

Information Extraction – Seminar Topics

WS 2020–2021, English topics

Jindřich Libovický
libovicky@cis.lmu.de

LMU München, CIS

November 10, 2020

Fake Reviews Detection

- Internet is full of fake reviews of products, restaurants, ...
- Humans are not particularly good in detecting what reviews are fake
- Can we do it automatically?
- Can fake reviews be automatically generated?



The screenshot shows a product listing for "100% Genuine Snake Oil" by "Scammer's Warehouse". The product image is a bottle of snake oil. The listing includes a 5-star rating from 42 customer reviews and a price of \$89.70 with Prime shipping. Below the product listing is a review by "Fake Jim" dated June 19, 2017, for the item "Snake oil, 4 oz.". The review text reads: "Very good product. I can't prove this for certain, but I think it cured my cancer. I feel like I'm 17 again." The "HUSTLE" logo is visible in the bottom right corner of the screenshot.

100% Genuine Snake Oil
By: [Scammer's Warehouse](#)
★★★★★ 42 customer reviews
Price: **\$89.70** ✓Prime

★★★★★ **AMAZING** healing qualities
By: [Fake Jim](#) on June 19, 2017
Item: Snake oil, 4 oz.
Very good product. I can't prove this for certain, but I think it cured my cancer. I feel like I'm 17 again.

HUSTLE

Fake Reviews Detection: Papers

- A. Mukherjee, V. Venkataraman, B. Liu, N. Glance, et al. **Fake review detection: Classification and analysis of real and pseudo reviews**. *Technical Report UIC-CS-2013-03, University of Illinois at Chicago, Tech. Rep.*, 2013
- S. Kennedy, N. Walsh, K. Sloka, A. McCarren, and J. Foster. **Fact or factitious? contextualized opinion spam detection**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 344–350, Florence, Italy, July 2019
- D. I. Adelani, H. Mai, F. Fang, H. H. Nguyen, J. Yamagishi, and I. Echizen. **Generating sentiment-preserving fake online reviews using neural language models and their human- and machine-based detection**. In L. Barolli, F. Amato, F. Moscato, T. Enokido, and M. Takizawa, editors, *Advanced Information Networking and Applications - Proceedings of the 34th International Conference on Advanced Information Networking and Applications, AINA-2020, Caserta, Italy, 15-17 April*, volume 1151 of *Advances in Intelligent Systems and Computing*, pages 1341–1354. Springer, 2020

Abstractive Text Summarization using Deep Learning

- Task: Write short abstract of a long text
- 2015: Summarization is special case of machine translation
- Later: Summarization is a different problem, specialized architectures
- Recent success of very large language models (BERT, GPT-2)

more abstractive summaries compared to prior work that employs a copy mechanism while still achieving higher rouge scores. *Note: The abstract above was not written by the authors, it was generated by one of the models presented in this paper.*

Abstractive Text Summarization using Deep Learning: Papers

- A. M. Rush, S. Chopra, and J. Weston. **A neural attention model for abstractive sentence summarization**. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal, Sept. 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1044. URL <https://www.aclweb.org/anthology/D15-1044>
- S. Subramanian, R. Li, J. Pilault, and C. J. Pal. **On extractive and abstractive neural document summarization with transformer language models**. *CoRR*, abs/1909.03186, 2019. URL <http://arxiv.org/abs/1909.03186>
- A. See, P. J. Liu, and C. D. Manning. **Get to the point: Summarization with pointer-generator networks**. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1099. URL <https://www.aclweb.org/anthology/P17-1099>

Task: find answer to question in text.

One of the most famous people born in Warsaw was Maria Skłodowska-Curie, who achieved international recognition for her research on radioactivity and was the first female recipient of the Nobel Prize. Famous musicians include Władysław Szpilman and Frédéric Chopin. Though Chopin was born in the village of Żelazowa Wola, about 60 km (37 mi) from Warsaw, he moved to the city with his family when he was seven months old. Casimir Pulaski, a Polish general and hero of the American Revolutionary War, was born here in 1745.

What was Maria Curie the first female recipient of?

Ground Truth Answers: Nobel Prize Nobel Prize Nobel Prize

What year was Casimir Pulaski born in Warsaw?

Ground Truth Answers: 1745 1745 1745

Who was one of the most famous people born in Warsaw?

Ground Truth Answers: Maria Skłodowska-Curie Maria Skłodowska-Curie Maria Skłodowska-Curie

Who was Frédéric Chopin?

Ground Truth Answers: Famous musicians musicians Famous musicians

How old was Chopin when he moved to Warsaw with his family?

Ground Truth Answers: seven months old seven months old seven months old

How do the datasets look like?

How do the models look like?

Answer Span Selection: Papers

- P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang. **SQuAD: 100,000+ questions for machine comprehension of text**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas, Nov. 2016. Association for Computational Linguistics. doi: [10.18653/v1/D16-1264](https://doi.org/10.18653/v1/D16-1264). URL <https://www.aclweb.org/anthology/D16-1264>
- M. J. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi. **Bidirectional attention flow for machine comprehension**. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=HJ0UKP9ge>
- P. Rajpurkar, R. Jia, and P. Liang. **Know what you don't know: Unanswerable questions for SQuAD**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 784–789, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: [10.18653/v1/P18-2124](https://doi.org/10.18653/v1/P18-2124). URL <https://www.aclweb.org/anthology/P18-2124>

Depressions in Online Forums

- People write various things on Reddit and other fora—e.g., that they depressed or have suicidal thoughts
- Can we use this data somehow to help people?
- Can we predict user will suffer with depression before he/she knows?

Depressions in Online Forums: Papers

- A. Yates, A. Cohan, and N. Goharian. **Depression and self-harm risk assessment in online forums.** In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978, Copenhagen, Denmark, Sept. 2017. Association for Computational Linguistics. doi: [10.18653/v1/D17-1322](https://doi.org/10.18653/v1/D17-1322). URL <https://www.aclweb.org/anthology/D17-1322>
- H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik. **Expert, crowdsourced, and machine assessment of suicide risk via online postings.** In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36, New Orleans, LA, June 2018. Association for Computational Linguistics. doi: [10.18653/v1/W18-0603](https://doi.org/10.18653/v1/W18-0603). URL <https://www.aclweb.org/anthology/W18-0603>
- J. Ive, G. Gkotsis, R. Dutta, R. Stewart, and S. Velupillai. **Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health.** In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 69–77, New Orleans, LA, June 2018. Association for Computational Linguistics. doi: [10.18653/v1/W18-0607](https://doi.org/10.18653/v1/W18-0607). URL <https://www.aclweb.org/anthology/W18-0607>