Summarizing Electronic Theses and Dissertations

Instructor: Prof. Edward A. Fox
Department of Computer Science, Virginia Tech
Blacksburg, Virginia - 24061

December 4, 2018

Team 17
Ashin Marin Thomas
Chinmaya Patnayak
Dhruv Sharma
Farnaz Khaghani
John Aromando
Contents

- Our Journey Through The Semester
- Final Judgement - GROBID vs. Scienceparse
- Extractive Baselines: Latent Semantic Analysis, LexRank, Reduction
- Seq2Seq - Model And Problems
- Pointer Generator Networks - Model, Problems And Output
- Hybrid Reinforce-Selected Sentence Rewriting - Model And Improvements
- Final Results
- Evaluation
- Major Takeaways
- Potential Future Work
The Journey

1. Tackling Noisy ETDs
2. Extractive Baselines with Sumy (Python)
3. Seq2Seq Model
4. Pointer Generator Networks
5. Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting
6. Model Tweaking
7. Evaluation
Final Judgement - GROBID vs Scienceparse

- Reasoning - The ETD text files included significant noise that made the data a mess to use
- GROBID: Older library, but cleaner data than the text files. Still contained noise
- ScienceParse: Newer library, converts PDF to JSON format. Contains just as much noise as GROBID. Difficult to set up!
- After running extractive baselines (ex: Latent Semantic Analysis) on extracted text using GROBID and Scienceparse, we found that Scienceparse did not warrant switching over to GROBID
The survey results also suggest that the VHSL should consider developing a plan to oversee game officials and their associations. The majority of participants responded that the VHSL should require head varsity and head junior varsity coaches participate in a coaching education course prior to the beginning of every season (63%). The survey results suggest that legislation regarding home-schooled student participation should be monitored closely by the VHSL and Virginia educators in the future. Recommendations for Future Research into the VHSL, and other state athletic associations, might focus in particular on the following areas based specifically on the results of this study: It will not be possible to identify you as the person who provided any specific information for the study. If you have any questions concerning this research study, please call me at my number listed below, or Dr. David Alexander, professor of education specializing in Educational Leadership and Policy Studies for Virginia Polytechnic Institute and State University, at (540) Completion and submission of the survey will be considered your consent to participate. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact Dr. David M. Moore, the Chair of Institutional Review Board, (540)231-4991. Thank you for your assistance. Sincerely, Scott Jefferies scottj77@vt.edu (540)817-0726
In conclusion, a combined 56% of survey participants responded that appeals should be heard by agencies other than the VHSL. Additionally, the survey results suggest that the VHSL should do more in regards to educating parents on high school athletics and activities. The survey results suggest that there is support for the current safety initiatives, policies, and rules that the VHSL has established and implemented. The majority of participants responded that home-schooled students should not be able to participate in athletics and activities at member schools in the areas where they live (78%). Additionally, more Virginia legislators participated in the Polakiewicz study (21) when compared to the number of Virginia legislators who participated in this study (9) (Polakiewicz, 1985). A larger percentage of survey participants in the Polakiewicz study (87.6%) responded that the VHSL should have the authority to impose a period of probation upon coaches, member schools, and participants when compared to the survey results of this current study (73%) (Polakiewicz, 1985). Additionally, a larger percentage of survey participants responded that the VHSL should have the authority to prohibit the participation of coaches, member schools, and participants in the Polakiewicz study (79.9%) when compared to this current study (70%) (Polakiewicz, 1985). The majority of participants responded in the Polakiewicz study (57.2%) and this current study (59%) that the VHSL imposes penalties that are appropriate (Polakiewicz, 1985). Additionally, high school principals from the Polakiewicz study (67.4%) and high school principals from this current study (67%) responded that the VHSL imposes penalties that are appropriate (Polakiewicz, 1985). The results of this current study suggest that the VHSL should consider either an outside agency established by the Virginia Department of Education (10%) or an impartial fact-finding panel (46%) to hear appeals on VHSL decisions or disciplinary action. The survey results suggest that while there is satisfaction with how the VHSL is performing, the number of complaints received by participants has increased when compared to the Polakiewicz study. The majority of participants in the Polakiewicz study responded that they were satisfied with the VHSL's administration of athletics (76.3%), while 79% of participants responded that they were satisfied with the VHSL's administration of athletics in this study (Polakiewicz, 1985). The survey results suggest that satisfaction with the VHSL's administration of athletics has increased when compared to the Polakiewicz study. In regards to the performance of the VHSL, 23.2% of participants from the Polakiewicz study assigned the VHSL a performance grade of 'A--excellent', while only 9% of the survey participants from this current study assigned the VHSL the same grade (Polakiewicz, 1985).
The purpose of this study was to answer the following questions: The majority of participants responded that the VHSL should have the authority to impose periods of probation on coaches and sponsors, member schools and participants (73%), and prohibit the participation of coaches and sponsors, member schools, and participants (70%). The survey results suggest that the authority of the VHSL to perform such duties is supported by the majority of the groups that participated in this study. In addition to supporting the authority of the VHSL to impose periods of probation and the authority to prohibit participation, the majority of participants (59%) responded that the VHSL currently imposes penalties that are appropriate. Additionally, more Virginia legislators participated in the Polakiewicz study (21) when compared to the number of Virginia legislators who participated in this study (9) (Polakiewicz, 1985). A larger percentage of survey participants in the Polakiewicz study (87.6%) responded that the VHSL should have the authority to impose a period of probation upon coaches, member schools, and participants when compared to the survey results of this current study (73%) (Polakiewicz, 1985). Additionally, a larger percentage of survey participants responded that the VHSL should have the authority to prohibit the participation of coaches, member schools, and participants in the Polakiewicz study (79.9%) when compared to this current study (70%) (Polakiewicz, 1985). The majority of participants responded in the Polakiewicz study (57.2%) and this current study (59%) that the VHSL imposes penalties that are appropriate (Polakiewicz, 1985). Additionally, high school principals from the Polakiewicz study (67.4%) and high school principals from this current study (67%) responded that the VHSL imposes penalties that are appropriate (Polakiewicz, 1985). The results of this current study suggest that the VHSL should consider either an outside agency established by the Virginia Department of Education (10%) or an impartial fact-finding panel (46%) to hear appeals on VHSL decisions or disciplinary action. The survey results suggest that while there is satisfaction with how the VHSL is performing, the number of complaints received by participants has increased when compared to the Polakiewicz study. The majority of participants in the Polakiewicz study responded that they were satisfied with the VHSL's administration of athletics (76.3%), while 79% of participants responded that they were satisfied with the VHSL's administration of athletics in this study (Polakiewicz, 1985). The survey results suggest that satisfaction with the VHSL's administration of athletics has increased when compared to the Polakiewicz study. In regards to the performance of the VHSL, 23.2% of participants from the Polakiewicz study assigned the VHSL a performance grade of 'A--excellent', while only 9% of the survey participants from this current study assigned the VHSL the same grade (Polakiewicz, 1985). In regards to potential issues the VHSL may face over the next five years, the participants identified the following main themes: home-schooled students participating in VHSL athletics and activities, private school participation in VHSL athletics and activities, the financial model of the VHSL, and transgender student policies.
Analysis Of Extractive Baselines

Latent Semantic Analysis: Focuses on results, future work and individuals involved in the study (and their contact information).

LexRank: Focuses on feedback from participants in surveys, specific numbers and the differences between past research and the current research. Includes a lot of what Latent Semantic Analysis produces, but builds off of the text further.

Reduction: Focuses on the same information LexRank focuses on - feedback from survey participants, specific numbers and the differences between past findings and the current findings. Includes some noise. Very detailed.
Seq2Seq: Model Description

- BBC dataset
- arXiv dataset
- DUC
  - Used Glove pre-trained vectors to initialize word embedding,
  - Used LSTM cell with stack-bidirectional-directional-RNN,
  - Used LSTM BasicDecoder for training, and BeamSearchDecoder for inference.

---

- Optimizer SGD -> Adam
- Number of nodes in each hidden layer: 150 -> 300
- Batch size: 64 -> 32
- Input size: 300 -> 750
- Output size: 50 -> 150
Seq2Seq - Problems

- Does not generate grammatically correct sentences always and generates <UNK> frequently
- Involves too much hyperparameter tuning - number of nodes, input-output sequence length, beam search length, embedding size, batch size
- Fixed size embeddings are easily overwhelmed by long inputs and long outputs
- Computationally expensive as attention model attends every word to decide it’s importance
Pointer Generator Networks
Model Description

Dataset:
- CNN/ Daily Mail
- arXiv

Hyper parameters:
- Learning Rate - 0.15
- Adagrad Optimiser
- Initial Accumulator - 0.1
- Vocab Size - 50k
- Hidden dimensional - 256
- Word embedding dimensional - 128
- Model trained - 33 epochs

Processed bin files -
https://drive.google.com/drive/u/1/folders/1MXrrXQTXDr6OrBcYsRU3IBA07XajqQU9
Pointer Generator Networks - Output And Potential Problems

Advantages:
- Removed <UNK> issue from Seq2Seq
- Removed factual errors
- Easy to accurately reproduce phrases
- Fits CNN/Daily Mail dataset

Issues:
- Uses pointer to copy words causing the summary to be more extractive rather than abstractive
- Repeats itself
- PGNs overfit on the arXiv dataset due to the small size (~ 4600 articles)
- No common vocabulary between news articles and ETDs
Abstractive Summarization With Reinforce-Selected Sentence Rewriting

Abstractive Summarization With Reinforce-Selected Sentence Rewriting

Advantages:

● Removes sentence repetitions in summary
● Sentence-level reinforcement learning takes into account the word-sentence hierarchy which better models the language structure
● The model adopts an approach that first extracts all the salient sentences and then rewrites them, avoiding the redundancy issues of attending to all the words

Scope of Improvement:

● Sensitive to noise and unexpected characters
● Length of the generated summary is still unpredictable
An analysis is done and patterns are discovered for the popularity-adjusted block model (PABM) model spectral decomposition. The traditional spectral clustering applicable to a PABM model graph is made by moving to a slightly higher dimensional space where $k^2$ dimensional space is considered. In this space, $k^2$ sub-communities are formed by performing a simple distance-based clustering. These $k^2$ columns need to be transformed to form $k$ communities indicating the communities in the PABM model. For higher numbers of communities, the existence of these $k^2$ sub-communities is proven, and a manual transformation of these $k^2$ sub-communities shows that PABM spectral clustering is accurate, given the general function to transform the $k^2$ sub-communities into $k$ communities. In addition, for the case of two-community PABM model, distance-based clustering in 4-dimensional space leads to 4 sub-communities but there is no function to transform these into two communities.

pabm is a more advanced model compared to sbm. but the existing normalized spectral method doesn’t yield good results in case of pabm. most of the real-world networks fit well to a model. spectral clustering is an existing algorithm for finding communities. so, we try to find a spectral algorithm that works better for pabm model. spectral clustering method exists for sbm model and a more general normalized clustering algorithm exists for dcbm which work very well.
The perceptions of both athletic and school administrators in Virginia regarding the statutes of the Virginia High School League (VHSL) are re-examined in light of the number of student-athletes dramatically increasing to approximately 175,000 since the Polakiewicz dissertation in 1985. As in the Polakiewicz study, the outcome identified and summarized the perceptions of a range of authority figures involved in Virginia education. Perceptions regarding the VHSL are compared and contrasted with the perceptions of those identified in the Polakiewicz study. Since the Polakiewicz study, the VHSL has faced numerous changes and challenges. High school students who participated in extracurricular activities were found likely to be academically successful. Furthermore, socioeconomically disadvantaged students benefited as much, or more than, those students who were not disadvantaged.

Past research has demonstrated that individuals in a promotion-focused state are more inclined to take risks for rewards. Unlike past regulatory fit research, the current study included a fit sensitivity and task performance dependent variable. The study sought to demonstrate that both promotion and prevention fit can lead to equal performance. Additionally, past regulatory fit research illustrates the enhancing effects that a fit-state has in many different variables (Aaker, 2004).
Golden standard

Recently, the private and commercial use of small UAS continues to expand and low altitude air traffic will become more congested raising the risk of mid-air collisions that may result in injuries to people or damage to property below. For many years, commercial manned aircraft have used the Traffic Collision Avoidance System (TCAS) to help ensure that aircraft do not collide in flight [1]. The TCAS and similar collision avoidance systems are useful, however, only when every aircraft in the airspace uses the technology. Small UAS typically operate at low altitude where collision threats include general aviation aircraft and other UAS, which may not include collision avoidance equipment [2]. In these scenarios, it is urgent that the small unmanned aircraft be able to sense and avoid these threats.

Generated summary

cooperaive types: transponder [1, 2, 2. automatic dependent surveillance-broadcast [6] 2. 2. non-cooperative types: radar [7], lidar 11] 11] cooperative types depend on communication between aircraft and aircraft. radar [7], radar [7], cooperative types. the tcas is most prevalent saa sensor since all commercial aircraft are required to equip this system in controlled therefore, required equipment for communication must be installed on each other. there are two main types of sensors that are used [3, 4]: the tcas, for example, requires a transponder. 
Final Results

Evaluation Metrics

ROUGE, or Recall-Oriented Understudy for Gisting Evaluation, is a set of metrics and a software package used for evaluating automatic summarization and machine translation software in natural language processing.

Average ROUGE scores:

- ROUGE-1: 0.1387
- ROUGE-2: 0.1224
- ROUGE-SU4: 0.0480

Possible reasons:

- Varying lengths of the two summaries
- The generated summary captures a very superficial level of the document
Major Takeaways

● Significance of summarization given the information boom
● Text Processing techniques
● Natural Language Processing via Deep Learning and traditional methods
● Data and task dependence of the model
Potential Future Work

- Train the hybrid model with a dataset that covers wider range of topics and vocabulary
- Consider the nature of each chapter in summarization (background, discussion, and result sections may be different for summarization)
- Some of the summaries that are generated use the first person, which needs to be dealt with
Thank you!

- Dr. Fox
- Liuqing Li
- Thank you teams 10 and 16.

Funding Info: GETAR "NSF: IIS-1619028"