1. Detailed Descriptions of the Used Benchmarks

The details descriptions of the datasets that we used in the paper are as follows:

- **AGNews**: Topic classification over four categories of Internet news articles (Del Corso et al., 2005). The four categories are **World**, **Entertainment**, **Sports** and **Business**. Each article is composed of titles plus descriptions classified. The training and test sets respectively contain 120k and 7.6k examples.

- **20newsgroups**\(^1\): The 20 Newsgroups data set is a collection of approximately 20,000 newsgroup documents, partitioned (nearly) evenly across 20 different newsgroups. The training and test sets respectively contain 11.3k and 7.5k examples.

- **Yahoo! Answers**: Topic classification over ten largest main categories from Yahoo! Answers Comprehensive Questions and Answers v1.0, including question titles, question contents and best answers.

- **Yelp Review Polarity (YelpP)**: This dataset is collected from the Yelp Dataset Challenge in 2015, and the task is a binary sentiment classification of polarity. Reviews with 1 and 2 stars are treated as negative and reviews with 4 and 5 stars are positive. The training and test sets respectively contain 560k and 38k examples.

- **IMDB**: This dataset is collected by Maas et al. (2011). This dataset contains an even number of positive and negative reviews. The training and test sets respectively contain 25k and 25k examples.

- **AAPD**: The arXiv Academic Paper dataset (Yang et al., 2018). It is a multi-label benchmark. It contains the abstract and the corresponding subjects of 55,840 papers in the computer science. An academic paper may have multiple subjects and there are 54 subjects in total. We use the splits provided by Yang et al. (2018).

1. **Handcrafted Templates**

In this section, we list templates for different categories for some of the datasets used in this work. Templates for 20 news categories are obtained from Wikipedia definitions:

- **comp.graphics**: Computer graphics is the discipline of generating images with the aid of computers. Today, computer graphics is a core technology in digital photography, film, video games, cell phone and computer displays, and many specialized applications. A great deal of specialized hardware and software has been developed, with the displays of most devices being driven by computer graphics hardware. It is a vast and recently developed area of computer science. The phrase was coined in 1960 by computer graphics researchers Verne Hudson and William Fetter of Boeing. It is often abbreviated as CG, or typically in the context of film as CGI.

- **comp.sys.ibm.pc.hardware**: A personal computer (PC) is a multi-purpose computer whose size, capabilities, and price make it feasible for individual use. Personal computers are intended to be operated directly by an end user, rather than by a computer expert or technician. Unlike large costly minicomputer and mainframes, time-sharing by many people at the same time is not used with personal computers.

- **comp.sys.mac.hardware**: The Macintosh (branded simply as Mac since 1998) is a family of personal computers designed, manufactured and sold by Apple Inc. since January 1984.

- **comp.windows.x**: Windows XP is a personal computer operating system produced by Microsoft as part of the Windows NT family of operating systems. It was released to manufacturing on August 24, 2001, and broadly released for retail sale on October 25, 2001.
Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the Internet using a web browser. Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine, which displays the same products availability and pricing at different e-retailers. As of 2016, customers can shop online using a range of different computers and devices, including desktop computers, laptops, tablet computers and smartphones.

A gun is a ranged weapon typically designed to pneumatically discharge solid projectiles but can also be liquid (as in water guns/cannons and projected water disruptors) or even charged particles (as in a plasma gun) and may be free-flying (as with bullets and artillery shells) or tethered (as with Taser guns, spearguns and harpoon guns). Disruptors may be free-flying (as with bullets and artillery shells) or tethered (as with Taser guns, spearguns and harpoon guns).

A motorcycle, often called a bike, motorbike, or cycle, is a two- or three-wheeled motor vehicle. Motorcycle design varies greatly to suit a range of different purposes: long distance travel, commuting, cruising, sport including racing, and off-road riding. Motocycling is riding a motorcycle and related social activity such as joining a motorcycle club and attending motorcycle rallies.

Baseball is a bat-and-ball game played between two opposing teams who take turns batting and fielding. The game proceeds when a player on the fielding team, called the pitcher, throws a ball which a player on the batting team tries to hit with a bat. The objective of the offensive team (batting team) is to hit the ball into the field of play, allowing its players to run the bases, having them advance counter-clockwise around four bases to score what are called runs. The objective of the defensive team (fielding team) is to prevent batters from becoming runners, and to prevent runners advance around the bases. A run is scored when a runner legally advances around the bases in order and touches home plate (the place where the player started as a batter). The team that scores the most runs by the end of the game is the winner.

Hockey is a sport in which two teams play against each other by trying to manoeuvre a ball or a puck into the opponents goal using a hockey stick. There are many types of hockey such as bandy, field hockey, ice hockey and rink hockey.

Politics is a set of activities associated with the governance of a country, state or an area. It involves making decisions that apply to groups of members.

Medicine is the science and practice of establishing the diagnosis, prognosis, treatment, and prevention of disease. Medicine encompasses a variety of health care practices evolved to maintain and restore health by the prevention and treatment of illness. Contemporary medicine applies biomedical sciences, biomedical research, genetics, and medical technology to diagnose, treat, and prevent injury and disease, typically through pharmaceuticals or surgery, but also through therapies as diverse as psychotherapy, external splints and traction, medical devices, biologics, and ionizing radiation, amongst others.

Religion is a social-cultural system of designated behaviors and practices, morals, worldviews, texts, sanctified places, prophecies, ethics, or organizations, that relates humanity to supernatural, transcendental, or spiritual elements. However, there is no scholarly consensus over what precisely constitutes a religion.

For the yelp dataset, the description are the sentiment indicators (positive, negative). For the IMDB movie reviews, the description are (a good movie, a bad movie). For the aspect sentiment classification datasets, the description are the concatenation of aspect indicators and sentiment indicators. Aspect indicators for BeerAdvocate and TripAdvisor are re-
Although the course used in the movie was not the actual site of the 1913 US golf open and made a film that is about much more than an extraordinary game of golf. The film also deals directly with the class tensions of the early twentieth century and touches upon the profound anti-Catholic prejudices of both the British and American establishments. But at heart the film is about that perennial favourite of triumph against the odds. The acting is exemplary throughout. Stephen Dillane is excellent as usual, but the revelation of the movie is Shia LaBoeuf who delivers a disciplined, dignified and highly sympathetic performance as a working class Franco-Irish kid fighting his way through the prejudices of the New England WASP establishment. For those who are only familiar with his slap-stick performances in “Even Stevens” this demonstration of his maturity is a delightful surprise. And Josh Flitter as the ten year old caddy threatens to steal every scene in which he appears. A old fashioned movie in the best sense of the word: fine acting, clear directing and a great story that grips to the end - the final scene an affectionate nod to Casablanca is just one of the many pleasures that fill a great movie.

Pos_Tem: a good movie
Neg_Tem: a bad movie
Pos_Ex: fine acting, clear directing and a great story
Neg_Ex: dummy2
Pos_Abs: a great movie
Neg_Abs: a bad movie

dummy1 dummy2 As a recreational golfer with some knowledge of the sport's history, I was pleased with Disney's sensitivity to the issues of class in golf in the early twentieth century. The movie depicted well the psychological battles that Harry Vardon fought within himself, from his childhood trauma of being evicted to his own inability to break that glass ceiling that prevents him from being accepted as an equal in English golf society. Likewise, the young Ouimet goes through his own class struggles, being a mere caddie in the eyes of the upper crust Americans who scoff at his attempts to rise above his standing. What I loved best, however, is how this theme of class is manifested in the characters of Ouimet’s parents. His father is a working-class drone who sees the value of hard work but is intimidated by the upper class; his mother, however, recognizes her son’s talent and desire and encourages him to pursue his dream of competing against those who think he is inferior. Finally, the golf scenes are well photographed. Although the course used in the movie was not the actual site of the historical tournament, the little liberties taken by Disney do not detract from the beauty of the film. There’s one little Disney moment at the pool table; otherwise, the movie does not really think Disney. The ending, as in "Miracle," is not some Disney creation, but one that only human history could have written.

Pos_Tem: a good movie
Neg_Tem: a bad movie
Pos_Ex: I was pleased with Disney's sensitivity
Neg_Ex: dummy2
Pos_Abs: a good movie
Neg_Abs: a bad movie

dummy1 dummy2 This is an example of why the majority of action films are the same. Generic and boring, there’s really nothing worth watching here. A complete waste of the then barely-tapped talents of Ice-T and Ice Cube, who’ve each proven many times over that they are capable of acting, and acting well. Don’t bother with this one, go see New Jack City, Ricochet or watch New York Undercover for Ice-T, or Boyz n The Hood, Higher Learning or Friday for Ice Cube and see the real deal. Ice-T’s horribly cliched dialogue alone makes this film grate at the teeth, and I’m still wondering what the heck Bill Paxton was doing in this film? And why the heck does he always play the exact same character? From Aliens onward, every film I’ve seen with Bill Paxton has him playing the exact same irritating character, and at least in Aliens his character died, which made it somewhat gratifying...Overall, this is second-rate action trash. There are countless better films to see, and if you really want to see this one, watch Judgement Night, which is practically a carbon copy but has better acting and a better script. The only thing that made this at all worth watching was a decent hand on the camera - the cinematography was almost refreshing, which comes close to making up for the horrible film itself - but not quite. 4/10.

Pos_Tem: a good movie
Neg_Tem: a bad movie
Pos_Ex: dummy1
Neg_Ex: dummy2
Pos_Abs: a good movie
Neg_Abs: This is a generic and boring movie

dummy1 dummy2 This German horror film has to be one of the weirdest I have seen. I was not aware of any connection between child abuse and vampirism, but this is supposed based upon a true character. Our hero is deaf and mute as a result of repeated beatings at the hands of his father. he also has a doll fetish, but I cannot figure out where that came from. His co-workers find out and tease him terribly. During the day a mild-manner accountant, and at night he breaks into cemeteries and funeral homes and drinks the blood of dead girls. They are all attractive, of course, else we wouldn’t care about the fact that he usually tears their clothes down to the waist. He graduates eventually to actually killing, and that is what gets him caught. Like I said, a very strange movie that is dark and very slow as Werner Pochath never talks and just spends his time drinking blood.

Pos_Tem: a good movie
Neg_Tem: a bad movie
Pos_Ex: dummy1
Neg_Ex: dummy2
Pos_Abs: This German horror film has to be one of the weirdest I have seen
Neg_Abs: a good movie

dummy1 This German horror film has to be one of the weirdest I have seen

We list four input movie reviews from the IMDB datasets, with the gold label of the first one being positive and the second being negative. For the template strategy, the descriptions for the two classes (i.e., positive and negative) are always copied from templates, i.e., a good movie and a bad movie. For the extractive strategy, the extractive model is able to extract substrings of the input relevant to the golden label, and uses the dummy token as the description for the label that should not be assigned to the input. For the abstractive strategy, the model is able to generate descriptions tailored to both the input and the class. For labels that
should not be assigned to the class, the generative model outputs
the template descriptions. This is due to the fact that the generative
model is initialized using template descriptions. Due to the fact
that we incorporate the copy mechanism into the generation
model, the sequence generated by the abstractive model tend to
share words with the input document.

References

Del Corso, G. M., Gulli, A., and Romani, F. Ranking a stream of
news. In Proceedings of the 14th international conference on

Lei, T., Barzilay, R., and Jaakkola, T. Rationalizing neural predic-

Li, J., Monroe, W., and Jurafsky, D. Understanding neural
networks through representation erasure. arXiv preprint

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and
Potts, C. Learning word vectors for sentiment analysis. In Pro-
cedings of the 49th annual meeting of the association for com-
putational linguistics: Human language technologies-volume 1,

McAuley, J., Leskovec, J., and Jurafsky, D. Learning attitudes and
attributes from multi-aspect reviews. In 2012 IEEE 12th Inter-
national Conference on Data Mining, pp. 1020–1025. IEEE,
2012.

Yang, P., Sun, X., Li, W., Ma, S., Wu, W., and Wang, H. Sgm:
sequence generation model for multi-label classification. arXiv