STAT 517 – Literature Review Critique

Adhar Partap Singh: “Text plagiarism classification using syntax based linguistic features”

Reviewed by Gavin Whitesitt

Length of Paper: Adhar chose the 33 page paper “Text plagiarism classification using syntax based linguistic features” written by Vani K a and Deepa Gupta. The paper consists of 7 sections an Introduction, Existing systems, Proposed approach, Data statistics of the smaller and larger test conditions, Evaluation measures general and advanced classification metrics, Results and discussions, and Conclusions and future work. The paper proposes a new method to detect text plagiarism using syntax based linguistic characters through natural language processing. Adhar did a good job of explaining this in the introduction of his presentation as well as in his paper.

Depth of Understanding: After reading the paper and comparing it to Adhars review I think he does an overall good job of hitting the main points of the paper. I think there could be some improvement in delivering the content in a more palatable during the presentation though. I found myself caught up in the jargon. This isn’t to say Adhar didn’t explain the paper accurately, I just couldn’t quite keep up.

Difficulty: This paper was difficult for me to understand and took more time than I am proud to admit for me to grasp, but I don’t have any background in natural language processing so that may be why. I think that Adhar deserves credit for taking on this difficult paper.

Comprehensiveness: This paper is very technical. Adhar addressed about as much as possible given the length constraints of the paper. However I would have preferred a little bit more focus of the general mechanism of the classification methods and less space used explaining them ­-­­­­­­mathematically. This is just personal preference though, there is certainly an audience that would prefer more mathematical notation to less.

Clarity: Due to the difficulty of the paper I think this is a difficult metric to evaluate. I think that the writing is honest and explains what is going on in the paper about as well as the paper itself does. I do think there is room for improvement however as I found myself reading and rereading sections of the writeup to make sense of it. Again considering the difficulty of the paper this is only a minor gripe

Presentation: Overall, I thought the presentation was good. To improve upon it I think more stimulating visuals could be added. Additionally I think that some of the slides could be improved upon by condensing some of the text from big blocks to brief bullet points. I also would have benefitted from brief definitions of some of the vocab at the beginning of the presentation such as n-grams. I thought Adhar answered questions posed by Dr. Lee well, and was adequately prepared for the presentation.