Chapter 6
Detecting Uncertainty in Spoken Dialogues: An explorative research for the automatic detection of speaker uncertainty by using prosodic markers

Jeroen Dral, Dirk Heylen and Rieks op den Akker

Abstract This paper reports results in automatic detection of speaker uncertainty in spoken dialogues by using prosodic markers. For this purpose a substantial part of the AMI corpus (a multi-modal multi-party meeting corpus) has been selected and converted to a suitable format so its data could be analyzed for a selected set of prosodic features. In the absence of relevant stance annotations on (un)certainty, lexical markers (hedges) have been used to mark utterances as (un)certain. Results show that prosodic features can indeed be used to detect speaker uncertainty in spoken dialogues. The classifiers can tell uncertain from neutral utterances with an accuracy of 75% which is 25% over the baseline.

6.1 Introduction

Each utterance we make comes with a particular degree of certainty we have about the state of affairs that is described in our utterance. We may feel reasonably confident or rather hesitant about whether there is any truth in what we are saying. We often express this degree of certainty in what we are saying through hedges ("I think"), modal verbs ("might"), adverbs ("probably"), tone of voice, intonation, hesitations. Our speech may be co-occur with gestures and facial expressions that can express the same hesitant or confident state of mind. This research will focus on the prosodic features of speech and will try to develop a method to automatically classify speech as being (un)certain. The purpose of this research is to (automatically) measure one's belief (or confidence or self-conviction) in the correctness of

Jeroen Dral
University of Twente, Enschede, the Netherlands, e-mail: j.dral@student.utwente.nl

Dirk Heylen
University of Twente, Enschede, the Netherlands, e-mail: heylen@ewi.utwente.nl

Rieks op den Akker
University of Twente, Enschede, the Netherlands, e-mail: infrieks@ewi.utwente.nl

a certain utterance. Even when the definition of uncertainty is clear, a number of questions can still be posed: how to state the degree of uncertainty? Is it certain or uncertain or are there shades of gray in between? And if so, how do we state them?

6.2 Related Work

Although uncertainty can be detected by both visual and non-visual means, this research, and the overview of the related work, will focus on the non-visual aspects of the detection of (un)certainty.

6.2.1 Defining (un)certainty

People's ability to accurately assess and monitor their own knowledge has been called the 'feeling of knowing' (FOK) by Hart [6]. Many experiments on this area are based on question-answering where respondents must answer certain (knowledge) questions and assess whether their answer is likely to be correct. A study by Smith and Clark [12] investigated FOK in a conversational setting and followed the method mentioned above. Respondents were asked to answer general knowledge questions, then estimated their FOK about these questions and finally were tested on their ability to recognize the correct answer. They found that FOK was positively correlated with recognition and with response latency when retrieval failed and negatively correlated when retrieval succeeded.

Another study by Brennan and Williams [5] used the research of Smith and Clark and in addition researched the sensitivity of listeners to the intonation of answers, latencies to responses and the form of non-answers. When looking at the 'feeling of another's knowing' or FOAK, Brennan and Williams state a listener can use several different sources of information to evaluate a respondent's knowledge:

- His/her own knowledge
- Assess the difficulty of the question for the average person or for the typical member of a particular community and use that information to judge a respondent's confidence.
- Information from their shared physical environment and from immediately previous conversation "mutual knowledge")
- Information about the respondent's ability or previous performance
- Paralinguistic information displayed in the surface features of respondent's responses (intonation, latency to response)

In their experiments Brennan and Williams concentrated on the paralinguistic information available. The result of their experiments supports the interactive model of question-answering and perhaps help in understanding respondent's metacognitive states when searching their memories for an answer. Brennan and Willaims' work also demonstrates the ability of listeners to use these cues. Their FOAK was affected by the intonation of answers, the form of non-answers and the latency to response (e.g. a rising intonation often accompanied a wrong answer).

Krahmer and Swerts [7] describe experiments with adults and children on signaling and detecting of uncertainty in audiovisual speech. They found that when adults feel uncertain about their answer they more likely to produce pauses, delays and higher intonation (as well as some visual signals, such as eyebrow movements, and smiles). For children similar results were found but appeared not as uncertain as the adults were. The children in this experiment were aged 7-8 which is younger than the children in Rowlands study [11], where the age was around 10 years. (see next section). Age matters: Krahmer and Swerts suggest that young children do not signal uncertainty in the way adults do because they do care less about self-presentation than adults. Our study is about adults only.

6.2.2 Linguistic pointers to uncertainty

Knowledge questions, like the ones mentioned above, can be seen as 'testing questions' where the focus may not be on revealing the truth but rather on exposing ignorance and thus adding pressure on the speaker, making him or her nervous and uncertain [1]. Since a common perception about mathematical propositions is that they are either right or wrong, Rowland analyses transcripts of interviews with children focused on mathematical tasks and looks at the children's use of language to shield themselves against accusation of error [11]. According to Rowland, children tend to use a certain category of words (called hedges) which are associated with uncertainty. These hedges are further divided in different types:

- Shield
    - Plausibility shield (I think, maybe, probably)
    - Attribution shield (According to, says...)
- Approximators
    - Rounders (About, around, approximately)
    - Adaptor (A little bit, somewhat, fairly)

While some hedges are obvious shields to protect against 'failure', others are more elusive and require some contextual information. For example, the word 'about' may be a shield when used in combination with a number (e.g. 'there are about 150 thousand people in Enschede') but is no such thing when used in a sentence like 'the story is about a small boy'.

Another research which looks at the use of hedges is that of Bhatt et al. [3]. In their research they study how students hedge and express affect when interacting with both humans and computer systems. It was found that the students hedge and apologize to human tutors often, but very rarely to computer tutors. Another important result of their research is that hedging is not a clear indicator of student uncertainty or misunderstanding, but rather connected to issues of conversational flow and politeness.

6.2.3 Prosodic markers of uncertainty

Prosody is important because a speaker can communicate different meanings not extractable from lexical cues by giving acoustic ‘instructions’ to the listener how to interpret the speech. A good example is the increasing pitch (high F_0) at the end of a question. By using this kind of intonation the speaker draws attention to his question. Other theories include the speaker taking a humble stance by imitating a younger person (with higher F_0 and formants) since he’s actually asking a favour to the listener (answering his question) [10, p. 277].

In their research Liscombe et al. investigated the role of affect (student certainty) in spoken tutorial systems and whether it is automatically detectable by using prosody [8]. They discovered that tutors respond differently to uncertain students than to certain ones. Experiments with Intelligent Tutorial Systems (ITS) indicate that it is also possible to automatically detect student uncertainty and utilize that knowledge for improvement of these ITS’s, making them more humanlike. During their research they not only looked at the current (speaker) turn but also compared this turn with the dialogue history. Among the features analyzed were mean, minimum, maximum and standard deviation statistics of F_0 and the intensity, voiced frames ratios, turn duration and relative positions where certain events occurred.

6.3 Problem Statement

Much of the research mentioned above limits itself to the answering of what appears to us as trivial questions, or questions with short answers. So, there are questions about the usefulness of the results in a broader/different context. Many applications using automatic recognition of the degree of certainty of a person with respect to what he or she is saying might require different input than ‘simple’ question/answer-pairs. Since the experiments as described above needed relatively short answers (a few words) in order to get a standardized intonation [9], one wonders what the effects will be on longer utterances like normal dialogues, statements or presentations. Also, a rising intonation (a sign of uncertainty when answering a question) then can also be meant as a question itself (so how to differentiate between the two?) and the latency before an utterance may be irrelevant since the (potential) uncertain utterance might be encapsulated in other utterances from the same speaker. Nonetheless, these short utterances derived from question answering sessions make it possible to research prosodic features of speech which may be correlated with (un)certainty.

There are arise two questions here: Can prosodic features be used to automatically assess the degree of (un)certainty in a normal spoken dialog? And which features, if any, qualify best as prosodic markers to the qualification of this (un)certainty?

From previous research we have seen that certain features (intonation, latency) can be used to assess the degree of (un)certainty in (short) answers to questions. While the applicability of these features on utterance derived from normal dialogue may be a bit more complex they are still expected to be valuable indicators. Uncertain utterances will probably have a rising intonation due to the questionable nature of these utterances ("Maybe we can make a green remote?"). Also, common sense would correlate uncertain utterances with longer pauses (latencies) between words.

Besides intonation and latency (or gaps between words in case of longer utterances) we can include intensity (softer, less conviction in case of uncertainty) and the speed of talking to be a factor for identifying uncertainty. In both cases some way of comparing the uncertainity to a mean value for these features will be needed since it would not be possible to state whether the utterance has a below/above average value for intensity or speed.

6.4 Data Selection

In order to be able to perform prosodic analysis and reach valid conclusions, it seemed logical to use an existing corpus which had already been annotated. The AMI Corpus, which we addressed during the preliminary phase of this project, not only had many hours of high quality voice recordings but also annotations on different levels (hand made speech transcriptions, time aligned words, dialog acts) which could be used for this research.

6.4.1 Selection of Meetings

After reviewing the available annotation data for the AMI Corpus [2] a choice had to be made as to which sets were to be analyzed. Since the ES, IS and TS sets were the only ones with complete coverage of the words and dialog acts annotations and the existence of these annotations was considered essential these three sets were chosen. Table 6.1 shows that the our dataset comprised 552 audio files with a total duration of about 280 hours.

	Groups	Meetings	Files	Duration
ES	15	60	240	118:52:35
IS	10	40	152	93:05:28
TS	10	40	160	92:54:05
Total	35	140	552	278:01:50

Table 6.1 Overview of selected audio files.

A limitation of the corpus used in our work is the lack of sufficient stance annotations needed for the identification of uncertainty in speech. Since there was no reliable and efficient way to mark uncertain utterances, it was decided to use lexical elements (hedges) to identify utterances which would have a high probability of being uncertain. We split the dialogue acts into three classes: *uncertain* (that contain uncertainty hedges), *certain* (that contain certain hedges), and *neutral* (that do not contain any hedges).

Uncertainty	**Certainty**
according (to)	absolutely
approximately	certainly
around	clearly
fairly	definitely
maybe	(in) fact
perhaps	must
possible	obviously
possibly	(of) course
probable	positively
probably	surely
somewhat	undeniably
(I) think	undoubtedly
usually	

Table 6.2 Overview of hedges for uncertainty and words indicating certainty.

In Table 6.2 an overview of indicators used can be seen. These groups of words are derived from previous studies as performed by Rowland [11] and Bhatt et al. [3]. This approach raises some questions. In their study Bhatt et al. already disputed hedges being the only indicators for uncertainty, and have suggested that hedges can also be used for polite conversation as well [3]. To make sure the assumption we made was valid 25 random dialog acts, marked as uncertain during this research, were ranked on a five point scale ranging from certain to uncertain: certain – probably certain – undecided – probably uncertain – uncertain. 80% of the utterances were scored as either uncertain or probably uncertain.

6.4.2 Data Preparation and Selection

In preparing the AMI data to run through PRAAT, a program for speech analysis and synthesis [4], certain errors in the data were found (missing end or begin times of words). Since the Dialog Act tiers are based on the word tiers therefore several Dialog Act intervals had missing start and/or end times also and had to be discarded. In Table 6.3 the total amount of valid and invalid items can be seen. Since the percentage of these incorrectly annotated words and dialog acts was very low it was decided to simply discard them from the dataset instead of trying to figure out the correct data (if possible at all).

	Words			**Dialog Acts**		
Series	Valid	Invalid	Invalid %	Valid	Invalid	Invalid %
ES	351615	42	0.01%	47251	35	0.07%
IS	198968	14	0.01%	26909	14	0.05%
TS	283208	695	0.24%	42394	419	0.98%
Tot	833791	751	0.09%	116554	468	0.40%

Table 6.3 Overview of converted words and dialog acts.

PRAAT was used for the prosodic analysis [4]. First a selection of the relevant prosodic features which had to be measured had to be made.

For each category of the prosodic properties mentioned in Subsection 6.2.3, several attributes were chosen and implemented in PRAAT. Beside these prosodic attributes some lexical attributes (like amount of words, the presence of ‘yeah (, but)’, ‘okay’) were added as well. In total 76 attributes were chosen for the analysis, of which 67 were prosodic.

The number of dialog acts including hedges consists of only 7.26% of the total (7317 dialog acts of a total of 100799), which means that simply classifying each dialog act as certain gives a score of about 93%. By balancing the dataset the script will take 4819 random other dialog acts and combine them with the ones containing hedges to form a new dataset.

6.4.3 Statistical Analysis

Since the dataset preparation script in phase 4 has been designed in such a way that different datasets can be created on the fly it is easy to compare different prosodic features of different classes. In phase 3 of the research, the actual prosodic analysis, the presence of several lexical markers or indicators was also checked. Among these markers were the hedges as mentioned before, the group of words (supposedly) indicating certainty, yeah and okay.

6.5 Experimentation

During the following experiments all datasets were leveled on a 50/50 basis so each ‘group’ was equally represented. As a result the baseline (computed with the ZeroR classifier) of all datasets is about 50%. Next, the datasets were classified with the J48 (tree) and NaiveBayes (NB) classifiers. Each classifier was evaluated for accuracy using 10-fold cross-validation. We used the implementation in the Weka toolkit [13]. To determine the key attributes being used for this classification the input data was also evaluated using the InfoGain attribute evaluator in combination with a Ranker search method.

6.5.1 Hedges –vs– No Hedges

First the dataset with the hedges was analyzed. Out of all 100799 dialog acts analyzed with PRAAT in phase 3 only 7317 contained one or more hedges (see also Table 6.4). These instances were complemented with the same (random) amount of dialog acts containing no hedges. Based on previous research it was expected that several prosodic features would be good indicators for uncertainty in speech. Among these features were a rising pitch, a declining intensity and a slower rate of speech (more pauses and/or longer average word-length).

Class	**Instances**
No Hedges	7317 (dropped 85502)
Uncertain Hedges	7317

Table 6.4 Properties of dataset Uncertain Hedges –vs– No Hedges.

In Table 6.5 the results of the analysis can be seen. Two classifiers were used (J48 and NaiveBayes); for each the improvement over the baseline (IOB) is included in the table. As anticipated the baseline is about 50% correct classifications. Two striking results are the overall improvement over the baseline score (with an average increase of about 17/18% based on which classifier has been used) and the high performance on the lexical features alone. The evaluation of the (key) attributes show the importance of attributes related to the length of the dialog act.

Baseline (ZeroR)	49.98%			
Features	**J48**	**IOB**	**NB**	**IOB**
Lexical features (LF)	74.67%	24.69%	71.27%	21.29%
Spectrum related features (SF)	67.70%	17.72%	64.59%	14.61%
Pitch related features (PF)	68.27%	18.29%	67.04%	17.06%
Intensity related features (IF)	63.61%	13.63%	61.20%	11.22%
Formant related features (FF)	66.80%	16.82%	67.97%	17.99%
All Prosodic Features	66.05%	16.07%	68.46%	18.48%
All features	**71.14%**	**21.16%**	**69.96%**	**19.98%**
Average Improvement		18.34%		17.23%

Table 6.5 Classification Performance of Hedges –vs– No Hedges including improvement over baseline (IOB).

The first 8 attributes, headed by the number of words (da_words) in the DA, are all related to the DA length, either indicating time or the amount of (voiced) frames or bins. Since the utterances in the corpus have been marked (un)certain by using hedges this is not very surprising: hedges are normally part of (longer) sentences. As a result, the length of a dialog act (shown by a number of attributes) is a good indicator since short dialog acts are often marked certain.

After the attributes indicating length in some way the type of dialog act is also important, taking a 9th place in the attribute ranking. Apparently the type of DA as annotated by the members of the AMI Project has some relation to uncertainty. More about the distribution of hedges over dialog acts can be seen in section 6.3. Next in the attribute ranking are several formant attributes headed by the minimum F2, maximum F1 and maximum F2. After several other formant attributes the standard deviation for the intensity during the 2nd half of the DA (intensity2_sd), the spectrum band energy (spectrum _band_energy) and the voiced frame ration during the 2nd half (pitch2_voiced_fr_ratio) and the total DA (pitch_voiced_fr_ratio) seem to be good indicators for uncertainty. When classifying the dataset with the J48 classifier and using only the formants' minimum and maximum values the performance result is 67,3%, even higher than when using all formant attributes. Classification based on the voiced frame ratios only gives a performance of 59,8%.

6.5.2 Uncertain Hedges –vs– Certain Hedges

Similar to the previous dataset where dialog acts with hedges were compared to dialog acts without these lexical markers another set was created which contained all dialog acts with words which should indicate certainty and compared to a similar sized group of hedged dialog acts. As can be seen in Table 6.6 the size of this dataset was significantly smaller.

Class	**Instances**
has_ hedge[1] = Hedges	663 (dropped 6654)
has_ hedge[2] = Anti-Hedges	663

Table 6.6 Properties of dataset Hedges –vs– Anti-Hedges.

Baseline (ZeroR)	49.77%			
Features	**J48**	**IOB**	**NB**	**IOB**
Lexical features (LF)	58.30%	8.52%	57.77%	7.99%
Spectrum related features (SF)	55.66%	5.88%	50.38%	0.60%
Pitch related features (PF)	55.13%	5.35%	52.26%	2.49%
Intensity related features (IF)	53.09%	3.32%	54.45%	4.68%
Formant related features (FF)	51.58%	1.81%	55.13%	5.35%
All Prosodic Features	55.28%	5.51%	54.90%	5.13%
All features	**56.41%**	**6.64%**	**55.51%**	**5.73%**
Average Improvement		5.29%		4.57%

Table 6.7 Classification Performance of Hedges –vs– Anti-Hedges including improvement over baseline (IOB).

In contrast with the expectations mentioned above the actual results show a lower performance of the classifiers with an average improvement of about 5%. Once again the lexical features score best, although the gap is smaller.

This time the attribute ranking shows the type of DA (da_type) being the most predictive attribute, followed by some length related attributes. The first prosodic feature is the mean F4 (6th place), followed by the minimum intensity (9th) and minimum pitch (12th). In contrast to the previous dataset where the formants played an important role, for this dataset the pitch values (mainly of the 2nd half of the DA) seem to be a better indicator for uncertainty.

6.5.3 Distribution of hedges over dialog acts

To see whether uncertain utterances occur more in particular dialog acts the distribution of dialog acts marked uncertain over the different dialog act classes has been looked into, the results of which can be seen in Table 6.8. For comparison, the distribution of all dialog acts has been included as well.

Dialog Acts (ID)	**Total Dialog Acts**	**Percentage of Total DA's**	**Hedges**	**Percentage of Hedges**	**Percentage of Dialog Act**
Minor	**30816**	**30.6%**	**670**	**9.2%**	**2.2%**
Backchannel (1)	10655	10.6%	33	0.5%	0.3%
Stall (2)	6983	6.9%	82	1.1%	1.2%
Fragment (3)	13178	13.1%	555	7.6%	4.2%
Task	**56438**	**56.0%**	**6094**	**83.3%**	**10.8%**
Inform (4)	29841	29.6%	2456	33.6%	8.2%
Suggest (6)	8610	8.5%	1645	22.5%	19.1%
Assess (9)	17987	17.8%	1993	27.2%	11.1%
Elicit	**6557**	**6.5%**	**396**	**5.4%**	**6.0%**
Elicit-Inform (5)	3743	3.7%	125	1.7%	3.3%
Elicit-Offer-Or-Suggest (8)	640	0.6%	45	0.6%	7.0%
Elicit-Assessment (11)	2016	2.0%	225	3.1%	11.2%
Elicit-Comment-Understanding (13)	158	0.2%	1	0.0%	0.6%
Other	**6988**	**6.9%**	**157**	**2.1%**	**2.2%**
Offer (7)	1370	1.4%	80	1.1%	5.8%
Comment-About-Understanding (12)	1942	1.9%	16	0.2%	0.8%
Be-Positive (14)	1856	1.8%	40	0.5%	2.2%
Be-Negative (15)	84	0.1%	3	0.0%	3.6%
Other (16)	1736	1.7%	18	0.2%	1.0%
Total	**100799**	**100.0%**	**7317**	**100.0%**	**5.5%**

Table 6.8 Distribution of (uncertain) dialog acts.

As can be seen in Table 6.8 most dialog acts are task oriented or minor (56% and 31% respectively). We can also notice that most dialog acts marked as uncertain (by containing hedges) belong to the task-category.

The class of minor acts contains significantly less hedges than the class of elicit acts ($\chi^2(df = 1) = 311.45$; $p < 0.001$) and the class of elicits contains significantly less hedges than the class of task acts ($\chi^2(df = 1) = 143.93$; $p < 0.001$).

6.6 Conclusions

Based on the results described above, with classification performance increases of more than 20%, it is feasible to conclude that the degree of (un)certainty in spoken dialogues can be assessed automatically. When looking at the features which qualify best as prosodic markers to uncertainty the textual features obviously score best. Due to the nature of the uncertain utterances (being based on hedges which most often require some sort of sentence) this result might be of no surprise. There also appears to be a connection between the type of dialog acts (as annotated by members of the AMI Project) and the degree of uncertainty since the presence of uncertain utterances in several dialog act types is clearly above average. A relatively high percentage of uncertain dialog acts are suggestions or assessments. Whether these dialog acts are really uncertain or whether politeness strategies play a role here is hard to establish.

We have found interesting results about which of the prosodic markers will best serve in the detection of uncertainty. It was predicted that a rising intonation, longer pauses (latencies) and a decreasing intensity would be good indicators for uncertainty. Based on the attribute evaluation of the different datasets these theories seem to be supported, showing important roles for the pitch and intensity features. Especially with the dataset ‘Hedges –vs– No Hedges’ the minimum and maximum values of the formants are good prosodic markers as well.

Even though the results seem straightforward, with impressive classifier improvements over the baseline performances, several questions still remain.

In the current research the feature extraction was based on previous research and the scope and functionality of PRAAT. While a broad range of features have been researched it could very well be certain additional features might be promising as well. Another improvement could be using custom settings in PRAAT. For now all settings have been kept on default, but it is known that, for optimal results, different settings should be used for men and women for example. Additional difficulty would be to either automatically detect the gender of a speaker and adapt the settings accordingly, or manually set gender-values for all 500+ files.

For future research on this topic it would be advisable to have a clear understanding of what the ‘uncertainty’ being researched entitles and how it can be measured. Having that information should provide a basis for reliable annotations, with which further research can be done.

Further research in hedges and/or other lexical markers as indicators for uncertainty looks promising. The results of combined feature sets already showed the best results and expanding those features with other indicators (also visual) will probably give the best results in the end (although not all types of information will be available in all situations).

Acknowledgements This work is supported by the European IST Programme Project FP6-033812. This article only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

References

1. Ainley, J. (1988). Perceptions of teachers' questioning styles (pp. 92–99). In: *12th International Conference for the Psychology of Mathematics Education*, Vezprém, Hungary.
2. Carletta, J. C. (2007). Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. *Language Resources and Evaluation*, 41(2), pp. 181–190.
3. Bhatt, K. and Evens, M. and Argamon, S. (2004). Hedged Responses and Expressions of Affect in Human/Human and Human/Computer Tutorial Interactions. In: *26th annual meeting of the Cognitive Science Society*, Chicago, Illinois, USA.
4. Boersma, P. and Weenink, D. (2008). Praat: doing phonetics by computer. Version 5.0.06 [Software]. Available: `http://www.praat.org/`.
5. Brennan, S. E. and Williams, M. (1995). The feeling of anothers knowing - prosody and filled pauses as clues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, vol. 34, pp. 383–398.
6. Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, vol. 56, pp. 208–216.
7. Krahmer, E. and Swerts, M. (2005). How children and adults signal and detect uncertainty in audovisual speech. *Language and Speech*, vol. 48(1), pp. 29–54.
8. Liscombe, J. and Hirschberg, J. and Venditti, J. J. (2005). Detecting Certainness in Spoken Tutorial Dialogues (pp. 1837–1840). In: *9th European Conference on Speech Communication and Technology*, Lisbon.
9. Ozuru, Y. and Hirst, W. (2006). Surface features of utterances, credibility judgments, and memory. *Memory & Cognition*, vol. 34, pp. 1512–1526.
10. Rietveld, A. C. M. and Van Heuven, V. J. (1997). *Algemene Fonetiek*. Bussum: Uitgeverij Coutinho.
11. Rowland, T. (1995). Hedges in mathematics talk: Linguistic pointers to uncertainty. *Educational Studies in Mathematics*, vol. 29, pp. 327–353.
12. Smith, V. L. and Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, vol. 32, pp. 25–38.
13. Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. 2nd ed. San Francisco: Morgan Kaufmann.