote.



Рис. 1: Sample figure caption.

Таблица 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	~ 100
Axon	Output terminal	~ 10
Soma	Cell body	up to 10^6

et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

3.3 Tables

See awesome Table 1.

The documentation for `booktabs` (‘Publication quality tables in LaTeX’) is available from:

<https://www.ctan.org/pkg/booktabs>

3.4 Lists

- Lorem ipsum dolor sit amet
- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

Список литературы

Carlo Strapparava and Rada Mihalcea. SemEval-2007 task 14: Affective text. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 70–74, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL <https://aclanthology.org/S07-1013>.

- Anoop K., Deepak P., Savitha Sam Abraham, Lajish V. L., and Manjary P. Gangan. Readers' affect: predicting and understanding readers' emotions with deep learning. *Journal of Big Data*, 9(1):82, Jun 2022. ISSN 2196-1115. doi:10.1186/s40537-022-00614-2. URL <https://doi.org/10.1186/s40537-022-00614-2>.
- Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991, Online, November 2020. Association for Computational Linguistics. doi:10.18653/v1/2020.emnlp-main.404. URL <https://aclanthology.org/2020.emnlp-main.404>.
- George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR)*, 2014 14th International Conference on, pages 417–422. IEEE, 2014a.
- Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.
- George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR)*, 2014 6th International Conference of, pages 312–318. IEEE, 2014b. doi:10.1109/SOCPAR.2014.7008025.