Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection

Christie M. Fuller
Department of MSIS
Oklahoma State University
Stillwater, OK 74078
Christie.fuller@okstate.edu

David P. Biros
Department of MSIS
Oklahoma State University
Stillwater, OK 74078
David.biros@okstate.edu

Dursun Delen
Department of MSIS
Oklahoma State University
Tulsa, OK 74106
Dursun.delen@okstate.edu

Abstract

Recent research has demonstrated the effectiveness of automated text-based deception detection. In this study, using a variety of data sets and common classification techniques, this has been shown to be an accurate technique. Previous results have shown the need to reduce the number of inputs to these models in order to prevent overfitting. While previous results have been promising, there is a need to improve accuracy and reduce the number of false positives. Using 5 classification models and 3 variable sets, we have achieved accuracy level of 76% in this study.

1. Introduction

In some fields, such as law enforcement, psychology, and human resources, it is necessary to be able to accurately assess the credibility of persons of interest. Though this skill may be important, most people are not very good at this task. Few people are able to detect deception at a rate better than chance [1]. Over the years, tools such as the polygraph, voice stress analyzer, Scientific Content Analysis (SCAN), and Behavioral Analysis Interview (BAI) have been developed to assist people with deception detection [2-8]. Aside from the polygraph, these tools are not very accurate. Further, they generally lack a theoretical basis. They are also subjective and can be invasive.

Automated text-based deception detection has been introduced as an alternative to these other methods. This technique has shown accuracy of up to 72% when trained on a very small sample of real-world lies [7]. This method has previously relied on the common classification methods of logistic regression, decision trees, discriminant analysis, and multilayer perceptron neural networks (MLP). Studies using these methods have found the MLP to be the most accurate classifier for linguistic-based cues [9, 10] . In addition to the MLP, four additional classification techniques not previously used in this research stream were chosen along with three cue sets in order to identify the model which maximizes the accuracy of classification using these cues.

2. Background

Attempts to detect deception have persisted for as long as people have been lying. Though this task is not new, humans have proven to not be very apt at determining veracity. An analysis summarizing the credibility assessment attempts of 23000 subjects found that average accuracy was only 54% [1]. Through the years, several tools have been developed to aid humans in making credibility assessments.

2.1 Previous Deception Detection Efforts

Previously, the primary automated deception detection tools were the polygraph and the computerized voice stress analyzer. The polygraph has shown accuracy from 72% to 92% in field studies [11-13]. Though it is probably one of the more accurate deception detection tools, the polygraph does have several drawbacks. The test can only be used in interviews with yes or no responses from the subject. The physiological measures used in a polygraph examination are automated, but these must be interpreted by a trained examiner. Of course, this also implies the availability of the necessary equipment and qualified examiner [8]. Voice stress analyzers (VSAs) were once touted as a replacement to the polygraph [6]. They can record any type of speech and are non-invasive. Unfortunately, like the polygraph, the results

1530-1605/08 $25.00 © 2008 IEEE

also depend on the person administering the exam. The VSA has been shown to only have an accuracy level about equal to that of chance [4]. Another issue for both of these techniques is that they largely lack a theoretical basis.

Numerous theories have been proposed to describe deception. These include Cue Leakage, Interpersonal Deception Theory (IDT), Information Manipulation Theory (IMT), Reality Monitoring (RM), Four Factor Theory, and the Self-Presentational Perspective. Cue Leakage proposes that deceivers leak cues that reveal either that deception is taking place or what the deception is [14]. Most known cues simply reveal that deception is occurring. According to IDT, deceivers strategically manipulate their message content in order to appear credible. Dimensions that may be manipulated include veracity, personalization, clarity, directness, and completeness [15, 16]. IMT examines deception using four conversational maxims which deceivers may violate: quality, quantity, relation and manner [17]. Reality monitoring suggests that imagined memories, associated with deception will differ from truthful, or actual memories, in the amount of perceptual and contextual information included in a statement or story [18]. The Self-Presentational Perspective summarizes cues to deception by placing them in five categories: liars are less forthcoming, tell less compelling stories, are less positive, more tense and fewer ordinary imperfections and unusual contents within their messages [3]. From these theories and deception research in general, a focus on cues to deception has emerged.

A recent meta-analysis identified 158 cues to deception [3]. This list includes both verbal and nonverbal cues to deception. Cues are those indicators or variables that can be observed and measured that are believed to be indicative of deception. While some of these cues are utilized in deception detection methods such as Content Based Criteria Analysis (CBCA) and SCAN, until recently they had not been used to automatically analyze text.

2.2 Automated Text-Based Deception Detection

Automated text-based deception detection has received some research attention as a possible supplement or replacement for other tools. Automated text-based text deception relies on text-processing and data mining software to build models for determining veracity. Past research has implemented as few as five and as many as 19 cues. Newman and colleagues used five cues as inputs to a logistic regression model, achieving an accuracy level of 61% [5]. Bond and Lee expanded upon this research by adding variables associated with Reality Monitoring (RM). Again using logistic regression, overall accuracy increased to 71.1% [19].

Zhou et al. studied a larger cue set drawn from IDT, IMT, RM, CBCA and other sources. This study implemented discriminant analysis, logistic regression, decision tree, and neural network algorithms. An implementation of C4.5 was selected for the decision tree and a standard multilayer perceptron was used for the neural network. When cues were reduced to those identified as important in the first iteration of training and testing the models, accuracy reached 80.2% [10]. This study, as well as those of Newman and Bond utilized 'mock lies'. A later study relied on the overall set of cues identified by Zhou et al. to classify a small real world sample (n = 18) with an overall accuracy of 72% [7]. This real-world study implemented an MLP neural network since it was the most accurate technique in the study that preceded it.

3. Methodology

This study utilized message feature mining to analyze a high-stakes real-world data sample. For the feature extraction phase, two programs were used to quantify cues belonging to three cue sets. For the classification step of message feature mining, five classification algorithms were implemented on each of the cue sets.

3.1 Data Sample

This study utilizes a sample of person of interest statements collected from two military bases. Person of interest statements are official written reports completed by a person involved in a crime. This person may be either a suspect or witness to the crime. Once completed, law enforcement personnel use the statements as a starting point for their investigations. All statements used in this study were classified as truthful or deceptive by base law enforcement personnel. Criteria such as confession, case resolution, conflicting evidence, verification or contradiction by a witness in law enforcement were used to classify the statements. For example, in one case, a military member who was driving erratically on base was pulled over for DWI. Before the police officer could get

from his car to the suspect's vehicle, the suspect and his wife, who was riding in the passenger seat, switched places (military members receive much more severe punishment on bases than civilians). Both wrote statements explaining that the wife was driving. However, later when the investigator reviewed the police cruiser's dashboard video, he noticed the switch and charged the military member with DWI, an offense punishable by loss of rank, fines, loss of driving privileges, and confinement.

These statements represent a unique opportunity to examine deception in a real world high stakes context. A high stakes context is one in which there are actual consequences associated with being judged truthful or deceptive [20]. Not only do military personnel who are found guilty of a crime receive punishment under the Uniform Code of Military Justice (UCMJ), but they could also be prosecuted for "false official statement" if found to have lied on a person of interest statement; another infraction of the UCMJ. Deception research has mostly been conducted in the laboratory using student subjects. The need for high stakes and real world samples has been previously noted [18, 25, 26].

Over 370 statements were collected over an 18-month period. Typed statements and those written on behalf of someone else were excluded from the current sample. While typed statements may be very similar to those handwritten, to maintain consistency in the sample, these were excluded. Further, the typed statements were nearly all labeled as truthful. If typed statements were only used for the truthful class, this might introduce bias if there were any difference between these and the written statements. This resulted in a final sample of 366 statements, including many more truthful than deceptive statements. The sample was then balanced to achieve better accuracy and generalizability [21]. The final sample included 79 deceptive and 84 truthful statements. All statements had to be transcribed for processing. A set of standardized procedures was developed to transcribe the data. These transcription procedures are detailed by Twitchell et al [22].

3.2 Message Feature Mining

The data was processed using a procedure known as 'message feature mining' [23]. The first major step of message feature mining is processing the text to extract the cues. Essentially, the selected program processes a statement to determine which cue or cues each word represents. It then calculates the number or proportion of words assigned to each cue. These values serve as the quantitative inputs for classification. For example, the program may first use a part-of-speech tagger to identify the verbs in each statement. Then, the program will count the number of verbs present in the statement. The figures below illustrate this process is accomplished in one of the two programs used, GATE.

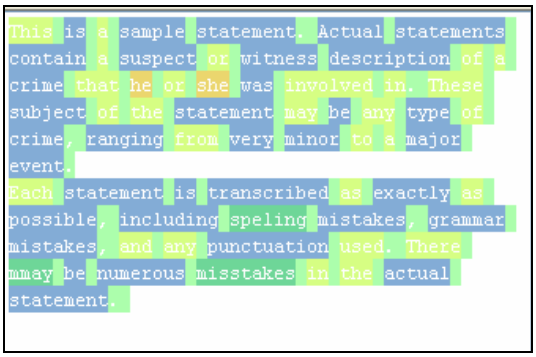

Figure 1: Identification of Cues within a Statement

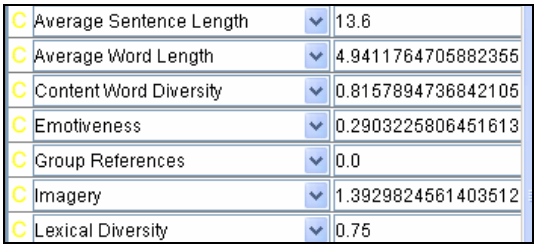

Figure 2: Quantifying the Cues Identified In a Statement

The diagram below (see Figure 3) illustrates the message feature mining cycle. This shows the cycle used here: collecting data, transcribing that data, selecting cues to use for classification, then identifying and quantifying those cues. Finally classification models are trained and tested. As is common in data mining problems, once the models are built, the process might be repeated with new data and new features.

The message feature mining process captures elements of text mining and typical data mining. Text mining (also known as text data mining [24] and knowledge discovery in textual databases [25]) can be described as the process of deriving novel information from a collection

of texts (also known as a corpus) such as documents, Web pages, short statements, article abstracts, etc. By novel information, we mean associations, hypotheses or trends that are not explicitly present in the text sources being analyzed. Even though text mining is considered a part of the general field of data mining, it differs from regular data mining.

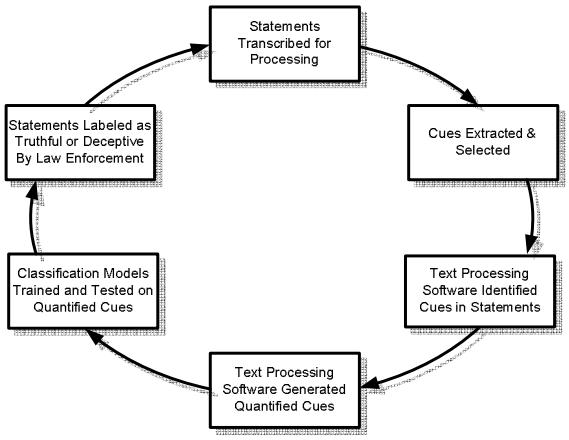

Figure 3: Message Feature Mining Model Development Cycle

The main difference is that in text mining, the patterns are extracted from natural language text rather than from structured databases of facts. Databases are designed for programs to process automatically; text is written for people to read. We do not have programs that can "read" and "understand" text (at least not in the manner human beings do). Furthermore, despite the phenomenal advances achieved in the field of natural language processing [26], we will not have such programs for the foreseeable future. Many researchers think it will require a full simulation of how the mind works before we can write programs that read and understand the way people do [27]. So what does text mining do? On the most basic level, it numerically represents (i.e., numericizes) the unstructured text sources and then, using data mining tools and techniques, extracts patterns from them.

To achieve this numerical representation of the text, two text processing tools were used: Agent 99 Analyzer (A99A) and Linguistic Inquiry and Word Count (LIWC). A99A is part of a larger system for deception detection developed at the University of Arizona known as Agent 99 [28, 29] [30, 31]. A key component of A99A is the Generalized Architecture for Text Engineering (GATE) [30]. GATE incorporates features such as tokenizers and part-of-speech taggers which enable processing of several cues. It is possible to load extra processing tools and dictionaries to manage additional cues. LIWC includes 74 dictionaries for processing emotional and cognitive content of text. Similar to Agent99A, additional cue dictionaries can readily be incorporated into LIWC [31].

3.3 Classification Algorithms

The second half of the message feature mining process includes training and testing classification models. The classifiers are implemented using ten-fold cross validation to reduce any bias in the results due to the selection of particular train and test partitions [32].

Five different classification models were implemented using Waikato Environment for Knowledge Analysis (WEKA) [33]. While there are numerous additional techniques that could be implemented, using five model types should be sufficient to provide insight into the potential for increased accuracy by changing the type of model. The first was the common multilayer perceptron neural network. This is a system of weighted nodes typically consisting of three layers: input, hidden and output. Through training, the weights are adjusted to maximize classification accuracy [21]. The second classifier was the radial basis function network. Radial basis functions (RBF) are a type of artificial neural network. The layers of an RBF use a radial activation function. RBF networks can model complex relationships with the input, hidden, and output layers that would require additional MLP layers [34].

The third classifier implemented was the naïve bayes algorithm. This algorithm makes classifications based on accumulated evidence of the correlation between a given variable and the other variables present in the model [35]. The fourth algorithm selected was random forest. This technique classifies using a combination of trees. Random forest uses randomly selected values to grow the tree [36]. Finally, the WEKA SMO implementation of the support vector machine algorithm was used. This classifier combines both linear and nonlinear techniques to find the hyperplanes that maximize the differences between output classes [33]. Each of these five classification algorithms was

combined with three different sets of cues to achieve maximum accuracy.

3.4 Feature Selection

Three cue sets were used in training and testing the models. A previous study which analyzed four different cue sets showed that reducing the number of model inputs greatly reduced overfitting, therefore creating more generalizeable results that may extend to more data of the type used here and other data sets [9]. This is also consistent with the results found by Zhou et al [10].

Two of the cue sets from previous research were used here. Based on neural network heuristics requiring five to ten examples per network weight and the size of the data set, we decided to use a feature selection procedure to select eight cues to form the first cue set [37]. From the overall list of over thirty cues that have been used in automated text based deception detection, a feature selection procedure based on the f-statistic was executed to rank the variables. The top eight variables were retained for the first cue set.

The second set of cues was selected based on a set of recently validated deception constructs. The constructs were validated using confirmatory factor analysis. Eleven cues were determined to be reliable indicators of four constructs: quality, specificity, affect, and uncertainty (see Table 1 for details). While data mining does not require a strict link to theory, if results show that the cues related to these constructs can be accurately used to determine veracity, it may provide further validation of the constructs.

Table 1: Validated Text-Based Deception Constructs

Construct	Brief Description	Indicator
Quantity	Length of message	Word Quantity, Sentence Quantity
Specificity	Amount and type of details in the message	Sensory Ratio, Temporal Ratio, Bilogarithmic Type-Token Ratio
Affect	Emotions present in the message	Activation, Imagery, Pleasantness
Uncertainty	Relevance, directness, and certainty of message	certainty terms, generalizing terms, tentative terms

While the first two cue sets have been used before and were shown to perform well, an additional cue set was selected to further explore the potential of feature selection.

The algorithms used in this study incorporate non-linear forms of analysis. Therefore, as an alternative to the first feature selection technique which is based on a linear relationship between the cues and the dependent variable, the third cue set will be based on a method that incorporates non-linear relationships, the multilayer perceptron. The third cue set resulted from sensitivity analysis on a multilayer perceptron built with thirty text-based deception cues derived from several previous deception studies. Again based on neural network heuristics, the eight most important variables were drawn from this network. The three cue sets are summarized in Table 2.

Table 2: Cue sets used for classification

Cue Set	Cues
Set 1	Word Count, Exclusive terms, 3rd person pronouns, Lexical diversity, Verb quantity, Content word diversity, Modifiers, Sentence Quantity
Set 2	Word Count, Sensory Ratio, Time Ratio, Generalizing Terms, Sentence Quantity, Certainty Terms, Tentative Terms, Bilogarithmic Type-Token Ratio, Activation, Imagery, Pleasantness
Set 3	Bilogarithmic Type-Token Ratio, Imagery, Average Sentence Length, Passive Verbs, 1st Person Singular Pronouns, 1st Person Plural Pronouns, Exclusive Terms, Lexical Diversity

Another solution to the overfitting seen in large data sets would be to increase sample size. However, difficulty in establishing ground truth, or actual truthful or deceptive classification, is inherent to deception research. Previous sample sizes in deception research have not exceeded 200, and it is anticipated that these limits on sample size will continue [11, 18].

Two key performance measures will be evaluated for each model and data set. The first is overall classification accuracy, or the number of both truthful and deceptive statements accurately classified. The second measure will be

the proportion of false positives. This evaluates the proportion of truthful statements incorrectly classified as deceptive statements. Research into the polygraph has demonstrated the need to evaluate both measures, as there is often a tradeoff between the two [2]. There is no specific guideline as to what is an acceptable tradeoff between false positives and overall accuracy. Rather, decision makers must assess this tradeoff for each particular sample and situation.

4. Results

The first measure evaluated for the models was overall accuracy. The percentage reported is that for the overall accuracy in classifying the test data partitions for cross validation. Overall, Cue Set 2, the set of cues based on the text-based deception constructs had the poorest performance on three of five algorithms studied. There are mixed results for the other two cue sets, depending on the classifier, though cue set three may be preferred as it showed the highest accuracy for any cue and algorithm combination, as well as relatively consistent performance across classifiers. While the Random Forest trained and tested with the third cue set had the best individual result, the SVM, Naïve Bayes and MLP seem to perform well consistently across data sets.

Table 3: Overall Accuracy %

	Cue Set 1	Cue Set 2	Cue Set 3
MLP	74.85	73.62	73.01
RBF	75.46	69.94	74.23
Random Forest	69.33	75.46	76.07
Naïve Bayes	74.23	73.01	74.85
SVM	74.85	73.62	74.85

As noted, polygraph research has demonstrated the need to evaluate the false positive rate along with overall accuracy, as there may be a tradeoff between the two measures. The best overall false positive rate was for the RBF algorithm implemented on the first cue set. Additionally, the Random Forest performed relatively well across the cue sets, having the lowest rate for two of the three sets. The worst overall performance was for the RBF built using the second cue set. There is not a clear pattern showing which cue set produces the fewest false positives, though the third cue set did not have any extremely high or low values, as were found for the other two cue sets. As the third cue set performed relatively well in terms of overall accuracy and false positive rate, it appears to be the best set of cues to use. The optimal combination of model and data set is the Random Forest algorithm with the third data set. This combination shows the best overall accuracy and one of the lowest false positive rates. It cannot be concluded from study of a single data sample of these cue sets and algorithms which combination or combinations of these should be implemented in any particular situation. However, the identification in this study of cues and algorithms that appear to have superior performance can guide future efforts to determine this ideal combination.

Table 4: False Positive %

	Cue Set 1	Cue Set 2	Cue Set 3
MLP	28.57	21.43	27.38
RBF	14.29	40.48	26.19
Random Forest	25.00	20.24	19.05
Naïve Bayes	32.14	32.14	33.33
SVM	29.78	30.95	30.95

5. Discussion and Conclusion

Previous research has shown that automated text-based deception detection has the potential to accurately determine veracity. Limited work has been done with this technology in deception detection studies, and with automated text-based deception detection in particular, using real world samples. Past studies have shown the MLP to be superior to techniques such as logistic regression and decision trees.

This study explored whether alternative classification algorithms could further improve the accuracy of this technique. Previous text analysis studies have failed to explore false positive rates though polygraph research shows that this measure may suffer when overall accuracy increases [2]. The results shown here illustrate that this is not necessarily the case. Not surprisingly, this study also shows that performance for the algorithms varies with the inputs used.

Limitations on the size of the data set likely affected the accuracy that could be achieved. This will likely continue in the stream of research. Cue set three seemed to be the best set of inputs. These were selected based on a previous MLP model. Since the algorithms used in this study all incorporate nonlinear techniques,

it is not surprising that they performed best on a cue set selected with a nonlinear method. Though the algorithms implemented here seem to improve accuracy, they do not readily provide an explanation of the importance of each cue or how each is related to the dependent variable, as a decision tree or logistic regression would. If these algorithms are to continue to be used, sensitivity analysis would need to be explored.

As this study has successfully demonstrated classifiers not previously used in deception research may show promise, additional algorithms might be explored in future studies. There are many more feature selection techniques that could also be implemented. This study also will be replicated in additional domains and cultures to determine the generalizability of the findings. A primary goal of this research is to provide a prototype tool to be implemented by law enforcement. The current accuracy may not be high enough to warrant implementation in the field. As noted above, incorporating additional data, refined cue sets and additional classifiers are anticipated to improve the current results. As the accuracy of automated text-based deception detection approaches maximum of 92% field accuracy of the polygraph, such an implementation can be achieved in the near future.

This study demonstrated that text-based deception detection analysis is indeed a promising technology to aid in credibility assessment. Through the use of real world data derived from conditions of high stakes consequences, we determined that a reduced cue set combined with the Random Forest algorithm produced commendable results. This tool was studied in a law enforcement domain where there may be severe consequences and penalties associated with credibility assessment decisions. Therefore, a tool implemented to aid investigations must be accurate. By determining the best classifier and the best cues to use with that classifier, the accuracy of automated text-based deception detection can be maximized.

6. References

[1]. Bond, C.F. and B.M. DePaulo, "Accuracy of Deception Judgments", Personality and Social Psychology Reports, 2006. pp. 214-234.

[2]. Nation Research Council, The Polygraph and Lie Detection, Committee to Review the Scientific Evidence on the Polygraph, The National Academies Press: Washington, DC, 2003.

[3]. DePaulo, B.M., J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to Deception", Psychological Bulletin, 2003, pp. 74-118.

[4]. Gamer, M., H.G. Rill, G. Vossel, and H.W. Godert, "Psychophysiological and Vocal Measures in the Detection of Guilty Knowledge", International Journal of Psychophysiology, 2006, pp. 76-87.

[5]. Newman, M.L., J.W. Pennebaker, D.S. Berry, and J.M. Richards, "Lying Words: Predicting Deception from Linguistic Styles", Personality and social psychology bulletin., 2003, pp. 665-675.

[6]. Rice, B., "The New Truth Machines", Psychology Today, 1978, pp. 61-64, 67,72, 74, 77-78.

[7]. Twitchell, D., D.P. Biros, N. Forsgren, J. Burgoon, and J.F. Nunamaker Jr, "Assessing the Veracity of Criminal and Detainee Statements: A Study of Real-World Data", in 2005 International Conference on Intelligence Analysis, 2005.

[8]. Vrij, A., Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice, John Wiley & Sons, New York, 2000.

[9]. Fuller, C., D.P. Biros, and R.L. Wilson, "Decision Support for Determining Veracity Via Linguistic Based Cues", Under Review, 2007.

[10]. Zhou, L., J.K. Burgoon, D.P. Twitchell, T.T. Qin, and J.F. Nunamaker, "A Comparison of Classification Methods for Predicting Deception in Computer-Mediated Communication", Journal Of Management Information Systems, 2004, pp. 139-165.

[11]. Honts, C.R. and D.C. Raskin, "A Field Study of the Validity of the Directed Lie Control Question", Journal Of Police Science And Administration, 1988, pp. 56-61.

[12]. Raskin, D.C., "Methodolgical Issues in Estimating Polygraph Accuracy in Field Applications", Canadian Journal of behavioral Science, 1987, pp. 389-404.

[13]. Suzuki, A., K. Ohnishi, K. Matsuno, and M. Arasuna, "Amplitude Rank Score Analysis of Gsr in the Detection of Deception: Detection Rates under Various Examination Conditions", Polygraph, 1979, pp. 242-252.

[14]. Ekman, P., Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage, WW Norton & Company, New York, 1985.

[15]. Buller, D.B. and J.K. Burgoon, "Interpersonal Deception Theory", Communication Theory, 1996, pp. 203-242.

[16]. Buller, D.B., J.K. Burgoon, A. Buslig, and J. Roiger, "Testing Interpersonal Deception Theory: The Language of Interpersonal Deception", Communication Theory, 1996, pp. 268-289.

[17]. McCornack, S.A., "Information Manipulation Theory", Communication Monographs, 1992, pp. 1-16.

[18]. Johnson, M.K. and C.L. Raye, "Reality Monitoring", Psychological Review, 1981, pp. 67-85.

[19]. Bond, G.D. and A.Y. Lee, "Language of Lies in Prison: Linguistic Classification of Prisoners' Truthful and Deceptive Natural Language", Applied Cognitive Psychology, 2005, pp. 313.

[20]. Frank, M.G. and P. Ekman, "The Ability to Detect Deceit Generalizes across Different Types of High-Stake Lies", Journal of Personality and Social Psychology, 1997, pp. 1429-1439.

[21]. Berry, M.J.A. and G.S. Linoff, Data Mining Techniques, 2 ed. Wiley Publishing, Indianapolis, Indiana, 2004.

[22]. Twitchell, D.P., D.P. Biros, M. Adkins, N. Forsgren, J.K. Burgoon, and J.F. Nunamaker Jr, "Automated Determination of the Veracity of Interview Statements from People of Interest to an Operational Security Force", in 39th Annual Hawaii International Conference on System Sciences, 2006.

[23]. Adkins, M., D. Twitchell, J.K. Burgoon, and J.F. Nunamaker Jr, "Advances in Automated Deception Detection in Text-Based Computer-Mediated Communication", in Enabling Technologies for Simulation Science VIII, SPIE, Orlando, FL, USA, 2004.

[24]. Hearst, M.A., "Untangling Text Mining", in 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, 1999.

[25]. Feldman, R. and I. Dagan, Knowledge Discovery in Textual Databases (Kdt), in Knowledge Discovery and Data Mining. 1995, p. 112-117.

[26]. Manning, D.M. and H. Schutze, Foundations of Statistical Natural Language Processing, MIT Press, Cambridge, MA, 2003.

[27]. Hearst, M., What Is Text Mining? SIMS, UC Berkeley, 2003.

[28]. Cao, J., J.M. Crews, M. Lin, J. Burgoon, and J.F. Nunamaker, "Designing Agent99 Trainer: A Learner-Centered, Web-Based Training System for Deception Detection", in Lecture Notes in Computer Science: Proceedings of Intelligence and Security Informatics: First Nsf/Nij Symposium, Tucson, Az, USA, June 2-3, 2003, Springer. p. 358-365.

[29]. Zhou, L., J.K. Burgoon, J. Nunamaker, Jay F., and D.P. Twitchell, "Automated Linguistics Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communication: An Empirical Investigation", Group Decision and Negotiation, 2004. pp. 81-106.

[30]. Cunningham, H., "Gate, a General Architecture for Text Engineering", Computers and the Humanities, 2002. pp. 223-254.

[31]. Pennebaker, J.W. and M.E. Francis, *Linguistic Inquiry and Word Count: Liwc 2001*. 2001, Erlbaum Publishers: Mahwah, NJ.

[32]. Weiss, S.M. and C.A. Kulikowski, Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufman Publishers, Inc., San Mateo, California, 1991.

[33]. Witten, I.H. and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java. Morgan Kaufman, San Francisco, 2000.

[34]. Bors, A.G., Introduction of the Radial Basis Function, Department of Computer Science, University of York: York, UK.

[35]. Tang, Z. and J. MacLennan, Data Mining with Sql Server 2005, Wiley Publishing, Inc., Indianapolis, Indiana, 2005.

[36]. Breiman, L., "Random Forests", Machine Learning, 2001, pp. 5-32.

[37]. Sarle, W., "What Are Cross-Validation and Bootstrapping?" 2004 [cited 2005; Available from: http://www.faqs.org/faqs/aifaq/neural-nets/part3/section-12.html.