A Hedging Index: Exploring the Confidence in Radiology Reports using Natural Language Understanding

Niketh Bandla, MD, VA Baltimore Medical Center, University of Maryland Medical Center; Eliot L. Siegel, MD, FSII; Gilan E. Saadawi, MD, PhD, MS

Hypothesis

An Natural Language Understanding (NLU) algorithm can be trained to recognize various forms of uncertain terminology in Radiology text reports including the ‘hedge.’

Purpose

Given that previous studies have explored the effectiveness of radiology reports at conveying information, hedging remains an elusive topic without standardization in its usage. Other radiologists have taken a stance against vague and ambiguous reports in general. Many authors have argued against the practice of hedging i.e, an attempt to protect oneself from the risk of commitment, although definitions and examples of this term lack consensus.

The Oxford definition of the word ‘hedge’ is “A way of protecting oneself against financial loss or other adverse circumstances” [1]. In relation to Radiology specifically, Ferris M. Hall defined a hedge as “an evasive statement to avoid the risk of commitment” [2]. Wallis, A. recommends using a standard where if the radiologist is 90% certain about a diagnosis, to completely avoid any ambiguity and hedge terms such as “no fracture seen” or “no pneumonia identified”[3]. On the other hand, Berlin used a definition which was “the making of calculatedly noncommittal or ambiguous statements”[4]. More important than the definition of the word however, is the impact of hedging on the profession of Radiology and the attitudes of referring clinicians on practice. Avoiding hedging terminology and focusing on clear and accurate communication is an important factor in risk management and malpractice defense in Radiology [4]. Vague and ambiguous terminology can be used against the radiologist in the court of law similar to absent communication[5].

Clinical context also plays a large role in whether a hedge is appropriate, for example Hall also wrote that a phrase such as “no fracture is identified”, implying that a fracture may have been missed, can sometimes be acceptable. This type of wording is appropriate in certain difficult to assess radiographs such as “ribs or externally rotated hips in osteoporotic women”, but may be inappropriate phrasing when looking at radiographs of long bones[6].

The purpose of this study was to determine (1) the extent to which radiologists agree on what constitutes hedging in a radiology report and (2) whether a computer using natural language understanding (NLU) software can be “taught” to recognize hedging which is predictive of the consensus of multiple radiologists.

Methods & Materials
One thousand radiology reports from multiple imaging modalities were randomly selected, which were annotated by human observers who were instructed to determine whether there was “hedging” in these reports and the results were stored in a database. The test set of 1000 reports was used to refine a NLU algorithm to categorize reports into groups that contain (1) uncertain concepts, (2) hedged concepts and (3) remote concepts. ‘Remote’ was defined as a type of negative hedge, an example being “pneumothorax not seen” versus “apparent pneumothorax” which would fall into the ‘hedge’ category. The initial performance of the algorithm in appropriately segregating the reports into these categories was tested against human observation of hedging. The NLU requires a set of baseline definitions prior to validation with human input. Uncertainty was commonly shown using terms such as “possible/probably”, “may”, “suspect”, and “likely/unlikely”. Some reports contained hedged concepts in the body, but not in the impression of the text report. In this situation, the report was not counted to have hedging terminology because the impression had full clinical certainty. If the impression of the report referred fully to the body, for example “see above” or “refer to findings”, the body of the report would be analyzed for hedging terminology.

Results

The reports were tested by the algorithm to see if they can fall into any of the previously defined categories. To test its initial performance, the results were validated by humans who analyzed all of the reports as well. There were a total of 173 reports identified with some concepts ‘hedged.’ The human testers disagreed with the algorithm in 21 reports, and agreed on the remaining 152. In the next category, there were a total of 261 reports with some concepts uncertain, with the human testers only disagreeing on 3 reports. In the last category there were 308 reports total with some concepts ‘remote’ and there was disagreement with the human testers in 8 reports.

Discussion

Clinical uncertainty and hedging terminology are historically challenging to define and detect. There was some overlap in the findings, since a single report could have both hedged and uncertain concepts.

Language used in this dataset of Radiology text reports often contained concepts that are either uncertain or hedged. The purpose was not to detect the accuracy of reports but rather the clarity by analyzing the type of language used in a given clinical context. Out of the 1000 reports analyzed by NLU technology, nearly one third of the reports contained negatively hedged concepts. These negative hedges were defined as ‘remote’ concepts, making this the most common type of hedge within the dataset. The NLU algorithm recognized ‘remote’ concepts by identifying examples such as “no definite pneumothorax”, “no evidence of pneumothorax”, and “no pneumothorax seen”. Some examples of positive hedge terms identified by the NLU algorithm include “appears to be a pneumothorax” or “pneumothorax seen”. Positively hedged clinical findings were the least common of the given categories.

Uncertainty is also common within the radiology text reports given that over a quarter of the reports in the dataset contained uncertain concepts. Some of the reports which contained full certainty were due to an impression that did not have any ambiguity, even though the body of the report could have still contained these concepts. It is possible that the frequency of hedged, remote, and uncertain concepts was actually higher than reported by the algorithm due to reports having firm and unambiguous impressions. The tolerance for ‘hedging’ terminology in the body versus the impression is also another area that can be explored. For example, Rothman contended that the body of the report is focused on the subjective experience of the Radiologist, where words ‘apparent, seen, identified’ are acceptable and common, however should not be present in the impression[7]. Since the body of the report is
traditionally used as a way to specifically describe what the radiologist sees rather than making strict diagnosis, this section is naturally prone to ambiguous terminology.

Conclusion

High quality radiology reports should clearly convey level of certainty in making findings while minimizing hedging. Given the challenge Radiologists have at making clinical diagnoses in the midst of patient complexity combined with the subjective nature of radiological findings, it can be difficult to communicate with clarity within the text report. These results show that a significant number of reports contain variants of ‘hedging terminology’ and NLU technology can serve as a quality control measure to track this type of communication. Making a report discretely intelligible and determining report quality requires more than just a semantic “understanding” of the terms and concepts but also requires evaluation of level of certainty and the use of hedging. NLU technology can be used to identify the occurrence of uncertainty and hedging in radiology reports, but when a continuous rigorous validation process is in place. Ultimately an index can be developed unique to each radiologist that can assess the number of reports that were hedged or ambiguous over a given period of time. This would however require a fully agreed upon definition of these concepts while also evaluating in which clinical circumstances this type of language should be avoided.

Figure 1

<table>
<thead>
<tr>
<th>Algorithm findings, # of Radiology text reports</th>
<th>Agree</th>
<th>Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some concepts 'hedged'</td>
<td>152</td>
<td>21</td>
<td>173</td>
</tr>
<tr>
<td>Some concepts uncertain</td>
<td>258</td>
<td>3</td>
<td>261</td>
</tr>
<tr>
<td>Some concepts 'remote'</td>
<td>300</td>
<td>8</td>
<td>308</td>
</tr>
</tbody>
</table>

References

Keywords

Natural Language Understanding, NLU, communication, clarity, informatics