G1 CAI SHIZHAN, SONG WENXIN

Topic: Disaster prediction of tweets with NLP

Report: 2

Difficult to follow along with analysis solely from the report. No quantifiable information is given for the training procedure (learning rate? number of epochs? which optimizer? Why?) Some concepts need to be more clearly defined or at least given an overview (BERT pretrain-finetune discrepancy). Some attention could be what distinguishes RoBERTa from BERT instead of just explaining BERT. In the conclusion paragraph, why does XLNet increase the amount of pre-training data? The team should also assess which procedures (eg autoregressive formulation in XLNet) could be applied to other datasets and problems.

Typos:

Figure 1 and Figure 2 labels are swapped (top 10 frequent keywords should be graph on right.) Some grammatical errors. RoBERTa paragraph: “it” is incorrectly capitalized.

Presentation: 4

Clear delivery and pacing from both speakers. Presentation managed to cover many points that were glossed over in the report. Good explanation of RoBERTa’s key differences from standard BERT. However, some typos in slides (3.9 “in essene”) and some slides (especially explanation of the attention mechanisms in section 3) could have less text and use point form. Also, speaking time between the two speakers is imbalanced.

Creativity: 3

Accuracy is good, but interpretability of the models used in the analysis leave much to be desired. No real change was proposed to the SOTA models. Perhaps analyzing common features of incorrect classifications could help the team further assess the strengths and limitations of the models they used and how they could be improved in the future.

Overall: 3

Confidence: 3

G5 PHAM TRUNG KIEN, DO VAN QUYET

Topic: Workers Supervision for Construction Safety

Report: 4.5

Report outline is clear. Methodology, data preprocessing and training process is clearly and succinctly described with relevant quantitative information (eg number of epochs, learning rate, etc.) There is a clear analysis of challenges faced and how the model could be improved (workers in low-light conditions not being detected.) There are some minor grammatical errors (mainly subject-verb agreement) are present in the text but they do not significantly impede readability, ex. in section 3.2, writing "EfficientNetV2 also surpass" instead of "EfficientNetV2 also surpasses." In section 2, "In fact, HABBOF has number of frames and people within a frame limited, also lack of challenging cases" is a bit unclear and could be rewritten as "Images in the HABBOF dataset have a small number of frames and a limited number of people within those frames, as well as a lack of challenging cases."

Presentation: 5

Good pacing and pronunciation from both speakers and content was well organized. Intuitive explanation of ad-hoc fisheye augmentation and how to construct large dataset (CEPDOF/HABBOF always no helmet/jacket, videos from construction sites always have helmet/jacket etc), as well as differences in what they changed between their midterm progress report and now.

Creativity: 5

End-to-end model results have good F-scores for both fisheye and non-fisheye case that matches the original hypothesis in the introduction section. The work proposes a high-performance solution to an intriguing and novel issue. I noticed that the loss in Figure 7 seems to converge in as little as around 40 epochs. Perhaps the team could use early stopping to prevent overfitting in the future?

Overall: 5

Confidence: 3

G7 SHAO ZHIHAO

Topic: Pawpularity prediction from images and tabular data

Report: 3

Report is well written with proper grammar and features a good amount of figures and EDA. However, report is lacking in detail (details are only in the presentation), eg justification for optimizer choice and learning rate schedule would be useful (why Cosine Annealing Warm Restarts?) A brief introduction into Swin Transformer in the report would help readers understand how it works and why it is a suitable model for this problem. Kaggle test score should also be included, and (if necessary) some analysis on why there is a significant discrepancy between CV RMSE and test RMSE. A section on "future analysis" (different model types? different augmentation scheme? etc) besides just the conclusion would help the presenter clearly demonstrate which ideas may or may not be of use in similar projects.

Presentation: 4

Good delivery and clear pronunciation. Strong explanation of Swin transformers (a model previously applied to NLPs) and its strengths and weaknesses compared to the standard CNN, and how self attention mechanism reduces computational complexity. Slides are clear and well elaborated. GradCam visualisation to explain why Swin predicts the way it does was very intuitive and easy to understand although it would help readers understand the effectiveness and limitations of Swin better if it were also on the report.

Creativity: 4

Used unique voting regression ensemble method and feature engineering methods (PCA to generate 16 features of high feature importance.) Kaggle test score and best CV RMSE score of 17.881 is a good result and the distribution of the predicted values via Swin does not seem to be as imbalanced as the training set distribution. SOTA model Swin was used but attempts could have been made to further improve it, or at least show what wouldn't work.

Overall: 4

Confidence: 3