Verb-Based Novelty Detection

Aurelio Arango,
Advisers: Lauren Stuart, Dr. Julia Taylor, and Dr. Victor Raskin
Purdue University West Lafayette, Indiana 47907

Introduction

Data is growing at very fast rates and there is a necessity to protect information. Also, there is a need to gather information on possible threats to society. Information Intelligence is essential for security. Focus on novelty is part of the process. So why focus on verbs only? There are not many verbs in a sentence. In an emergency, vast amounts of data might need to be processed and if a system only works with verbs, then processing time will be less than processing whole sentences. Also, once a flag has been triggered by the system an expert can determine if further investigation is needed.

Methodology

CNN Original Article Excerpt

1. Collection of phone records stirs debate: Valuable tool or 'beyond Orwellian'?
2. (CNN) -- Frightening government overreach or valuable law enforcement tool?
3. That's the question politicians in Washington, and millions of citizens around the United States, asked Thursday thanks to a jolting report suggesting the government has been collecting millions of Americans' phone records.
20. He further stressed the importance of ensuring "we have the tools we need to confront the threat posted by terrorists (and to) protect the homeland."
21. "That is his top priority," Earnest said of domestic security. "But ... we need to balance that priority with the need to protect civil liberties and constitutional rights of American people.

Find All verbs & Remove everything else

1. Stirs
2. Frightening, overreach
3. Is, asked, suggesting, has been collecting
20. Stressed, have, need, confront, posted, protect
21. Is, said, need, balance, protect

Unique, Recurring, and Synonyms Verbs

1. Stirs
2. Frightening, overreach
3. Is, asked, suggesting, has been collecting
20. Stressed, have, need, confront, posted, protect
21. Is, said, need, balance, protect

Analyzed Text

1. Collection of phone records stirs debate: Valuable tool or 'beyond Orwellian'?
2. (CNN) -- Frightening government overreach or valuable law enforcement tool?
3. That is the question politicians in Washington, and millions of citizens around the United States, asked Thursday thanks to a jolting report suggesting the government has been collecting millions of Americans' phone records.
20. He further stressed the importance of ensuring "we have the tools we need to confront the threat posted by terrorists (and to) protect the homeland."
21. "That is his top priority," Earnest said of domestic security. "But ... we need to balance that priority with the need to protect civil liberties and constitutional rights of American people

Results

Table 1 provides information on the number of verbs in the article in relation to the total words in the article. In these two articles "a," "an," and "the" were removed since those three elements did not provide novel information. Table 2 shows the categories in which some verbs fall under, such as unique verbs, verb's novelty, and synonyms. Table 2 also keeps count of those categories. Additionally, verbs might fall under a general category, such as "inform," or "security" and Table 2 keeps count of the different categories.

Article	FOX	CNN
Fillers	275	572
Total verbs	168	335
Total words (No nouns)	443	907

Table1

Article	FOX	CNN
Unique Verbs	70	101
Verbs with same meaning	22	42
General Categories	10	28
Novelty	48	60
New verb but not concept	22	41
Total Verbs	168	335

Table 2
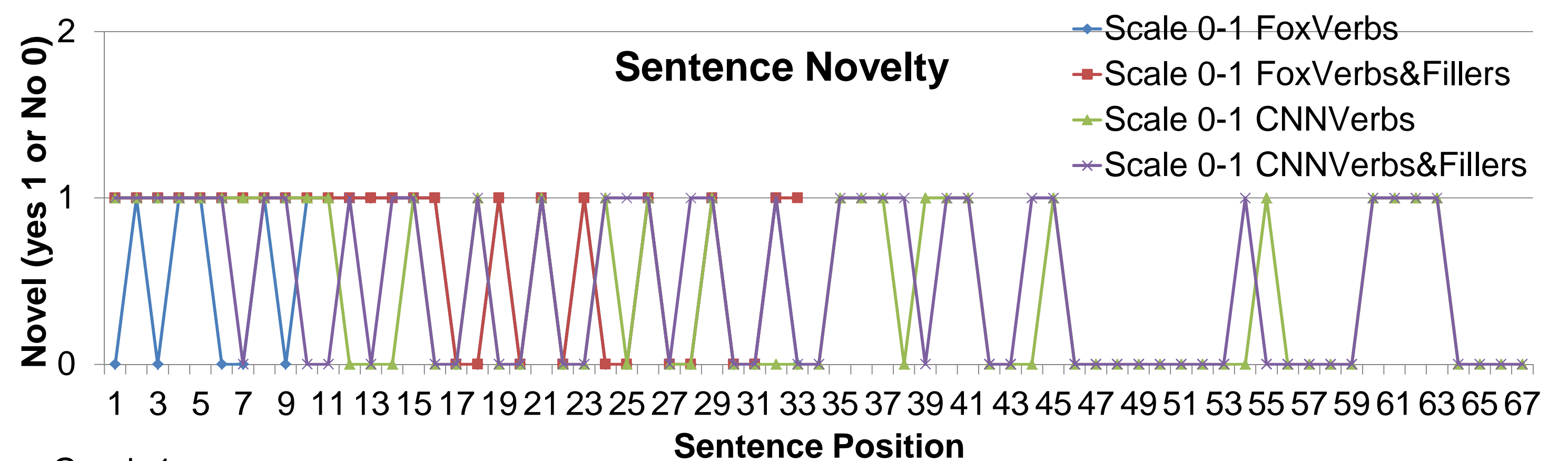
Graph 1
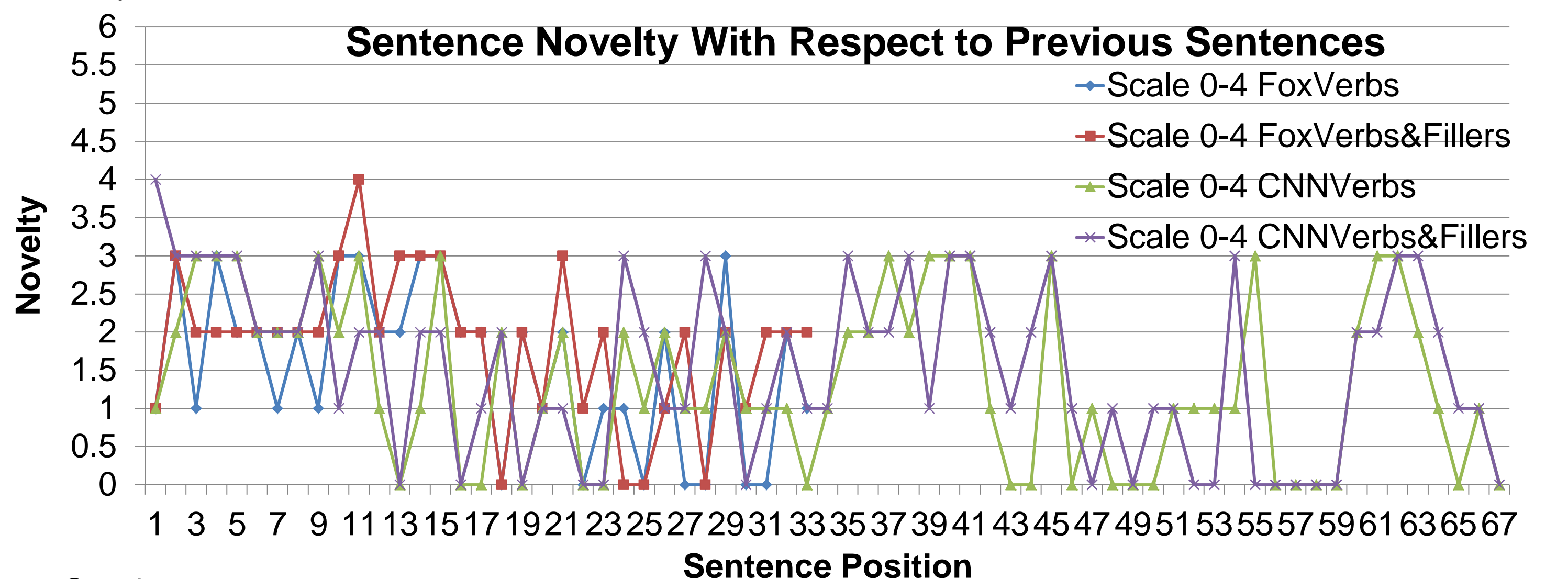
Graph 2

Graph 1) This graph represents novelty with respect to previous sentences yes(1) or no(0). One line focuses on only verbs (diamond or triangle marker). The second line focuses on verbs and fillers (square or x maker). Graph 2) This graph illustrates novelty at each sentence with respect to previous sentences. Zero indicates no novelty and a four indicates very novel.

Filler: All elements of a sentence not containing verbs, nouns, "a," "an," and "the."

Conclusion

There is not enough evidence to support the claim that verb-based novelty detection contributes enough information. Table 2 shows the different general concepts found in the articles and in a verb-based novelty detection system those concepts can lead to different interpretations. Contribution of other elements that are part of the sentence helped define the subject. High peaks in the second graph are indicative of novelty and most of the novelty can be found at the beginning of the text. An observation worth noting is that there are relations between words that contribute enough information to know about the theme of the text.