Proceedings of the 37th Hawaii International Conference on System Sciences - 2004

An Exploratory Study on Promising Cues in Deception Detection and Application of Decision Tree

Tiantian Qin, Judee Burgoon, Jay F. Nunamaker, Jr.
University of Arizona
{tqin, jBurgoon, jnunamaker}@cmi.arizona.edu

Abstract

Automatic deception detection (ADD) becomes more and more important. ADD can be facilitated with the development of data mining techniques. In the paper we focus on decision tree to automatic classify deceptions. The major question is how to select experiment data (input data for training in decision tree) so that it maximally benefits the decision tree performance. We investigate promising level of the cues of experiment data, and then adjust the applications in decision tree accordingly. Five comparative decision tree experiments demonstrate that tree performance, such as accurate rate and complexity, is dramatically improved by statistically and semantically selecting cues.

1. Introduction

Deception means that messages are transmitted to cause a false impression or conclusion [3]. The challenge of detecting untrue information becomes more significant after the event of 911. For this purpose, tremendous efforts are needed to analyze messages, compare with previous records, and figure out the suspicious information. Furthermore, the empirical studies provide evidence that humans are typically very poor at detecting deception and fallacious information [11]: especially when the messages are sourced from text-based, computer mediate, where the accuracy is little better than chance [9]. Tools that facilitate human deception detection are therefore valuable, such tools would also benefit to law enforcement in dealing with criminal investigations.

Previous literatures offer many prospective cues that might be useful to distinguish deception and truth in text-based messages, while manageable by computer software tools [2,7] Such linguistic cues include number of sentences, number of words, sentence complexity etc. Judging whether one is deceptive is equal to decide the classification (true or deceptive) of one's message, from a set of attributes (linguistic cues). Considering the suspicious message as a data (record) with a list of attributes and an unknown classification (true or deceptive), deception detection is nothing more than a data classification process.

Many data mining techniques are now available to classify data: such as neural networks, Bayesian networks, k-means, decision trees, etc [8,7]. All these data mining techniques require training process where data with known classification are input and their attributes are auto analyzed. The training process then produce a classification baseline, which could be a tree structure (decision tree) or a network structure (Bayesian network), and so on. Future data can be classified based on the baseline.

Research on cues and decision trees make deception detection objectives possible. Briefly speaking, the final research objective is to decide automatically whether a text-based message (such as an email) is true or not. In data mining, the goal is to classify the data (message) into one of 2 categories (true or deceptive) based on its attributes (linguistic cues). Among many data mining methods, we focus on decision tree (C4.5) in this paper, because it is powerful and the tree structure shows interactions of cues. The most challenging part of the goal is the training process, i.e., constructing a baseline structure, or threshold values distinguishing between deception and truth.

Training data is critical to the training process. In the current stage, we obtain training data from experiments [2]. However, we claim that directly applying experiment data as training data (a quick and dirty method) for two major reasons does not guarantee a reliable baseline:

First, deception behavior (represented as a cue) is consistent in all contexts. Cues (attributes) need to be validated according to their context, since deceptive behavior (cues) perform significantly variously, sometimes even oppositely under different circumstances [9]. For example, people in vocally chatting might communicate more informally than in text chatting, where messages contain more complicated and/or longer sentences. It is necessary to semantically analyze experiment data, and to construct many baselines depending on the corresponding contexts.

Second, even in the same context, characteristics of cues may differ. It has been noted that selection of attributes tremendously influences data mining [7]. Some cues are more promising than others for use in detecting deception. For example, some may be more significant in a statistical sense. Because including superfluous attributes (cues) in training data would decrease the performance of decision tree, cue selection is also an important issue in selecting training data.

In this paper, we concentrate on how to select cues (attributes) according to context and promise for improving training data (i.e., the preprocessing of experiment data in order to generate a qualified training data set for a decision tree). We demonstrated that, semantic and statistic analysis of the experiment data can result in drastically improves in decision trees. Meanwhile, consistent (under some context) and promising cues can maintain a focus on important indicators of the direction future deception detection experiments should take. As the first paper on semantic selection of training data for decision-tree implementation in a deception detection context, it sheds light on questions and discussions in this area.

This paper is organized into the following sections: The Method section describes deception detection experiments that provide source data; linguistic cues (attributes). In results and discussion section, we analyze cues and summary on their potential power in application; we also point out preprocessing methods that could enhance the decision performance. Application section shows five comparative decision tree experiments that support our preprocessing methods in previous section. Next, we take a close look at the decision trees generated, and explain the tree under deception detection context. We then finish the paper by summary the findings in conclusion section.

2. Method

The Mock theft experiment [2] was designed as a pilot study. One of the purposes of this study is to reveal useful linguistic cues to detect deception. In this experiment, participants staged a mock theft and were subsequently interviewed by untrained and trained interviewers via text chat (Txt), audio conferencing (Audio) , or face-to-face (Ftf) interaction. The FtF interactions were later transcribed, and the transcripts and chats were submitted to linguistic analysis such features as number of words, number of sentences, number of unique words (lexical diversity), emotiveness, pronoun usages, plus several others that are available in the Grammatik tool within WordPerfect. Due to small sample size, only a few of the differences between innocents (truth tellers) and thieves (deceivers) were statistically significant. urther more, because the effect of interaction modality (text chat, face-to-fact or Audio chat), cues might perform differently even opposite between truth tellers and deceivers under different modality. An example is: truthful tellers had more sentences than deceivers in Txt, but less in Audio, i.e., cues are not "consistent" across different modalities.

In the mock theft experiment, students were recruited from a multi-sectioned communication class by offering them credit for participation and the chance to win money if they were successful at their task. Half of the students were randomly assigned to be "thieves," i.e., those who would be deceiving about a theft, and the other half became "innocents," i.e., those who would be telling the truth.

Interviewees in the deceptive condition were assigned to "steal" a wallet that was left in a classroom. In the truthful condition, interviewees were told that a "theft" would occur in class on an assigned day. All of the interviewees and interviewers then appeared for interviews according to a pre-assigned schedule. We attempted to motivate serious engagement in the task by offering interviewers $10 if they could successfully detect whether their interviewee was innocent or guilty and successfully detect whether they were deceiving or telling the truth on a series of the interview questions. In turn, we offered interviewees $10 if they convinced a trained interviewer that they were innocent and that their answers to several questions were truthful. An additional incentive was a $50 prize to be awarded to the most successful interviewee.

Interviewees were then interviewed by one of three trained interviewers under one of three modalities— Face to Face (FtF), text chat, or audio conferencing. The interviews followed a standardized Behavioral Analysis Interview format that is taught to criminal investigators.

There are three segments of questions; the first segment is questions on previous class question, such as, what is your favorite class? In this segment every subject gave true answers. The second segment is on question on previous jobs, such as: describe your last job. In this segment, subjects acting as "thieves" are supposed to fake on the job stories, and had to make up one if they did not have any job experience. The third segment is the questions on the "stealing" event: described your experience in that classroom where the wallet was missing. These three segments recorded the pattern of subjects when they were: telling a truth, telling low risk lies on familiar topics, and telling high risk lies. The third segment is the situation that we are most interested in, since the deceptive behavior is close to those could bring the most dangerous (high risk) consequence in the real world. In application section, the decision tree will be trained on data in the third segment.

Interviews were subsequently transcribed and submitted to linguistic analysis. Clusters of potential indicators, all of which could be automatically calculated with a shallow parser (Grok or Iskim) or could use a look-up dictionary were included. The specific classes of cues and respective indicators were as follows:

1. Quantity (number of syllables, number of words, number of sentences, number of short sentences, number of simple sentences)

2. Vocabulary level Complexity (number of big words, number of syllables per word, lexical complexity, and lexical complexity)

3. Sentences level Complexity (Flesh-Kincaid grade level, average number of words per sentence, sentence complexity, number of conjunctions)

4. Specificity and Expressiveness (emotiveness index, rate of adjectives and adverbs, number of affective terms, sensory and RM terms)

5. Informality (total number of flagged errors (from Word Perfect))

The reason for these cue classes is that they distinguish truth from deception while they are machine extractable. In general, deceivers are higher on quantity, specificity and expressiveness, and less on complexity [10]. For informality, deceivers are less informality than truth-tellers [3].

The data set of Ftf is small since the transcription process has not finished, therefore we did not consider Ftf. We applied a design for statistical test, where there are 3 segments, 2 modalities, and 2 conditions (truth teller and deceiver). Total 58 subjects: 28 are for Txt (16 are "thieves"), and 20 (9 are "thieves") are for Audio.

We want cues that are both consistent and significant. "Consistent" means that people's deceptive behaviors are identical in some contexts. This paper consider context only in the two categories: segment (low-risk, high-risk), and modality (text, Audio). Therefore there are 4 contexts: low-risk Text, low-risk Text, high-risk Audio, and high-risk Audio. As an example of consistent cue: we seem sentence complexity consistent if deceivers use simpler sentences (lower sentence complexity), in the 4 contexts. "Significant" means that the values of cues are statistically different between deceivers and truth-tellers. In reality, it is usually difficult to find a cue that is consistent under all contexts, while remains significant. We need to decide from the combined performance of cues and come up with a set of good cues. These good cues are referred to as promising cues in rest of the paper. In the next section, we will discuss how to select promising cues.

3. Results and Discussion

In this section, we will discuss issues that influence the promising level of cues: significance (statistically) level of the cues, relationships among the cues, and consistency of cues (in contexts of high-risk Txt, and high-risk Audio). We summarized promising cues and predict that considering only those promising cues (attributes) in defining the training data should result in better decision trees having a reliable correct classification rate and less complexity, than using all 19 cues. Empirical evidence supports such a prediction in the next section, in which several decision trees, built using different training data, are compared.

3.1. Relations among cues

Including redundant attributes in training data cannot improve the tree performance, but adds in unnecessarily computational complexity. For example, if number-of-syllables and number-of-words are highly correlated and can be statistically demonstrated to be substitutable one for the other, retaining both cues is actually almost the equivalent of using the same cue twice. Therefore, we decide to check the relationships among cues in order to eliminate superfluous ones.

The 19 cues can be grouped into 5 dimensions: quantity, vocabulary-level complexity, sentence-level complexity, Specificity and Expressiveness, and informality. Table 1 shows that cues in the first column can be replaced by corresponding cues in the second column, for the reasons shown in the third column, where correlations between cues and reliability test are listed. For example, number-of-simple-sentences and number-of-short-sentences are highly correlated with number-of-sentences; number-of-syllables is highly correlated with number-of-words. Reliability testing confirmed that these variables measuring similar information and can be replaced one for the other. This suggests that eliminating # simple sentences, # short sentences, and # syllables can reduce complexity without losing information.

Table 1. Cues relation

Duplicated variables	Represented by	Reason	
		Correlation	Reliability test
Simple sentences	Sentences	0.692**	0.7074
Short sentences		0.863**	0.8045
Syllables	Words	0.993**	0.9376

**: Significant at .05 level.

For simplicity, these 3 cues: number-of-simple sentences, number-of-short sentences, and number-of-syllables will be defined as "duplicated". Getting rid of duplicated cues is expected to improve, or at least not reduce, the performance of trees, since noise is reduced by keeping just sufficient information into the training set.

3.2. Statistically significances

A cue is considered statistically significant if the difference in its occurrence between deceivers and truth-tellers is statistically significant, and we have reason to believe therefore that this difference is not due to chance. Significant cues are more important because they represent systematic difference. However, because decision tree (and many other existing data mining tools) cannot automatically determine the statistical significance of attributes (cues). we rely on traditional statistical methods.

A series of GLM and independent sample t-tests were applied. F and P values are shown in table 2. The multivariate testing of cues occurring frequently produced no significant multivariate effects ($p>0.01$). T-tests provided weak support significance of number-of-words ($p = .096$).

Multivariate testing of complexity at both the sentence-level (simple sentences, long sentences, short sentences, sentence complexity, Flesch-Kincaid grade level, number of conjunctions, average-words-per-sentence (AWS)) and the vocabulary level (vocabulary complexity, number of big words, average-syllables-per-word (ASW)) did not showed significance. T-test provided evidence for effects on deception condition for several individual variables (AWS, with p=.021; Flesch-Kincaid grade level, with p=.056; and sentence complexity, with p=.082). This implies that sentence complexity cues are helpful for distinguishing between deceptive and true messages.

Table 2. F-(p-) values of 19 cues

Cues	Multivariate test between subjects	Independent Samples t-Test	Cues	Multivariate test between subjects	Independent Samples t-Test
Syllables	1.842(.182)	1.502(.140)	Sentence complexity	2.055(.159)	1.779(.082)*
Words	2.407(.128)	1.702(.096)*	Vocabulary complexity	.512(.478)	-.997(.324)
Sentences	.001(.972)	.111(.912)	# of Conjunctions	2.569(.116)	1.426(.163)
Short sentences	.588(.447)	-.725(.472)	Rate Adjectives and Adverbs	.329(.569)	-.596(.554)
Long sentences	6.566(.014)*	2.781(.008)*	Emotiveness	.054(.818)	-.233(.817)
Simple sentences	.002(.969)	.061(.951)	Lexicon complexity	.771(.404)	.618(.542)
Big words	.288(.594)	.616(.541)	Sens&RM	.568(.457)	.769(.447)
Average syllables per word	1.703(.199)	-1.668(.102)	Total flagged errors	3.945(.056)*	2.174(.037)*
Average words per sentence	4.368(.042)*	2.414(.021)*	Affect	3.291(.214)	1.630(.110)
Flesch-Kincaid grade level	2.690(.108)	1.958(.056)*			

*: Significant at .1 level

For specificity and expressiveness (adjectives and adverbs, emotiveness, Sens& RM, and affect), the multivariate and t test showed no significance, p>.1.

Informality of message (total flagged errors) was significantly different between deception and truth by both multivariate test (p=.056) and t test (p=.037), implying that informality might also be useful in training data.

In general, number-of-words, number-of-long-sentences, AWS, Flesch-Kincaid grade level, sentence complexity and total-flagged-errors can be considered to be more significant for cue significance than other cues. We refer to them as significant for simplicity. This investigation confirmed that deceivers behave differently from truth tellers in text chat and/or Audio chat communications.

3.3. Consistency

A cue is considered consistent if it shows the same behavior under different circumstances (modality and segment). For example, if deceivers speak or write less than truth tellers in response to both low-risk and high-risk questions (segment), and/or in both Txt and Audio situations (modality), we consider the measure to be consistent measure.

We combine modality and segment effects to decide the consistency of cue. Consistency is clearly visible on profile-plots such as that for number-of-words in Figure 1.

In Figure 1 Modality =1 refers to Txt and 2 refers Audio. Question 1, 2 and 3 represent segment, i.e., question 1 is the segment where all subjects told the truth; questions 2 and 3 are segments in which subjects acting "thieves" told low- and high- risk lies, respectively. We focus on segments 2 and 3. In the Txt situation, deceivers said or wrote fewer words in both segments 2 and 3. In the audio situation, however, whether deceivers said or wrote fewer words varies on segments, i.e., more in low-risk context and fewer in high-risk context. Therefore, number-of-word was not a consistent cue because it depended on the segments. Although this method is subjective, it provides a vivid and effective way to look at the consistency of cues. Using the comparison method on all cues in the quality dimension (number of syllables, number of words, number of- sentences, number of short sentences, number of simple sentences), reveals trend in which quality are consistent only in Txt, since the deceiver has less quality than truth teller in both context of low-risk Txt, and high-risks Txt.

Continuing the consistency comparison for the remaining cue dimensions (Lexical level complexity, sentence level complexity, expressiveness, specificity, and informacy), we observe that sentence complexity is the most consistent. All profile plots of sentence complexity are shown in figure 2. The only inconsistent-point happens in number-of-conjunctions, in Audio modality. The inconsistence implies that Audio modality is less reliable than Txt, which remains true for all cue dimensions other than sentence complexity. We then predict that using only Txt data as training set could result in better decision trees than using data with both modalities, because Audio modality is not consistent and will bring in noise in the decision-trees.

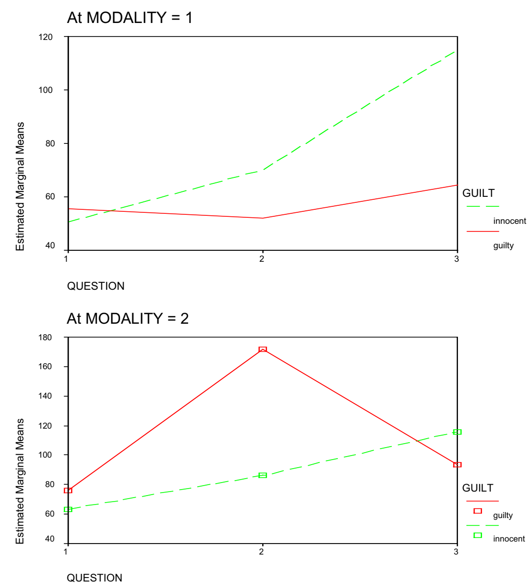
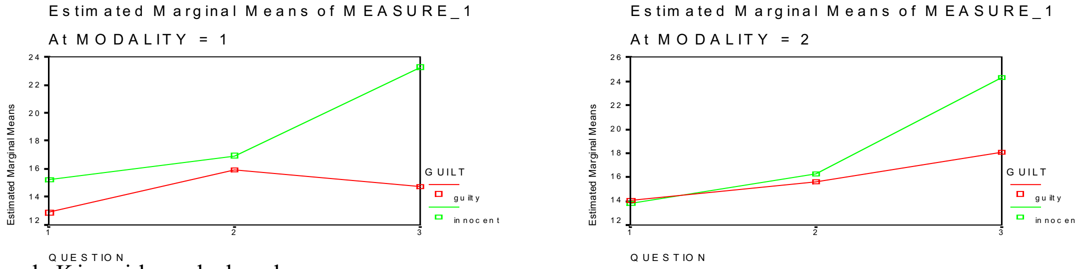


Figure1. Profile plot for number-of-words

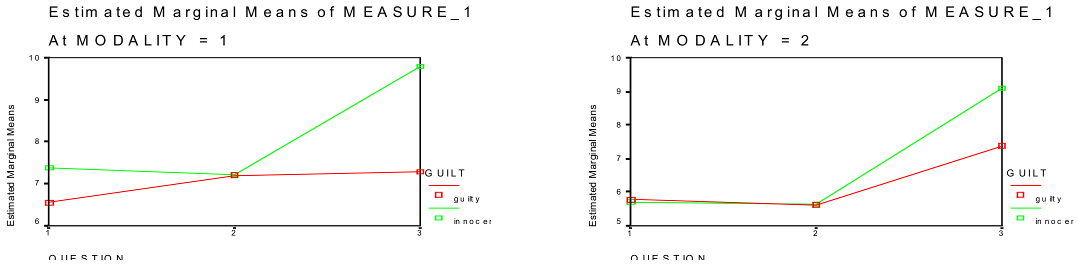
3.4. Summary

In table 3, we summarized promising cues by combining significance and consistency.

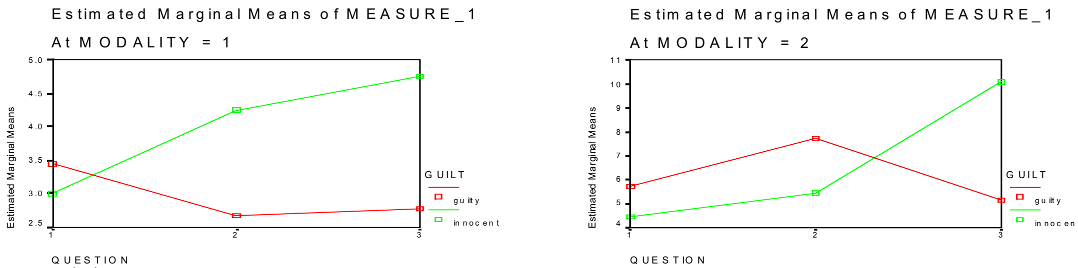
Words per sentences:



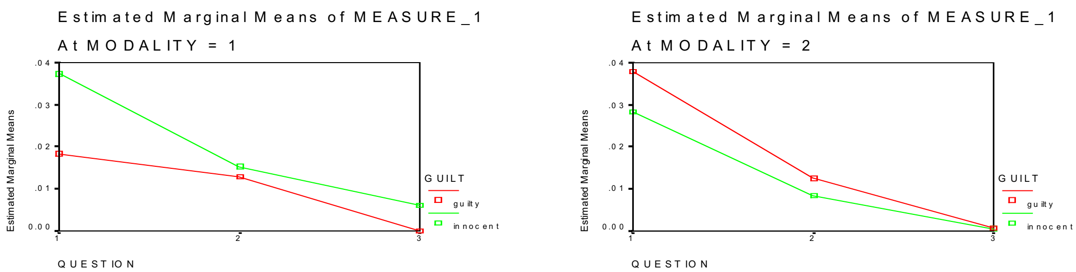
Flesch-Kincaid grade level:



#Conjunctions:



RateModals:



Sentence complexity:

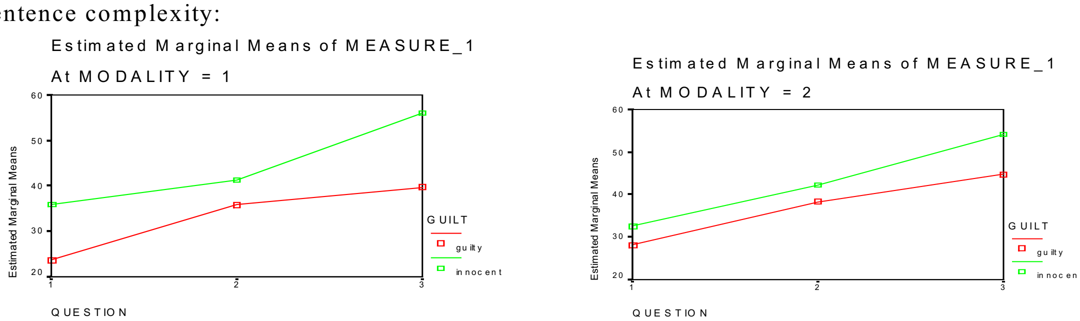


Figure 2. Profile plots of sentence level complexity

Although expressiveness and specificity seemed not to be consistent and significant in this pilot study, they are important because they represent a new information dimension. The reason for their not showing significance could be a consequence of the small data set and dictionary used to calculate these cues. We need to continue to evaluate them with more experiment data and a better dictionary.

In short, getting rid of the three duplicated cues could decrease the complexity of the decision tree. The six significant cues are potential to be the most efficient cues for training data. Cues are more consistent in a Txt situation, suggesting that using only Txt data as a training set could construct better decision trees.

Table 3. Summary of promising cue

Cue Class	Representative Cue	Significant Difference(D/T)	Consistency	Promising level
Quantity	# Words	Yes	Better under txt	mid-high
Lexical level complexity		No	No	low-mid
Sentence level complexity	AWS, FK Grade, Sentence complexity	Yes	Good	high
Expressiveness	Could be good indicators. Need more tests			
Specificity				
Informacy	Errors	Yes	Better under txt	mid-high

4. Application

The technique for constructing decision tree is provided in C 4.5 (Weka). Cross validation has been used for more reliable results [7]. We used the data in segment 3 because this type of deception (high risk) is more interesting.

Figure 3 shows the results of five experiments. In "Original" we used all 19 cues and both Txt and Audio as training set; In "No Duplicate" we used 16 cues without the 3 that are duplicates (number of -simple sentences, short sentence, and syllables). In "significant" we use only the 6 statistically significant cues (number of words, long sentences, AWS, Flesch-Kincaid grade level, sentence complexity, total flagged errors); the next two tests contain only Txt data, with no duplicated and significant cues, respectively. Each test was repeated 20 times and we recorded the highest prediction rate (the rate at which tree successfully classified deception and truth). For Audio-only data is not displayed because previous analysis and profile plots revealed inconsistency in the Audio data. However, ways to auto-distinguish deception in Audio context may exist and the Audio context is a highly sensitive scenario that needs further semantic analysis and deeper refinement of the data.

As shown in the figure, the prediction performance rate is increased among all the experiments. Training with Txt was significantly better than that with combined data from Txt and Audio. This means that deceptive behavior in Txt and Audio are so different that combining them is likely produce more noise in the training set, thus damaging the prediction rate. That cues are also more consistent in Txt meaning that Txt data are more reliable.

Although improvement is not significant, organizing training data without duplicated cues is slightly better than training with all cues. Simplifying cues so that they represent sufficient information in messages can help to improve performance, and the performance is even better when using only significant cues.

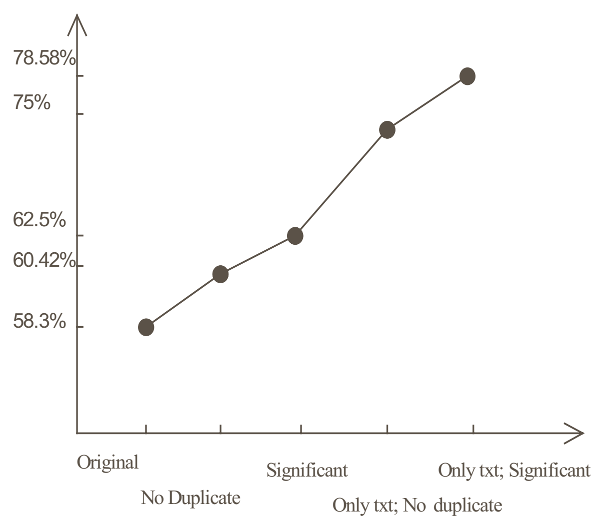


Figure 3. Performance (prediction rate) of C4.5

5. A close look at decision trees

In this section, the two best decision trees are displayed and explained in detail: 1.) a tree that was built with Txt-only data, plus 16 no duplicated cues; and 2.) a tree that was built with Txt-only data, 6 significant cues. The two trees displayed little noise in training set and therefore, exhibit more robust structures than other trees. All 5 trees are shown in the appendix.

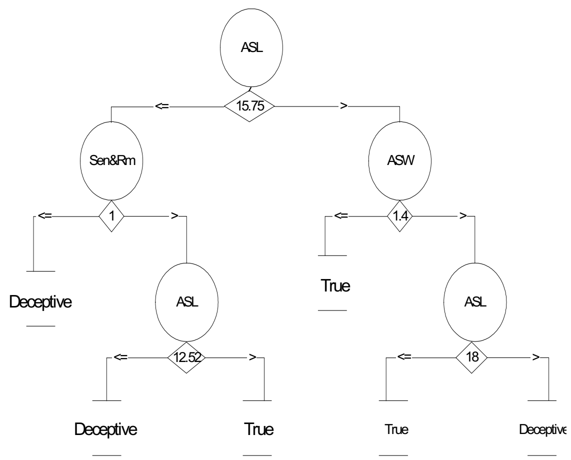


Figure 4. Txt only, with 16 no duplicated cues

As shown in figure 4, the decision tree picked up average_words_sentence, Sens&RM, and average_syllables_per_word (ASW), and organized them in a hierarchical way. Except for average_syllables_per_word (ASW), most of them were significant cues.

For simplicity, we explain that, in figure 5, if average sentence length was greater than 15.75, the message will be considered true; also, the message is deceptive if sentence complexity is greater than 22; and true if sentence complexity is less than 22.

In figure 5, where only 6 significant cues are used in training, the decision tree is much simpler, yet no less accurate than that in figure 4. This supports the expectation that semantic analysis and pre-selection of the cues in a training set can reduce the complexity, since decision trees cannot automatically filter out noisy data. For instance, the non-significant cue, ASW, did not increase the accuracy but introduce more complexity. Only through semantic analysis, can a noisy cue be detected and purged.

These two trees demonstrate the importance of sentence-level complexity in deception detection. Average sentence length (ASL) is the most important cue. In figure 5, a truth-teller uses longer sentences than deceivers. This implies that more information is available in each sentence. On the other hand, deceivers using shorter sentences, implies that they pause more often make up fake stories, possibly under a heavy cognitive load [1]. Sentence complexity (more compound sentences, and longer words) also plays a role in deception detection. A truth-teller, who feels at ease and undergoes less cognitive load, uses simpler sentences while recalling a previous experience. Compared with truth-tellers, deceivers strive to make a credible impression [9]. As a result, they use more formal, and more complex sentences, hoping that formal writings appear more credible. In short, the structures of decision trees show reasonable patterns that coincide with previous research.

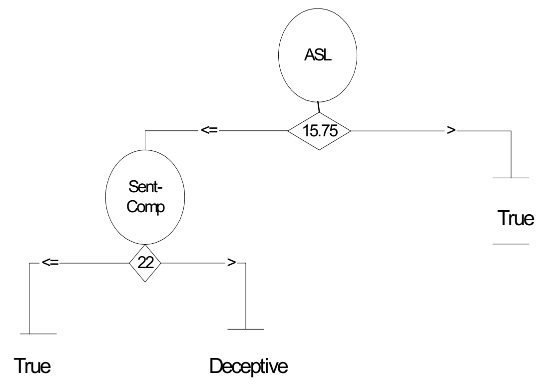


Figure 5. Txt, with 6 significant cues

The examples of trees are for Txt-only data in a high-risk context. Further analysis needs to be conducted for other contexts, since deception detection is sensitive to contexts

6. Conclusion

This paper reports on a preliminary study of selection of cues to generate more reliable training data. We describe a method of purifying experimental data by eliminating unpromising cues. We demonstrate that the purifying method actually enhanced performances of decision trees, with the best decision tree resulting from using only Txt data.

Given such a small data set, the current experiment showed big variances in tree structure and prediction performance. However, light is shed on the potential power of selecting the training data semantically and statistically.

7. Reference

[1] Buller, D. B. & Burgoon, J. K. (1996). “Interpersonal Deception Theory.” Communication Theory, 6, 203-242.

[2] Burgoon, J., Blair, J., & Moyer, E. (2003, November). “Effects of Communication Modality on Arousal, Cognitive Complexity, Behavioral Control and Deception Detection during Deceptive Episodes.” Paper submitted to the annual meeting of the National Communication Association, Miami.

[3] Burgoon, J. K., Buller, D. B., Ebesu, A., & Rockwell, P. (1994). "Interpersonal deception: V. Accuracy in deception detection." Communication Monographs, 61, 303-325.

[4] Burgoon, J.K., Buller, D.B., Guerrero, L.K., Afifi, W.A., & Feldman, C.M. (1996). Interpersonal Deception: XII. Information management dimensions underlying deceptive and truthful messages. Communication Monographs, 63, 52-69.

[5] Burgoon, J., Marett, K., Blair, J. (in press). “Detecting Deception in Computer-Mediated Communication.” In J. George (Ed.), Social Issues of Computing.

[6] Levine, T., & McCornack, S. (1992). "Linking love and lies: A formal test of the McCornack and Parks model of deception detection." Journal of Social and Personal Relationships, 9, 143-154.

[7] Quinlan, J. R. (1993). C4.5. San Mateo, CA: Morgan Kaufmann Publishers.

[8] Spangler, W., May, J., & Vargas, L. (1999). “Choosing Data-Mining Methods for Multiple Classification: Representational and Performance Measurement Implications for Decision Support.” Journal of Management Information Systems, 16, 37-62.

[9] Vrij, A., Edward, et al. (2000). "Detecting deceit via analysis of verbal and nonverbal behavior." Journal of Nonverbal Behavior, 24, 239-263.

[10] Zhou, L. Twitchell, D., Qin, T., Burgoon, J. K. & Nunamaker, J. F., Jr. (2003). “An Exploratory Study into Deception Detection in Text-based Computer-Mediated Communication.” Proceedings of the 36th Annual Hawaii International Conference of System Sciences, Big Island. Los Alamitos, CA: IEEE.

[11] Zuckerman, M., DePaulo, B., & Rosenthal, R. (1981). “Verbal and nonverbal communication of deception.” In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol 14, pp.1-59). NY: Academic Press.

Appendix Decision tree output

Original data of Txt and Audio, 19 cues

```
Long_sentences <= 1
|   FK_grade <= 5.408709: 2 (9.0)
|   FK_grade > 5.408709
|   |   Affect <= 0
|   |   |   Average_syllables_per_word <= 1.44
|   |   |   |   LexComp <= 3.679012: 2 (6.0/1.0)
|   |   |   |   LexComp > 3.679012
|   |   |   |   |   total_flagged_errors <= 7: 1 (10.0/1.0)
|   |   |   |   |   total_flagged_errors > 7: 2 (2.0)
|   |   |   Average_syllables_per_word > 1.44: 2 (9.0/1.0)
|   |   Affect > 0: 1 (6.0/1.0)
Long_sentences > 1: 1 (6.0)
```

Original, no duplicated, 16 cues

```
Long_sentences <= 1
|   FK_grade <= 5.408709: 2 (9.0)
|   FK_grade > 5.408709
|   |   Affect <= 0
|   |   |   Average_syllables_per_word <= 1.44
|   |   |   |   LexComp <= 3.679012: 2 (6.0/1.0)
|   |   |   |   LexComp > 3.679012
|   |   |   |   |   Flagged <= 7: 1 (10.0/1.0)
|   |   |   |   |   Flagged > 7: 2 (2.0)
|   |   |   Average_syllables_per_word > 1.44: 2 (9.0/1.0)
|   |   Affect > 0: 1 (6.0/1.0)
Long_sentences > 1: 1 (6.0)
```

Original, significant, 6 statistically significant cues

```
FK_grade <= 5.408709: 2 (9.0)
FK_grade > 5.408709
|   Long_sentences <= 0
|   |   Sent_comp <= 34
|   |   |   Flagged_error <= 2
|   |   |   |   Average_words_per_sentence <= 13.7: 2 (3.0)
|   |   |   |   Average_words_per_sentence > 13.7: 1 (2.0)
|   |   |   Flagged_error > 2: 1 (6.0)
|   |   Sent_comp > 34: 2 (10.0/2.0)
|   Long_sentences > 0
|   |   Average_words_per_sentence <= 18: 2 (3.0)
|   |   Average_words_per_sentence > 18
|   |   |   Average_words_per_sentence <= 31: 1 (12.0/1.0)
|   |   |   Average_words_per_sentence > 31: 2 (3.0/1.0)
```

Txt only, with 16 no duplicated cues

```
Average_words_per_sentence <= 15.75
|   Sens&RM <= 1: 2 (10.0)
|   Sens&RM > 1
|   |   Average_words_per_sentence <= 12.52: 2 (2.0)
|   |   Average_words_per_sentence > 12.52: 1 (2.0)
Average_words_per_sentence > 15.75
|   Average_syllables_per_word <= 1.4: 1 (10.0)
|   Average_syllables_per_word > 1.4
|   |   Average_words_per_sentence <= 18: 1 (2.0)
|   |   Average_words_per_sentence > 18: 2 (2.0)
```

Txt, with 6 significant cues

```
Average_words_per_sentence <= 15.75
|   Sent_comp <= 22: 1 (3.0/1.0)
|   Sent_comp > 22: 2 (11.0)
Average_words_per_sentence > 15.75: 1 (14.0/2.0)
```