Group 19

Reliability Prediction for Health-Related Content

Tang Wanjia | Mirjana Sadikovikj | Tobias Abraham Haider | Anum Hafeez

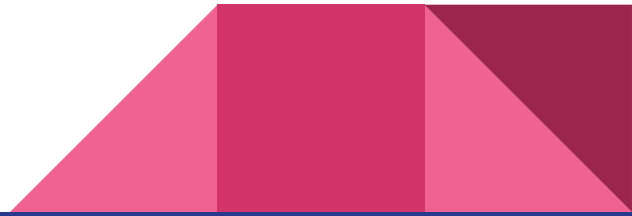
Strategy

- Recreating:
  - A data processing and classification experiment
  - The model detects medical disinformation web pages
- Reason for choosing this paper:
  - Well-defined topic and experiment setting
  - Detailed reports of the performance metrics
- Necessary steps to run the code
  - Run the provided jupyter notebook of the adapted code repository

Work progress

- Issues:
  - Custom implementations and outdated dependencies
  - Static file paths incompatible with Windows
  - High hardware requirements -> one experiment can not be reproduced

- Achievements:
  - Elimination of an outdated dependency (svmlight)
  - Minor code refactoring into one structured jupyter notebook
  - Successful reproduction of the first experiment cases



First results

					Weighted accyracy (%).		
Features	SVM cost factor	F1	F1(reliable class	F1 (non reliable class)	1	2	3
Links	1	**0.94**	**0.97**	**0**	**93.75**	—	—
	2	0.94	0.97	0	—	88.26	—
	3	0.94	0.97	0	—	—	83.4
Links + Commercial	1	**0.94**	**0.97**	**0**	**93.75**	—	—
	2	0.94	0.97	0	—	88.26	—
	3	0.94	0.97	0	—	—	83.4
Words (removing stopwards)	1	**0.91**	**0.95**	**0**	**91.25**	—	—
	2	0.91	0.95	0	—	85.88	—
	3	0.91	0.95	0	—	—	81.13
Words (keeping stopwords)	1	**0.91**	**0.95**	**0**	**91.25**	—	—
	2	0.91	0.95	0	—	85.88	—
	3	0.91	0.95	0	—	—	81.13
All (removing stopwards)	1	**0.91**	**0.95**	**0**	**91.25**	—	—
	2	0.91	0.95	0	—	85.88	—
	3	0.91	0.95	0	—	—	81.13
All (keeping stopwords)	1	**0.91**	**0.95**	**0**	**91.25**	—	—
	2	0.91	0.95	0	—	85.88	—
	3	0.91	0.95	0	—	—	81.13

Next Steps

- Test a few more cases
- Document the technical changes made to the experiment
- Compare the prediction accuracies of various models
- Report and interpret the results