Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 3109–3113
Marseille, 11–16 May 2020

A Lexicon-Based Approach for Detecting Hedges in Informal Text

Jumayel Islam[1], **Lu Xiao**[2], **Robert E. Mercer**[1]
[1]Department of Computer Science, The University of Western Ontario
[2]School of Information Studies, Syracuse University
jislam3@uwo.ca, lxiao04@syr.edu, mercer@csd.uwo.ca

Abstract

Hedging is a commonly used strategy in conversational management to show the speaker's lack of commitment to what they communicate, which may signal problems between the speakers. Our project is interested in examining the presence of hedging words and phrases in identifying the tension between an interviewer and interviewee during a survivor interview. While there have been studies on hedging detection in the natural language processing literature, all existing work has focused on structured texts and formal communications. Our project thus investigated a corpus of eight unstructured conversational interviews about the Rwanda Genocide and identified hedging patterns in the interviewees' responses. Our work produced three manually constructed lists of hedge words, booster words, and hedging phrases. Leveraging these lexicons, we developed a rule-based algorithm that detects sentence-level hedges in informal conversations such as survivor interviews. Our work also produced a dataset of 3000 sentences having the categories Hedge and Non-hedge annotated by three researchers. With experiments on this annotated dataset, we verify the efficacy of our proposed algorithm. Our work contributes to the further development of tools that identify hedges from informal conversations and discussions.

Keywords: Hedging, Informal conversation, Discourse Markers

1. Introduction

People use hedging when they try to avoid criticism or evade questions in conversations (Crystal, 1988). Using hedging gives interviewees an opportunity to organize their thoughts and make a suitable response in a one-to-one interview. This is frequent when there is a disjuncture between the interviewer and the interviewee. Layman (2009) showed how interviewees employ such strategies during an oral history interview to avoid answering sensitive questions. Most of the time, in such cases, their responses are dismissive and filled with hedging. This leads to interviewers' judgment whether to press the interviewee when it becomes evident that the interviewee is reluctant to answer certain questions. Thus, it is important to analyze such phenomenon and build tools that can identify hedging in interview transcripts which will help researchers to understand the dynamics of such interviews and give them more control of the situation. The following two examples from a conversational interview transcript demonstrate the use of hedging for these purposes:

> (1) *I think* that every survivor's story must be heard in the singularity of experience that it recounts.
> (2) *I don't know* if I *would* address the young survivors specifically.

The use of hedge terms *"I think"*, *"I don't know"* and *"would"* demonstrates the instability in their narrative.

Besides hedge words, people use discourse markers to hedge in conversations. These can be an utterance or a word or a phrase (such as *"oh"*, *"like"*, *"well"*, and *"you know"*) that either direct or redirect the flow of conversation without adding any significant meaning to the discourse (Schiffrin, 1987). For example, the discourse marker *"well"* has been shown to serve various purposes in conversations, such as delaying the response, mitigating the face threat, and marking insufficiency (Jucker, 1993; Ponterotto, 2018). People use it to show a slight change in topic, or when what they want to say is not quite what is expected, or as a pause filler in the face of an interactive difficulty, etc.. For example, "I think..., *well*, I've never compared them, it's a bit of a difficult question, but I think it's different." Layman (2009) discussed how necessary it is to be conscious of these circumstances so that the interviewer can better judge whether the interviewee should be questioned. For example, the use of discourse markers such as *'not really'*, *'not that I remember'* or *'well, anyway'* in responses shows how hedging through the usage of discourse markers in an interview might be influential.

In this study, we are interested in examining the presence of hedging words and phrases in identifying the tension between an interviewer and interviewee. Usually, this involves moments when the interviewer wants the conversation to go in one direction but the survivor either doesn't want to go 'there' (deflection) or wants to go in another direction (booster). It also includes moments of outright, though often subtle, disagreement (Ahn, 2010). While there have been studies on hedging detection in the natural language processing literature, all existing work has focused on structured texts and formal communications. For example, the CoNLL 2010 shared task (Farkas et al., 2010) completely focused on hedging in articles such as scientific papers and Wikipedia articles. Formal and informal language each serve a different purpose. Depending on the situation and the formality, the choice of words, the sound and how each word is put together will differ. Informal language is much more spontaneous and casual than formal language. It allows for the display of emotion or empathy. The first example below represents how hedging is used in formal text, whereas the second example shows how people use hedging in informal communication.

> (1) Most people *think* that dogs are smarter because they are loyal and make humans feel like

the center of the universe.
(2) That's a shame, I *think*, it truly is unfortunate, because many resources are left unused in this way.

Hence, we investigated a corpus of eight unstructured conversational interviews about the Rwanda Genocide and identified hedging patterns in the interviewees' responses, as a first step towards a computational tool that automatically identifies hedging in unstructured and informal communications. Specifically, we constructed three lists[1] of hedge words, booster words, and hedging phrases, and developed a rule-based algorithm that detects sentence-level hedges in these informal conversations with these lexicons. In the rest of the paper, we first review the related studies about hedging detection. We then present our method as well as the experiment that compared the performance of our approach against the annotations provided by three researchers.

2. Related Work

Light et al. (2004) constructed a dictionary of hedge cues to identify speculative (hedged) sentences in MEDLINE abstracts. They also used a Support Vector Machine (SVM) as a classifier to determine speculative sentences in the abstracts. Medlock and Briscoe (2007) treated the problem of determining speculative sentences as a classification task. Their training samples were collected from biomedical articles. They used single words as feature for their model. Szarvas (2008) used the same dataset, but they used bigrams and trigrams instead as features for their maximum entropy model classifier. Ganter and Strube (2009) proposed a hedge detection system based on word frequency measures and syntactic patterns of the weasel words in Wikipedia articles. In Özgür (2009)'s supervised learning approach, they used various features such as keywords, positional information of the keywords, and the contextual information of the keywords. They also used syntactic structures of the sentence to determine the scope of the hedge cues. Agarwal and Yu (2010) used a conditional random field (CRF) algorithm to train models in order to identify hedge cue phrases in biological literature. They performed experiments on the BioScope corpus (Szarvas et al., 2008) and showed the efficacy of their model in the biological domain.

The problem of detecting hedges was addressed in the CoNLL 2010 shared task (Farkas et al., 2010). However, the datasets that have been used contain only formal texts though. More recently, Ulinski et al. (2018) proposed a set of manually constructed rules which allowed them to identify hedged sentences in forum posts in an unsupervised manner. Theil et al. (2018) expanded a lexicon of uncertainty trigger words utilizing domain specific word-embedding models and used TF-IDF (Term Frequency - Inverse Document Frequency) for representing features. Their extended lexicon improved the performance of uncertainty detection significantly in financial domain when used with machine learning models. Ponterotto (2018) discussed different hedging strategies that have been employed by Barack Obama, the former president of the United States, in political interviews. Through defining hedging-related discursive approaches, they provided a thorough analysis of the president's responses. They discussed hesitation strategies, such as, pauses and repairs *(yes, no)*, restarts *(I won't ... I won't say)* and discourse markers *(anyhow, anyway, I mean)*.

Our review of the relevant literature suggests that most of the works that have been done so far in identifying hedging is on formal communication or structured text. For our work, we focus on identifying such phenomenon in unstructured conversations.

3. Methodology

Algorithm 1 shows the pseudo-code of our rule-based hedging detection algorithm. The algorithm leverages lexicons we compiled for hedge words, discourse markers and booster words.

We used Jaccard distance, complementary to the Jaccard index, to measure the similarity between the discourse markers of our lexicon and phrases from the input sentences. The lower the distance, the more similar the two strings.

Algorithm 1 Hedge Detection Algorithm

```
function isTrueHedgeTerm(t)
    Rules to disambiguate hedge terms
    if t is true hedge term then
        return True
    end if
    return False
end function
function isHedgedSentence(s)
    DM ← List of discourse markers
    HG ← List of hedge words
    P ← List of n-grams from s
    B ← List of booster words
    status = False
    JD ← Jaccard Distance
    for A in DM do
        for B in P do
            if 1 - JD(A,B) ≥ threshold then
                status = True
            end if
        end for
    end for
    for hedge in HG do
        if hedge in s AND isTrueHedgeTerm(hedge) then
            status = True
        end if
    end for
    for booster in B do
        if booster in s and booster is preceded by not or without then
            status = True
        end if
    end for
    return status
end function
```

[1] https://github.com/hedging-lrec/resources

We used this measure, shown in Equation 1, in our hedge detection algorithm.

$$d_J(P,Q) = 1 - J(P,Q) = \frac{|P \cup Q| - |P \cap Q|}{|P \cup Q|} \quad (1)$$

Here, $d_J(P,Q)$ represents Jaccard distance between P and Q, where P and Q are sets of words. We experimented with a number of thresholds between 0 and 1 in order to decide whether a discourse marker is similar enough with a phrase. We achieved the highest F1-scores when the threshold was set between 0.78 and 0.81.

Hedge words. We compiled a list of 76 potential hedge words. Words that reflect the speaker/writer's mental state or internal actions are known as epistemic words. We included different epistemic words in our hedge words lexicon that show their hedging act, such as verbs *(suppose, think, presume)*, adverbs *(arguably, barely, seemingly)*, adjectives *(unlikely, unsure, unclear)* and modal verbs *(might, maybe)*. With epistemic modality, a speaker's level of confidence on his/her proposition can be determined. We also included various approximators such as *(generally, usually)* in the lexicon.

Hedge words that are composed of multiple words are simply called multi-word hedges. For example, the sentence *"In my view*, this attitude produced through social discourse can also change things within families." shows how the multi-word hedge can be used during conversations. The phrase *"in my view"* acts as an important indicator of hedging here. The words *"in"*, *"my"* and *"view"* though can not show any hedging when used independently.

Discourse markers. As discussed in the introduction section, people also use discourse markers when hedging in conversations. These markers have a variety of functions. For example, when making an unexpected contrast (*even though; despite the fact that*), making a contrast between two separate things, people, ideas, etc. (*anyway; however; rather)*, clarifying and re-stating (*in other words; in a sense; I mean*), to change topic or return to the topic (*well, anyway*) or indicating a difference of opinion (*yes, but*). We constructed a list of such discourse markers.

Boosting words. Boosting, using terms such as *absolutely*, *clearly* and *obviously*, is a communicative strategy for expressing a firm commitment to statements. Holmes (Holmes, 1984) provides an early definition of boosting. According to him, "Boosting involves expressing degrees of commitment or seriousness of intention (p. 347)". It allows speakers to express their proposition with confidence and shows their commitment to statements. It also restricts the negotiating space available to the hearer. Boosting plays a vital role in creating conversational solidarity (Holmes, 1984) and in constructing an authoritative persona in interviews (He, 1993). Interestingly, if booster words are preceded by negation words such as *"not"*, or *"without"*, they can act as hedges. For example, "I'm still *not sure* if I would go back, I don't know what it would be like." Here, *"sure"* is a booster word. However, since it is preceded by a negation word *"not"*, it changes the meaning completely. We handle this kind of situation in our proposed algorithm by compiling and including a list of booster words in the algorithm.

Rules for Disambiguation. Hedging disambiguation is an important part of our algorithm, as some commonly used hedge terms in the conversational interviews have non-hedge senses as well. We apply rules to disambiguate these terms based on the syntactic structure of the sentences. Our rules are an extension and modification of the set of rules proposed by (Ulinski et al., 2018). We used the Stanford CoreNLP (Manning et al., 2014) parser to parse the sentences[2]. What follows is a brief analysis of some of the rules used in our study with examples derived from our interview datasets.

Hedge Term: Feel, Suggest, Believe, Consider, Doubt, Guess, Hope
Rule: If token *t* is (i) a *root* word, (ii) has the part-of-speech *VB** and (iii) has an *nsubj* (nominal subject) dependency with the dependent token being a first person pronoun *(i, we)*, *t* is a hedge, otherwise, it is a non-hedge.
Hedge: I don't think it's been a failure, but I **hope** that I'm on the right track.
Non-hedge: I'm still living with it, but without **hope** that I would find anyone.

Hedge Term: Think
Rule: If token *t* is followed by a token with part-of-speech *IN*, *t* is a non-hedge, otherwise, hedge.
Hedge: I **think** it's difficult to make generalizations about this kind of relationships.
Non-hedge: Even if it's difficult, I always say, **think** about your children.

Hedge Term: Assume
Rule: If token *t* has a *ccomp* (clausal complement) dependent, *t* is a hedge, otherwise, non-hedge.
Hedge: I **assume** they were responsible for this.
Non-hedge: They have **assumed** the role of parents and are doing their best to fulfill it.

Hedge Term: Suppose
Rule: If token *t* has an *xcomp* (open clausal complement) dependent *d* and *d* has a mark dependent *to*, *t* is a non-hedge, otherwise, it is a hedge.
Hedge: I **suppose** he was present during the discussion.
Non-hedge: I could see that they were skewing the real truth, the one they are **supposed** to tell me.

Hedge Term: Tend
Rule: If token *t* has an *xcomp* (open clausal complement) dependent, *t* is a hedge, otherwise, it is a non-hedge.
Hedge: We **tend** to never forget.
Non-hedge: All political institutions **tended** toward despotism.

Hedge Term: Appear
Rule: If token *t* has a *ccomp* (clausal complement) or *xcomp* (open clausal complement) dependent, *t* is a hedge, otherwise, it is a non-hedge.
Hedge: It **appears** that there were people who wanted to attack the school and that the nuns knew that beforehand.
Non-hedge: I had to do all I could to **appear** like an old lady, like someone who has no life, someone of no interest to you.

Hedge Term: Likely

[2]https://stanfordnlp.github.io/CoreNLP/download.html

Rule: If token *t* has relation amid with its head *h* and *h* has part of speech *N**, *t* is a non-hedge, otherwise, it is a hedge.
Hedge: They will **likely** visit us in the future.
Non-hedge: He is a fine, **likely** young man.
Hedge Term: Should
Rule: If token *t* has relation *aux* with its head *h* and *h* has dependent *have*, *t* is a non-hedge, otherwise, it is a hedge.
Hedge: That's precisely the message that **should** be sent to people who label others, isn't it?
Non-hedge: They **should** have been more careful.
Hedge Term: Rather
Rule: If token *t* is followed by token *than*, *t* is a non-hedge, otherwise, it is a hedge.
Hedge: I never had the opportunity to go, but i know people who have gone and who came back **rather** depressed.
Non-hedge: He would have protected his flock **rather** than shoot at them.

4. Experiments

4.1. Data

We collected our data from the living archives of Rwandan exiles and genocide survivors in Canada[3]. The life story interviews, which vary in duration from ninety minutes to twelve hours, were recorded between 2007 and 2012, by the Montreal Life Stories project, a COHDS-based partnership project that recorded 500 life stories of Montrealers displaced by war, genocide and other human rights violations. The digital repository contains those life stories of Rwandan genocide survivors and has been made publicly accessible for the researchers. In this study, we worked with eight transcribed interviews which are translated into English.

One of the main limitations in this study is the lack of readily available annotated data for model evaluation. In order to rectify this limitation, we randomly collected 3,000 sentences from the translated interviews and three researchers annotated the sentences as either hedged or non-hedged sentences independently. We understand that the amount of annotated data is not huge. However, it needs to be realized that we are constrained by resources and the process of annotation is time consuming.

4.2. Results

As the existing hedging detection techniques have focused on structured texts and formal communications, we anticipated that they would not perform well with informal conversations and discussions. Hence, we took the effort of developing this rule-based algorithm. The data used for our experiments consists of samples from eight annotated interview transcripts. We compared the performance of our algorithm with the annotated dataset through two different tasks. In the first task, we used a voting system to determine the final annotation of a sentence. If at least two out of the three annotators agreed on a label, we picked that label as the final label for that particular sentence. This produced 247 hedged sentences and 2,753 non-hedged sentences. With this "gold standard", we calculated the precision, recall, and F1-score of our algorithm. As we can see in Table 1, our algorithm achieved a precision of 32.1%, a recall of 88.3% and a F1-score of 47.0% for the category Hedge and a precision of 98.7%, a recall of 83.2% and a F1-score of 90.3% for the category Non-hedge.

	P	R	F1
Hedge	32.1	88.3	47.0
Non-hedge	98.7	83.2	90.3
Avg	65.4	85.8	68.7

Table 1: Results (in %) of our hedge detection algorithm in comparison with the 3 annotations where we used majority voting to finalize the label.

Acknowledging the fact that, the concept of hedging is subjective, we considered another scenario in the comparison - a sentence is tagged as hedged if any of the researchers annotated it as such. This produced 604 hedged sentences and 2,396 non-hedged sentences. Our interest is to examine whether the algorithm is able to identify all possible hedged sentences from the dataset. As we can see in Table 2, our tool achieved a precision of 57.6%, a recall of 64.9% and a F1-score of 61.0% for the category Hedge and a precision of 90.9%, a recall of 87.9% and a F1-score of 89.4% for the category Non-hedge.

Our dataset is very imbalanced, that is, there are over ten times more non-hedged sentences than hedged sentences. This affected the performance greatly, especially the precision of non-hedged sentences. The relatively much higher recall of the hedge sentences implies that when people hedge these lexicons cover the language features they tend to use when hedge. On the other hand, the low precision of the hedged sentences, in both evaluation measures (Table 1 and Table 2), suggests that in this one-to-one unstructured interview context, the use of discourse markers or hedge words may be for purposes other than hedging and cue disambiguation is a reasonable next step to improve the performance.

	P	R	F1
Hedge	57.6	64.9	61.0
Non-hedge	90.9	87.9	89.4
Avg	74.3	76.4	75.2

Table 2: Results (in %) of our hedge detection algorithm in comparison with the 3 annotations where we provide the label "Hedge" for a sentence if at least one of the annotators tagged the sentence having that category.

5. Conclusion and Future Work

Hedging plays an important part in conversational management. To help identify hedging in informal communications, we constructed lexicons for hedge words, discourse markers and booster words. We also discussed rules to handle ambiguous hedge terms. With the 3,000 annotated sentences, we evaluated the performance of our hedging detection tool. We will also annotate more sentences in informal conversations for further testing and validation. We also

[3]http://livingarchivesvivantes.org/

plan to explore the use of the hedging discourse markers in various contexts. For instance, we expect that hedging indicators in discussions of political issues are different from that of trip planning. With our open sourced lexicons, we contribute to studies that need to detect hedges in their data and for computer scientists in the development of hedging detection techniques for unstructured texts and informal communications.

6. Bibliographical References

Agarwal, S. and Yu, H. (2010). Detecting hedge cues and their scope in biomedical text with conditional random fields. *Journal of Biomedical Informatics*, 43(6):953–961.

Ahn, J. J. (2010). *Exile as forced migrations: A sociological, literary, and theological approach on the displacement and resettlement of the Southern Kingdom of Judah*, volume 417. Walter de Gruyter.

Crystal, D. (1988). On keeping one's hedges in order. *English Today*, 4(3):46–47.

Farkas, R., Vincze, V., Móra, G., Csirik, J., and Szarvas, G. (2010). The conll-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*, pages 1–12. Association for Computational Linguistics.

Ganter, V. and Strube, M. (2009). Finding hedges by chasing weasels: Hedge detection using wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 173–176. Association for Computational Linguistics.

He, A. W. (1993). Exploring modality in institutional interactions: Cases from academic counselling encounters. *Text-Interdisciplinary Journal for the Study of Discourse*, 13(4):503–528.

Holmes, J. (1984). Modifying illocutionary force. *Journal of Pragmatics*, 8(3):345–365.

Jucker, A. H. (1993). The discourse marker well: A relevance-theoretical account. *Journal of Pragmatics*, 19(5):435–452.

Layman, L. (2009). Reticence in oral history interviews. *The Oral History Review*, 36(2):207–230.

Light, M., Qiu, X. Y., and Srinivasan, P. (2004). The language of bioscience: Facts, speculations, and statements in between. In *HLT-NAACL 2004 Workshop: Linking Biological Literature, Ontologies and Databases*.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Medlock, B. and Briscoe, T. (2007). Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 992–999.

Özgür, A. and Radev, D. R. (2009). Detecting speculations and their scopes in scientific text. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1398–1407. Association for Computational Linguistics.

Ponterotto, D. (2018). Hedging in political interviewing. *Pragmatics and Society*, 9(2):175–207.

Schiffrin, D. (1987). *Discourse markers (Studies in Interactional Sociolinguistics, 5)*. Cambridge: Cambridge University Press.

Szarvas, G., Vincze, V., Farkas, R., and Csirik, J. (2008). The bioscope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 38–45. Association for Computational Linguistics.

Szarvas, G. (2008). Hedge classification in biomedical texts with a weakly supervised selection of keywords. *Proceedings of ACL-08: HLT*, pages 281–289.

Theil, C. K., Stajner, S., and Stuckenschmidt, H. (2018). Word embeddings-based uncertainty detection in financial disclosures. In *Proceedings of the First Workshop on Economics and Natural Language Processing*, pages 32–37.

Ulinski, M., Benjamin, S., and Hirschberg, J. (2018). Using hedge detection to improve committed belief tagging. In *Proceedings of the Workshop on Computational Semantics beyond Events and Roles*, pages 1–5.