AUTOMATED SCORING OF INTEGRATIVE COMPLEXITY USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

by

AARDRA KANNAN AMBILI

(Under the Direction of Khaled Rasheed)

ABSTRACT

Conceptual/Integrative complexity is a construct developed in political psychology and clinical psychology to measure an individual's ability to consider different perspectives on a particular issue and reach a justifiable conclusion after consideration of said perspectives. Integrative complexity (IC) is usually determined from text through manual scoring, which is time-consuming, laborious and expensive. Consequently, there is a demand for automating the scoring, which could significantly reduce the time, expense and cognitive resources spent in the process. Any algorithm that could achieve the above with a reasonable accuracy could assist in the development of intervention systems for reducing the potential for aggression, systems for recruitment processes and even training personnel for improving group complexity in the corporate world. The proposed approach produced classification accuracies ranging from 75% to 83%, which is a first in the literature for automated scoring of integrative complexity.

INDEX WORDS: integrative complexity, text classification, semantic similarity, support vector machines

AUTOMATED SCORING OF INTEGRATIVE COMPLEXITY USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

by

AARDRA KANNAN AMBILI

B.Tech, Amity University, 2011

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2014

© 2014

Aardra Kannan Ambili

All Rights Reserved

AUTOMATED SCORING OF INTEGRATIVE COMPLEXITY USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

by

AARDRA KANNAN AMBILI

Major Professor: Khaled Rasheed

Committee: Walter D. Potter
Adam Goodie

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

AUTOMATED SCORING OF INTEGRATIVE COMPLEXITY USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

Aardra Kannan Ambili

November, 20, 2014

DEDICATION

I dedicate this thesis to my mother and my father. Without their constructive criticism, infectious joy and boundless love, I may not have become the person that I am today.

ACKNOWLEDGEMENTS

I deeply appreciate and acknowledge the help and support provided to me by Dr. Khaled Rasheed. His sagely advice, sparkly humor and loads of positive energy have been extremely helpful. I would also like to thank all my committee members for their mentorship and timely advice. I would like to take this opportunity to appreciate and acknowledge the time and effort taken by all my teachers who have taught me since kindergarten through high school till graduate school. Finally I would like to thank all my friends for their valuable support and companionship.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Integrative Complexity is one of "the most-researched operations of the complexity of human thought" (Conway, 2008). Integrative complexity (IC) is used as a measure of the intellectual style used by individuals or groups in processing information, problem solving, and decision making. It is a measure of the structure of one's thoughts, regardless of the contents. The examples below (Statement 1 & Statement 2, both have low IC) are given to show that two text samples of the same topic but with competing views could have the same IC score. Integrative Complexity scorers are therefore required to ignore the content of the text samples and focus on the cognitive strategies involved in formulating the structure of the samples.

Statement 1:

> *"The physical poses of Hatha yoga are physically challenging. But the breathing exercises help to balance the energy in your body. The combination of breath work and physical exercises prepare your body for spiritual growth."*

Statement 2:

> *"Kundalini yoga focuses on breathing and meditation. For kundalini energy to flow freely, your body must be in harmony. The way to achieve this harmony comes with practicing Hatha yoga. However over-emphasis of Hatha may lead to injury."*

Low levels of integrative complexity are characterized by rigid and simplistic perspectives on events. Such narrow simplistic thinking will typecast a single point of

view as the correct one and view all other perspectives as illegitimate, flawed, or ridiculous (Suedfeld, Tetlock, & Streufert, 1992). In contrast, thinking characterized as having high levels of IC acknowledge multiple perspectives on an issue and further recognize how these divergent viewpoints connect and contribute to the conclusion. Complex thinkers are more resistant to the influences of singular events and are less susceptible to suggestion and manipulation than simple thinkers. They may also be better able to accommodate stress (Suedfeld & Piedrahita, 1984) and are also better at predicting the behaviors of others and are less prone to projection (Bieri, 1955).

1.1 Theoretical Background

The psychological theorization that there does exist stable differences in the cognitive processes among individuals became a salient theory in the 1960s. These differences could exist in the way we process information, evaluate data and in making decisions with the available data. This theorization became possible due to the precursor ideas that already existed back then (Leary & Hoyle, 2009). One of the prominent ones that contributed to the hypothesis was the idea that characteristics like intelligence and authoritarianism were found to be unchangeable traits for decades. Later on, with the subsequent transformation of psychology with the cognitive revolution, the idea that thinking could have the same characteristics was considered a possibility.

Subsequently, several theories were proposed to capture these distinct differences and on why these differences emerge in cognition. "*Cognitive Complexity subsumes a variety of specific approaches, but the general foundation is the idea that a nonhomeostatic variable can be identified that involves how people deal with the flow of information that impinges on them throughout their lives*" (Leary et al. 2009). That stable differences exist

in the way individuals react when information flow becomes too meager or too lavish was the hypothesis behind this theory. When information is too meager, individuals tend to selectively prefer certain parts of the information or generate their own, whereas when information is too lavish, they tend to reject certain parts of the information, while preferring others, or clump information into categories to reduce the number of distinct portions, or try to merge information coming from different inputs while ignoring difference to obtain a unified view. The three approaches to cognitive complexity are: the need for closure, the need for cognition and conceptual complexity.

Conceptual complexity theory (Schroder, Driver, & Streufert, 1967), like other cognitive complexity approaches, asserted that ways of processing information constitute an individual difference dimension that can be usefully considered a personality trait. One of the prominent derivatives of conceptual complexity theory was the conceptual/integrative complexity construct, whose problem of automation forms the research focus of this thesis.

The conceptual/integrative complexity construct is also a descendant of Kelly's (1955) personal construct theory. The closest relatives of integrative complexity are cognitive complexity (Bieri, 1971) and cognitive structure (Scott, Osgood, Peterson & Scott, 1979; Suedfeld, P., Tetlock, P. E., & Streufert, S. 1992). Bieri's inception of the concept of cognitive complexity-simplicity was based on the theory of psychological constructs by Kelly (1955).

Although the framework for Integrative Complexity was partly derived in the 1960's from Pidget's theory, the framework got its most comprehensive statement in a book by Harold Schroder, Michael Driver and Siegfried Streufert (Schroder, Driver & Streufert

1967). It was during the 1980's and 1990's that the framework obtained its popularity and significance through a large number of publications in psychology issues. Consequently, the framework went through significant modifications and developments and was subsequently used in empirical studies on political psychology, in particular by Philip Tetlock and Peter Suedfeld and their associates (Coren & Suedfeld, 1995; Suedfeld & Bluck, 1988; Suedfeld & Piedrahita, 1984; Suedfeld & Tetlock, 1977). Jordan (1998) in his paper, explains the reasons for Integrative Complexity's popularity as due to the stringent formulation of the framework, its accessibility to people less familiar with psychological theory and its ease of use in empirical studies.

Suedfeld (2010) demarcates between conceptual complexity and integrative complexity in two ways: theoretical and methodological. Theoretically, Integrative Complexity is used to measure Differentiation and Integration at a particular point of time. The level of integrative complexity expressed by an individual at any point is thought to be jointly determined by personality (conceptual complexity), by other internal factors (e.g., fatigue, emotional arousal), and by external situational factors such as danger and time pressure. In the Cognitive Manager model (Suedfeld, 2010), conceptual complexity is regarded as the trait component (as a part of the individual's personality, something that is unchangeable throughout the person's life) and integrative complexity is regarded as the state component (pertaining to the individual's complexity at that point of time, which could be easily influenced by stress, environment, disease etc.). The second difference is methodological: in the measurement of IC, the material being measured is archival, produced during day-to-day "real life" activities of the source rather than in the context of research. Thus the measure is non-intrusive. Since the emphasis is on Integrative complexity's dynamic

qualities, much of the research is focused on the effects of the changes on information processing and decision making.

One of the earliest theories and closest relative to Integrative Complexity, Cognitive Complexity (as conceptualized by Kelly and Bieri) is often correlated with Integrative Complexity in empirical studies for its ease of measurement. Differentiation, one of the integral parts of IC, is essentially equivalent to the variable of cognitive complexity. Therefore, while the measurement of Cognitive Complexity could be correlated to the measurement of Integrative Complexity, they are not equivalent.

In 1967, the term Cognitive Complexity (CC) was re-defined as a characteristic of information processing in cognitive systems (Schroder, Driver, & Streufert, 1967), while IC reflected the information processing capabilities of the individual. CC is the complexity of the knowledge structures in a cognitive system, and it describes the sophistication of those cognitive structures that are used for organizing and storing cognitive contents. High CC reflects a flexible and adaptive orientation in information processing. In a cognitive system characterized by high CC information processing is defined by the use of many constructs with many relations among them.

Integrative complexity has been widely researched in the domain of political psychology and has also received some attention within clinical psychology (Bruch, Juster, & Kuethe, 1985; Liotti, 1987; Raz-Duvshani, 1986; Strohmer, Biggs, Haase, & Bruch, 1983). Although integrative complexity has not been widely used in the clinical literature, it is undoubtedly relevant for clinical research and practice.

1.2 Scoring of Integrative Complexity

Originally, the paragraph completion test (PCT) and the picture story exercise (PSE) were used to measure IC from text. The PCT was recommended by Suedfeld, Tetlock & Streufert (1992). Participants were instructed to complete prepared sentences, such as "When I am confused…" or "When a friend acts differently…". Then expert coders were asked to code the Integrative Complexity of their responses. In PSE, participants were asked to draw descriptions about ambiguous pictures which was subsequently followed by coding by expert coders (Tetlock, Peterson, and Berry, 1993).

The most preferred method for measuring IC from text is through content analysis. Content analysis is "a research method that uses a set of procedures to make valid inferences from text" (Weber, 1990, p. 9). Research on integrative complexity represents a "best-practice" case of how content analysis can be used to assess complex cognitive styles in individuals. The bulk of research on integrative complexity uses content analysis, and it illustrates clearly many of the benefits of this approach. In political psychology and other related fields, the systematic scoring of archival materials is a core methodology, mainly because of the cumbersome and adverse nature of obtaining access to data or text written by high level political leaders or other personnel. The relationship between aspects of cognitive processing and decision making and several other variables could be investigated through such methods. Samples are extracted from verbal/written/recorded output of the person and then after selecting the variable of interest, detailed scoring methods are used to extract quantitative data. For scoring IC, there might be detailed scoring manuals for measuring the presence or level of the variable. The procedure is methodologically rigorous for a number of reasons. Initially, identifying material is usually removed from the samples

prior to scoring, then manual scorers are qualified through painstaking training courses, then their inter-scorer reliabilities are tested for each study and repeatedly over time, and then data can be evaluated by normal inferential statistics. Thematic Content Analysis can be adapted to the variables that can be measured. Scoring systems are available for a wide range of personality, motivational, attitudinal, and cognitive and other dimensions. (See Gottschalk, 1995; Smith, 1992).

Content analysis has allowed researchers to assess the Integrative complexity of individuals who are typically not available for psychological study. Some of the subjects who were studied using IC scoring were deceased individuals (e.g., pre-Civil War individuals, public figures), public figures (e.g US senators, US presidents) etc. (Porter & Suedfeld, 1981). The measurement of IC among some subjects could also be used to design intervention programs for certain variables. For example, Lint and Savage (2013) demonstrated that an intervention program designed, to prevent violent extremism among young UK Muslims, was able to increase value spread and integrative complexity. At the end of the program, it was noted that in group discussions and in written responses, the conflict resolution style shifted towards collaboration and compromise. There have also been studies that analyzed autobiographical texts of authors and determined that Integrative Complexity changed in accordance with stressors. Studies have determined IC of leaders in countries other than the US (e.g., Tetlock & Boettger, 1989; Liht & Savage, 2013), other cultures and across historical periods (e.g., Suedfeld & Bluck, 1988; Suedfeld et al., 1977;Tetlock, 1985). Content analysis also allows researchers to measure IC levels of groups, (Gruenfeld & Hollingshead, 1993; Tetlock, 1979; Walker & Watson, 1994) and

even entire governments and countries (Suedfeld & Tetlock, 1977; Tetlock & Boettger, 1989).

Conceptual/integrative complexity can be scored from almost any verbal material: books, articles, fiction, letters, speeches and speech transcripts, video and audio tapes, and interviews. For example, Integrative Complexity has been measured from speeches (Suedfeld, "Tetlock, & Ramirez, 1977; Tetlock, 1983), diplomatic documents (Suedfeld & Tetlock, 1977), interview transcripts (Tetlock, 1984), policy statements (Tetlock, 1983, 1985; Tetlock & Boettger, 1989; Walker & Watson, 1994), and personal letters (Porter & Suedfeld, 1981; Suedfeld & Bluck, 1993).

It is therefore a numerical score which measures the extent to which an individual demonstrates certain inclinations on consideration of events or issues. It is measured using two cognitive structural variables: Differentiation and Integration. Differentiation relates to the capacity of individuals to adopt and to apply a variety of perspectives on gauging an issue. This variable is indicative of an individual's ability to understand and appreciate another individual's views and make informed decisions. Integration refers to the capacity of individuals to recognize interweaving connections and similarities across perspectives. Hence, when integrative complexity is low, individuals tend to form simple and rigid attitudes and perceptions and are often unable to appreciate or absorb the views of other individuals. (Suedfeld, Tetlock & Streufert, 1992).

Many researchers find it difficult to accept the validity of results returned from content analysis, because of connotations of bias and fuzziness. Some researchers tend to shy away from content analysis as the coding can be painstaking (Lee & Peterson, 1997).

Considering these factors, automating the integrative complexity scoring process would mean that researchers can save considerably on time, expenses and cognitive energy.

Integrative complexity scoring can be performed on whole documents or paragraphs. Typically we could consider a paragraph as a single unit. IC is scored on a 1-7 point scale where each point signifies specific levels of integrative complexity, i.e. specific levels of differentiation and integration. (1 = low differentiation and almost non-existent integration, while 7 = high differentiation and high integration). Without a non-zero amount of differentiation, it is impossible to expect any text sample to show any integration because for integration to be present, differing perspectives must exist (for more information, see Baker-Brown, Ballard, Bluck, de Vries, Suedfeld, & Tetlock, 1992). Expert coders who are appointed to score/judge text samples receive appreciable training to gauge IC. These coders who are given the coding scheme and the definitions of differentiation and integration beforehand, are given 8 sets of paragraphs for scoring IC. In order to be deemed reliable, the correlation between their rating and accepted ratings of the samples must exceed a set threshold. The accepted criterion is 0.80 (e.g., Feist, 1994). Team integrative complexity can be thought of as the intellectual flexibility of a team, which corresponds to the capacity of teams to recognize differing perspectives on an issue, and their internal flexibility to adjust their notions in response to additional information. In this respect, the intellectual flexibility of a team is equivalent to team integrative complexity. The former is measured using groups dynamics Q sort (e.g., Peterson et al., 1998).

Table 1: Examples of Integrative Complexity from Senatorial Speeches on Abortion.[1]

Instances	***Comment***
"Abortion is a basic right that should be available to all women. To limit a woman's access to an abortion is an intolerable infringement on her civil liberties. Such an infringement must not be tolerated. To do so would be to threaten the separation of church and state so fundamental to the American way of life."	IC score of 1. Low Differentiation and Low Integration. This text indicates low levels of IC. There is consideration of a single perspective, while viewing all others as illegitimate.
"Many see abortion as a basic civil liberty that should be available to any woman who chooses to exercise this right. Others, however, see abortion as infanticide."	Medium Score of 3. Moderate to high differentiation, but no integration. Acknowledgment of two perspectives, so moderate differentiation. However, the individual hasn't linked the perspectives or contrasted them.
"Some view abortion as a civil liberties issue-that of the woman's right to choose; others view abortion as no more justifiable than murder. Which perspective one takes depends on when one views the organism developing within the mother as a human being."	Score of 5. Reflect moderate to high differentiation and moderate integration. There is explicit comparison of contrasting alternative perspectives on the issue.
"Some view abortion as a civil liberties issue; others see abortion as tantamount to murder. One's view of abortion depends on a complicated mixture of legal, moral, philosophical, and perhaps scientific judgments. Is there a constitutional right to abortion? What criteria should be used to determine when human life begins? Who possesses the authority to resolve these issues?"	High Score of 7. Scores of 7 reflect high differentiation and high integration. Explicitly acknowledges that there are different viewpoints and develops complex rules to compare and contrast them.

For example, the sample text given below shows moderate Integrative Complexity. On a range of 3 - 4. It depicts moderate differentiation and moderate integration. (Excerpt taken

[1] *Note.* Adapted from "Content Analysis of Archival Data," by Fiona Lee and Christopher Peterson, 1997, *Journal of Consulting and Clinical Psychology,* 65, p.959-969

from an article written by Shamshad Akhtar, Under-Secretary –General of the United Nations and Executive Secretary of the Economic and Social Commission for Asia and the Pacific.)

> *"Countries in the Asia-Pacific region continue to drive the global economy. The region has demonstrated great resilience during the economic crisis. Yet, regional growth is now in a challenging phase. Widespread poverty, rising inequality, social inequity and environmental degradation are hurdles to be cleared. Regional growth dynamics are being influenced by anemic recovery in the developed world, given the weak implementation of policies. Yet, overcoming domestic structural impediments is also vital."*

Manual coders who undergo the Integrative Complexity training workshops (for a duration of two weeks) typically attain an ideal 85% agreement with experts. (Suedfeld et al. 1992).

1.3 Issues faced in designing an Automated Scorer for measuring Integrative Complexity

Upon designing the automated scorer, several factors have to be kept in mind. One is that the manual scoring of Integrative/Conceptual Complexity is actually subject to the scorer's preferences and understanding. What some manual scorers identify as Differentiation or Integration, some others might see as mere qualifications. Therefore it makes more sense to automate the scorer, as the computer program would assign a single score and make no further judgments or further assertions. In some of the scoring material where manual scoring has been done, more than one score has been assigned to a particular text, which means that a particular text sample can have more than one score. For example, the following text had two scores assigned to it. The text assigned to the scorer is given in the next page:

"I am an impulsive person and have found myself in numerous interesting situations due to this aspect of my nature. My older brother, who is very close to me, finds this quality in

me rather frustrating at times (he is a cautious individual) especially when he wants to say "I told you so" but doesn't in order to avoid a quarrel. He does enjoy watching to see the outcomes of my impulsive situations and I know he does evaluate some of his own choices in life based on my experiences. We live in each other's lives vicariously, I think he is in mine more than I in his."

The text was assigned two individual scores of 4 and 5. Either of them would suffice according to the scoring manual (Baker-Brown et.al). Provided below is an explanation of the scoring process given in the scoring manual.

"There is clear Differentiation between the author's description of himself and that of his brother. The last two sentences hint at Integration with "he does evaluate some of his own choices in life based on my experiences" and "We live in each other's lives vicariously". Some have argued that the last sentence is a more explicit statement of Integration. Such disagreement is a reminder that complexity scoring is, to a large degree, dependent upon the scorer's interpretation of the material. Such interpretation is necessary, but a good scorer is always careful not to read too much into the text. In this case, the disagreement is based on a legitimate alternate estimation of the author's intent."

Another example that shows the ambiguity in the scoring process is:

"Advances made in the chemistry of antiseptics and the techniques of surgery are not wholly responsible for the new standards of lifesaving in war. An alert and courageous system of fully equipped yet highly mobile surgical units following close behind the assault troops has resulted in an immense saving of time between the battlefield and the operation table. In surgery time-saving is akin to lifesaving."

The extract is the explanation in the manual as to why the score should be a 4 instead of 3. The scoring manual (Suedfeld et al., 1992) outlines the confusion that was involved in the scoring process as follows:

"To show how even expert scorers can be imperfect, this paragraph has been listed for years as a "3" because one Differentiation was seen as comprising the bulk of the paragraph: "Advances made in the chemistry of antiseptics" as contrasted from "An alert and courageous system . . ." and related time savings, both considered factors contributing to the saving of lives during the war. At a workshop in May of 2000, however, it was decided that the final sentence hints at an over-arching schema, enough to allow a score of 4. This score garners further support from the phrase "not wholly responsible" in the first sentence."

There are certain issues that have been raised regarding the validity of scoring of IC through content analysis from archival materials. One of the concerns was pertaining to the validity of IC scores obtained from texts that have been written by individuals standing in place of the identified source (e.g. speechwriters). This concern was particularly highlighted with regards to a source who wanted to deliberately manipulate the score to project a different image (stems from the known fact that most modern leaders use materials/speeches prepared by aides), or if the source wanted to be deceptive, or if the source has expertise in a particular area. It is undoubtedly imperative to IC research to understand whether scores vary. However, studies have discovered that there is little, if any difference between the complexity scores of private and public materials originated by an aide, where the latter might be originated by an aide (Suedfeld & Tetlock, 1977). Another concern that was whether the impression management hypothesis bore any weight to IC scoring. The hypothesis states that any leader in his attempt to appear as decisive and determined may try to endorse low levels of IC in his communications, whereas a leader who wishes to project an image of a leader who is flexible, moderate and considerate may

try to endorse communications that reflect high levels of IC. The hypothesis was discredited after studies found that consistency was maintained across public and private utterances.

1.4 Integrative Complexity as a Predictor for aggression

Thomas Jordan in his paper (Jordan, 1998) titled "*Structures of geopolitical reasoning*: *Outline of a constructive-developmental approach*" summarized the need for a measure for differentiating between varying levels of reasoning in a geo-political context. It is this relationship between cognitive processes and complexity of thought that is being attempted to be captured by the Integrative Complexity framework. Studies have found that in general, when integrative complexity is low, aggression and hostility often prevail (Bruch, McCann, & Harvey, 1991; Winter, 1993).

One study (Winter, 1993) showed that if police officers do not exhibit moderate to high levels of integrative complexity, they are more inclined to act violently in stressful contexts. Intuitively this makes sense as when integrative complexity is low, other options to solve problems are less likely to be considered. A good deal of research suggests a high level of power motive imagery, a low level of concern for affiliation, a low level of responsibility and a low degree of Integrative complexity with other things being equal, tend to predispose decision makers and key political actors towards war rather than peace (Winter, 2007). Tetlock (1983) showed that simplistic black-and-white thinking that is characterized as having low Integrative Complexity is associated with more aggressive crisis outcomes. Alternately resolving crises peacefully requires differentiation and integration-in short, complex thinking.

One study (Matsumoto, D., & Hwang, H. C, 2013) examined the role of language used

by world leaders and leaders of ideologically motivated groups as indicators of whether the group escalated into acts of aggression (AoA) or into Acts of Resistance (AoR) and found that AoA speeches decreased in cognitive complexity across time near the focal event, whereas AoR speeches increased in cognitive complexity, and this is consistent with previous research on cognitive complexity and integrative complexity (Abe, 2012; Suedfeld & Bluck, 1988; Suedfeld, Tetlock & Ramirez, 1977). An intuitive explanation suggested by the authors was that cognitive complexity decreased in speeches, because the AoA itself became the solution and there was less need to comprehend and consider other competing solutions, whereas in the case of AoRs levels of cognitive complexity increased or were maintained as aggression was avoided and peace had to be maintained. (For more reviews, see Conway, Suedfeld & Tetlock, 2001; Suedfeld, 2010).

Determining the markers of aggression in text has significant implications: (Suedfeld et al., 1977). Theoretically, the identification of such markers would improve our understanding of the mental state of the author. The determination of the relationship between complexity and aggression would also provide a way to assess the potential for aggression by others. This open up the potential for the development of intervention systems that could monitor the dynamic nature of aggression and further organize an intervention to avoid the negative consequences of aggression and acts of violence.

1.5 Integrative Complexity in Politics

Many studies have examined the relationship between political preferences and integrative complexity (Tetlock, 1983, 1984; Tetlock, Hannum, & Micheletti, 1984). They have shown that liberal or moderate politicians exhibit complex thinking, i.e. thinking

characterized as having higher IC rather than simple thinking styles which were exhibited by their conservative counterparts.

Studies have shown that some leaders who have been shown to exhibit high levels of Conceptual complexity in their communications, experience drop in their IC levels when the situation becomes adverse, while others maintain their high IC levels even during moments of stress. The study of the political career of Andrei A Gromyko (Wallace & Suedfeld, 1988) revealed that he managed to retain high IC even when a crisis approached (Which was unlike his peers), which led to the speculation that his immunity to disruptive stress might have been the reason for his long career through years of changing leaders and circumstances. Interestingly, for leaders who exhibit high IC, it sometimes creates difficulties for them during governance. For e.g. the current US president Barack Obama is credited with having a high IC in his communications, in contrast to the former US president George W. Bush, who was considered to have low IC. However as Jonathan Haidt, a professor of social psychology at the University of Virginia remarked in a recent newspaper article on the Washington Post, "What distinguishes Obama particularly is the depth and carefulness of his thinking, which renders him somewhat unfit for politics," (Milbank, 2011). That the US President does not have a single organizing theme dominating over all his decisions may render him to be more careful and cautious with decision making, which may give off the impression of being unprincipled and confused.

The assumption that leaders who exhibit higher IC in their communications are more prone to success or to better decisions is erroneous at best. High IC thinking may In fact waste precious cognitive resources on unimportant, trivial easily solved problems or on processing irrelevant information (see, e.g., Tetlock & Boettger, 1989). An example is

when Neville Chamberlain's IC during the fateful negotiations at Munich was twice as high as Hitler's (Suedfeld, Leighton, & Conway, 2006). Leaders probably operate at higher levels of IC because of greater accountability. (e.g., Suedfeld & Leighton, 2002; Wallace, Suedfeld & Thachuk., 1993).

A major decrease in the level of Integrative Complexity can also be noted in the wake of strategic surprise attacks emanating from the leaders of the eventual attacker side (beginning about 6 months prior to the attack). This can be seen as a very important finding that could be used in planning an intervention system for reducing the possibilities of strategic surprise attacks, by monitoring the IC levels of potential attackers. The importance of IC in decision making and information processing has been established beyond any reasonable doubt.

1.6 Integrative Complexity and Performance

Studies have shown that IC is associated with effective performance in a variety of domains. Intuitively this could be attributed to the fact that complex thinkers attend to more information, in particular contradictory notions (Winter, 1996).Studies have found that activities which require some difficulty, in which team members often need to collaborate to solve the problems, demand thinking that can be characteristic of high IC (Gruenfeld & Hollingshead, 1993). Another study found that scientists involved in research who had high IC were cited more frequently and were perceived as eminent in their research fields, but they were perceived as hostile and exploitative. However scientists in teaching capacities who showed the same form of thinking were perceived as warm (Feist, 1994).

A research project investigating whether a diverse group of individuals are better at solving complex cognitive tasks than a homogeneous group of individuals found that

moderate levels of group disparity showed the highest level of complexity for groups (Curşeu, P. L., Schruijer, S., & Boroş, S.,2007). Conceptual networks (cognitive mapping) was used to illustrate the cognitive complexity of the group. Group Cognitive Complexity (GCC) was positively influenced by average individual complexity. The groups with members that have highly complex maps and experienced effective teamwork processes while working as a group had the highest GCC.

1.7 Concluding Remarks

The use of Integrative Complexity in various applications of political psychology and social psychology have been mentioned in the previous sections. It is of outmost importance that the construct be studied and automated, not only to allow researchers in psychology to use it more freely, but also to understand it in a way, that would allow research in artificial intelligence to obtain a deeper understanding of how our minds consider different information sources during decision-making.

CHAPTER 2

AUTOMATED SCORING OF LEVELS OF INTEGRATIVE COMPLEXITY USING MACHINE LEARNING[2]

[2] Aardra Kannan Ambili, Khaled Rasheed. Accepted by the 13th International Conference on Machine Learning and Applications. Reprinted here with permission of the publisher.

2.1 Abstract

Integrative complexity is a construct developed in political psychology and clinical psychology to measure an individual's ability to consider different perspectives on a particular issue and reach a justifiable conclusion after consideration of said perspectives. Integrative complexity (IC) is usually determined from text through manual scoring, which is time-consuming, laborious and expensive. Consequently, there is a demand for automating the scoring, which could significantly reduce the time, expense and cognitive resources spent in the process. Any algorithm that could achieve the above with a reasonable accuracy could assist in the development of intervention systems for reducing the potential for aggression, systems for recruitment processes and even training personnel for improving group disparity in the corporate world. In this study we used machine learning to predict IC levels from text. We achieved over 78% accuracy in a three way classification

2.2 Introduction

The need to measure an individual's inclination to examine different views on a particular issue and form a rational conclusion was captured in the construct termed as Integrative complexity (IC). It has been touted as the most used and widely validated measurement of complex thinking. IC is the function of two constructs, Differentiation and Integration. Differentiation relates to the capacity of individuals to adopt and to apply a variety of perspectives on an issue. On the other hand, integration refers to the capacity of individuals to recognize interweaving connections and contrasts across these perspectives. The levels of IC scored from the author's text or speech has been found to be indicative of reasoning skill, intelligence and its correlates. This measure has been said to be able to

capture the underlying mechanisms of the complexity of thought on a broad level regardless of variables that may influence the cognitive strategies used in formulating the text.

Low levels of IC are characterized by rigid and simplistic views on ideas. Such thinking would consider all other views as illegitimate, flawed and ridiculous (Suedfeld et al., 1992).Thinking classified as characteristic of high IC often acknowledges multiple perspectives on an issue and is also capable of recognizing how these views might contribute to a logical conclusion.

Complex thinkers whom have been attributed to have high IC have also been associated with effective performance in a variety of domains, This could be accredited to the fact that complex thinkers tend to attend to more information, paying special attention to data that may hold contradictory perspectives (Tetlock, Hannum, & Micheletti, 1984). In 1994, Feist showed that scientists who were complex thinkers were often cited more frequently (Feist, 1994). Studies examining the association between personality and IC, found that complex individuals also report elevated scores on openness and low scores on compliance and conscientiousness. The same individuals were gauged by a semi projective test called the Picture Story Exercise (PSE) to show more motivation to seek power(Coren & Suedfeld, 1995; Tetlock, Peterson & Berry, 1993).

The relationship with group IC and task performance have been studied (Gruenfeld & Hollingshead, 1993). While high IC levels are preferred in situations where individuals need to collaborate to solve individual problems, low IC levels are preferred in situations that require rapid decision taking and negotiation skills (Suedfeld, 1992). High IC top management teams (TMTs) were found to have improved corporate social performance

(CSP) (Wong, Ormiston & Tetlock, 2011). These results have a significant implication in the corporate world, since CSP is related to firm financial performance (Margolis& Walsh, 2003). Therefore teams in centralized organizations can improve their financial performance either by decentralizing or by adjusting their IC (Suedfeld, 1992). Such insights if obtained without laborious manual scoring could even augment recruitment processes in government, schools and universities. Additionally, to foster creativity and novel thinking in teams, managers should ensure that many of the individuals in a group espouse the same attitude towards some issues while maintaining two or three individuals to express divergent perspectives, since high levels of complexity were exhibited by groups with moderate levels of group disparity (Curşeu, Schruijer & Boroş, 2007).

Studies have also linked relationships between IC and political preferences [24] (Tetlock, 1983; Tetlock, Bernzweig & Gallant, 1985). Liberal or moderate politicians often exhibit complex rather than simple thinking styles. There have also been several studies confirming the proposition that individuals who adopt a majority position might show an increase in IC (Tetlock, 1984; Tetlock, Hannum, & Micheletti, 1984)

IC scores can be used to predict future events of aggression, rather than being merely used as explanations for past events. This has two major implications for national decision makers. The first one coaxes officials to maintain high levels of differentiation and integration during periods of crisis and the other is to be aware that officials of other countries might in fact be processing information at a lower level of integrative complexity than is normal-which could plausibly affect the decisions surrounding both countries (Suedfeld, Tetlock & Ramirez, 1977). The determination of the IC scores to predict aggression would provide a way to assess the future potential for aggression by others,

making possible the development of intervention systems that could monitor the dynamic nature of aggression and plan an intervention to avoid the negative consequences of aggression and acts of violence. In 2013, Liht and Savage demonstrated than an intervention program aimed to increase integrative complexity among youngsters in the UK, managed to shift the conflict resolution style towards collaboration and compromise in group discussions and written responses (Liht & Savage, 2013).

Thus, we propose a machine learning approach for automating the scoring of Integrative Complexity from text. Our approach uses Logistic regression, Support Vector Machines and Multi-class classifiers. Our paper is a first in addressing the gap in current literature for accurate reliable IC scoring from paragraphs of text through machine learning algorithms.

2.3 Related work

Integrative Complexity and Cognitive Complexity

The term cognitive complexity (CC) was first proposed by James Bieri in 1955. His work described a system of constructs as cognitively complex if it differentiates highly among persons and a system as cognitively simple in structure if it provides poor differentiation among persons (Bieri, 1955). The technique deployed by Bieri for measuring the degree of cognitive complexity among one's perception of others was the Role Construct Repertory Test (RCRT). One of the components of IC, differentiation, is essentially equivalent to the variable of cognitive complexity. The concept of CC can also be applied to groups as groups are viewed as socio-cognitive systems.

IC is defined as the function of two components, namely Differentiation and Integration. Linguistic Inquiry and Word Count (LIWC) is the first text analysis program that measures

CC (Pennebaker, Francis & Booth, 2001). It counts words in psychologically meaningful categories (Tausczik & Pennebaker, 2010). It measures CC by counting words which belong in two categories- exclusion words and conjunctions. Exclusion words (e.g., but, without, exclude) are helpful in making distinctions among different sentences. Conjunctions (e.g., and, also, although) join multiple sentences and contribute to CC (Graesser, McNamara, Louwerse, & Cai, 2004). In 2011, Abe used the LIWC measure of CC to examined changes in Alan Greenspan's language use across the economic cycle by analyzing his testimonies and speeches (Abe, 2011).

Integrative complexity is usually measured through content analysis of archival data (Lee & Peterson, 1997).Weber defined Content analysis as a "research method that uses a set of procedures to make valid inferences from text" (Weber, 1990). Data found in spoken and written reports are convenient sources of information about the concerned individual's thinking processes, although they should not be equated with the individual's cognition. IC has been measured from speeches, diplomatic documents, interview transcripts policy statements and personal letters.

The procedure is methodologically rigorous. Initially identifying material is removed from the material before scoring. Scorers are qualified through a rigorous training course, where their inter-scorer reliability is tested for different studies repeatedly over time (Baker-Brown et al., 1992). The scores are then evaluated through normal inferential statistics. Archival data is first divided into units, where each unit is defined as a section of text pertaining to a single idea, belonging to a single author (while measuring individual IC). Typically, a unit consists of a single paragraph. IC is scored on a 1-7 scale, where each scale point is characterized by specific levels of differentiation and integration (1=low

differentiation and low integration, 7= high differentiation and high integration). Trained coders usually obtain a correlation of .80 with expert coders. Traditionally, manual coders are required to take workshop training sessions lasting upto two weeks.

Coders often have to make difficult judgments on whether differentiation or integration exist in a snippet of text. Since it is extremely hard for coders to become objective while making these judgments, it is often the case that a single text could have different scores attributed by different scorers. For example, it is sometimes difficult to say whether a justification qualifies as an alternate perspective. For example, the following text (taken from the coding manual) (Suedfeld & Eichhorn, 2013). had two scores assigned to it. It was assigned two individual scores of 4 and 5. Either of them would suffice according to the scoring manual. Such passages usually merit transition scores of 2, 4, and 6; implying implicit differentiation or implicit integration.

"I am an impulsive person and have found myself in numerous interesting situations due to this aspect of my nature. My older brother, who is very close to me, finds this quality in me rather frustrating at times (he is a cautious individual) especially when he wants to say "I told you so" but doesn't in order to avoid a quarrel. He does enjoy watching to see the outcomes of my impulsive situations and I know he does evaluate some of his own choices in life based on my experiences. We live in each other's lives vicariously, I think he is in mine more than I in his."

2.4 Methodology

Data Selection

The data for the project consisted of 165 text samples along with the scores provided by manual coders. The data was taken from Suedfeld's Complexity Materials Download Page [19]. Initially each paragraph was scored a value between 1 and 7. After conducting experiments on the multi- classification problem with 7 classes and 165 instances, we decided that we could potentially improve performance by binning IC scores ranging from

1-7 to three bins (as shown in Table II). Intuitively, this also made sense, as 2, 4, and 6 were transition scores. Three different sets of features were selected for experimentation. Results are shown for the most successful set of features

Data Preprocessing

Initially, the data was cleaned and converted into an ARFF (Attribute Relation File Format) file format. Feature extraction methods played a huge role in this text-classification problem. The string in the text attribute of each instance is converted to a set of attributes representing word occurrences, using the filter in Weka (Hall et al., 2009) called the String to Word Vector filter. The number of attributes obtained after step was 1253. This was reduced to 28 using the Attribute Selection filter. The Correlation-based Feature Subset Selection evaluator was used.

TABLE 2.1: Binning of IC scores

Integrative Complexity Scores	***Class name***
1 - 2	Low IC
3 - 5	Moderate IC
6 - 7	High IC

Learning Methods

We have used for the purpose of this study classification algorithms such as the multilayer perceptron, support vector machines, bagging and the logistic regression models. For this purpose, we used the University of Waikato's WEKA (Hall et al., 2009) software. We used a validation approach in which we used the N-fold cross validation, where we set N to 10. For the validation approach, the data is partitioned into N disjoint subsets, then trained on

the training set (a set of N-1 disjoint tests) after the hold out set is removed. Then, the model is tested on the holdout set for validation. This process is repeated N times and the average accuracy is reported. Cross validation is generally used when the size of the data set is small.

Support Vector Machines are the first classifiers used for experimentation. These were trained using John Platt's sequential minimal optimization algorithm (Platt, 1999) implemented in Weka. Since the concerned problem is a multi-class classification problem, pairwise classification is followed. Polykernel, Radial Basis Function (RBF) kernel and the Puk kernels (The Pearson VII function-based universal kernel) were used.

The second classification algorithm that was used was a multinomial logistic regression model with a ridge estimator (Le Cessie & Van Houwelingen, 1992). Commonly mistaken as a regression algorithm, logistic regression is actually a classification algorithm. The multinomial logistic regression algorithm is a generalization of the logistic regression algorithm for multi-class classification problems. The probabilities describing the outcomes of an instance are modeled as a function of its features, using a logistic function. The experiments are carried out with optimization procedures for the search for parameters. They are the Broyden–Fletcher–Goldfarb–Shannon (BFGS) algorithm and the Conjugate Gradient Descent optimization algorithm. The Conjugate Gradient Descent optimization algorithm is used for faster updates, when there are many parameters.

Experiments were also carried out using the multilayer perceptron (MLP) as the classification algorithm. An MLP consisting of multiple layers of nodes in a directed graph, uses a supervised learning techniques called backpropagation for training the classifier. The MLP used in this work contained only nodes that had sigmoid functions as activation functions. The learning rate was set at 0.3 and momentum was set at 0.2.

A Voting classification algorithm such as Bagging was also used in this task. Bagging has been shown to be very successful in improving the accuracy of certain classifiers for artificial and real-world datasets (Breiman, 1996; Freund & Schapire, 1996). Base classifiers that were used were the Support Vector machine and the Multinomial Logistic regression model with a ridge estimator.

The last classifier we experimented with was the Multi-class classifier- suitable for the multi-class classification problem. The Meta classifier used binary classifiers to solve the 3 – class classification problem. The binary classifiers used for experimentation were the logistic regression and the multi-layer perceptron. Popular multi-classification methods like 1-against-1 and 1-against-all were used, with and without pairwise coupling.

2.5 Evaluation and Results

Evaluation approach

The performance of our classifiers is tested through stratified 10-fold cross-validation. Considering the limited amount of data, in a multi-class classification problem, the standard way of predicting the error rate of a learning technique is to use stratified 10-fold cross-validation. The complete data set is divided randomly into 10 portions. Stratification ensures that each class is represented in approximately the same proportions as in the original data set in each of the 10 portions. Thus, the training and testing is performed a total of 10 times on 10 different sets. Finally each individual error estimated are averaged to yield an overall error estimate. We chose 10, because of the empirical evidence behind it proving that it is the right number of folds to get the best estimate of error (Manning, Raghavan & Schütze, 2008).

We have used classification accuracy as one of the performance measures for this problem. But our focus is more on precision, recall and F-1 measures, as they tend to be better measures when evaluating small classes (Manning et al., 2008). Effectiveness is a generic term for these three measures. On the same note, we should compute macro-averaged results, to get a better sense of effectiveness over small classes.

Results

Table 2.2 shows the performance accuracies of the classifiers on the dataset. We can see that the highest classification accuracy was delivered by the multinomial logistic regression model and the multi-class classifier with 1 vs 1 method. Since we are dealing with a relatively small dataset, we have also reported the Precision, Recall and F-measures along with it. Table 2.3 and Table 2.4 report these scores. The same classifiers also have the highest reported precision, recall and F-measures. The SVM‘s F-measure shows that the SVM does 1.3% worse than the Multi-class classifier and the Multinomial Logistic regression, which amounts to a relative decrease of just 1%.

Table 2.5 and 2.6 show the confusion matrices for the multinomial logistic regression model and the multi-class classifier respectively. From the confusion matrix, it can been seen that the classifiers have little trouble in discerning between high IC and low IC. Out of the 89 instances having mid IC values, 71 were classified correctly, which is a considerable portion of instances. Note that most of the inaccuracies happened when the classifier was not able to discern very well between instances belonging to mid IC and low IC, or between mid IC and high IC, which is understandable considering that this might where the ambiguity of IC scoring played a role. Traditionally, we had transition IC scores like 2, 4, and 6 to display the ambiguity that an instance could have two IC scores. For a

human rater to qualify as reliable, the individual would have to have a minimum inter rater reliability of 0.8. Considering this statistic, an inter rater reliability of 0.78 from an algorithm is acceptable.

Table 2.2.: Classification accuracies

Algorithm	Specifications	Accuracy
SVM	1.Kernel: Polykernel 2. Logistic models are fit to the outputs to allow probability estimates	76.83%
SVM	1.RBF kernel 2.Logistic models are fit to the outputs to allow probability estimates	75%
SVM	Puk kernel	72.56%
Multilayer Perceptron		72.56%
Multinomial Logistic regression with a ridge estimator	Uses Conjugate Gradient Descent	78.0488%
Multi-class Classifier	Logistic regression as the base classifier Method:1-against-1	78.0488
Bagging		76.8%

Table 2.3. Effectiveness measures for Multilayer Perceptron, Multinomial Logistic regression model and the Multi-class classifier

Class	Multilayer Perceptron			Multinomial Logistic regression			Multiclass classifier		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
Low	0.691	0.717	0.704	0.774	0.774	0.774	0.759	0.774	0.766
Mid	0.747	0.764	0.755	0.807	0.798	0.802	0.807	0.798	0.802
High	0.722	0.591	0.650	0.696	0.727	0.711	0.727	0.727	0.727
Weighted avg.	0.726	0.726	0.725	**0.781**	**0.780**	**0.781**	**0.781**	**0.780**	**0.781**

Table 2.4. Effectiveness measures for SVMs with different kernels

Class	**SVM-Polykernel**			**SVM-RBF kernel**			**SMO- Puk kernel**		
	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
Low	.759	.774	.766	.758	.472	.581	.741	.755	.748
Mid	.789	.798	.793	.708	.843	.769	.787	.787	.787
High	.700	.636	.667	.760	.864	.809	.619	.591	.605
Weighted avg.	**.767**	**.768**	**.768**	.731	.726	.714	.749	.750	.750

Table 2.5. Confusion Matrix for the Multinomial Logistic regression model

Low	***Mid***	***High***	***Classified as***
41	12	0	***Low***
11	71	7	***Mid***
1	5	16	***High***

Table 2.6. Confusion matrix for the Multi-class Classifier

Low	***Mid***	***High***	***← classified as***
41	12	0	***Low***
12	71	6	***Mid***
1	5	16	***High***

2.6 Future Work

An extension to this work would focus on improving the algorithm's accuracy. An optimistic extension to the algorithm could produce results that exceed the average manual rater's accuracy, to approach expert coder proficiency. Such work could try to include measures that involve natural language processing techniques. We could also include

features that measure a paragraph's coherence, the sequential nature of an argument or perspectives, all of which could help in deriving a better model for prediction of IC levels from text.

Future work could build upon this study to help build an intervention system to predict potential acts of aggression and subsequently plan interventions and resolve conflicts. Further studies could help design a system to improve creativity among individuals, by using Integrative complexity as an indicator [28]. Prediction of IC through machine learning algorithms could also lead to development of recruitment systems that could aim at improving a team's integrative complexity. The biggest advantage to the development of an efficient algorithm that could accurately predict IC scores is that research scientists in political psychology, social psychology and management studies could eliminate the laborious manual scoring process. This could also aid in the development of a system that could measure an individual's IC, cognitive complexity and other correlates such as inclination to seek violence, motivation to seek power, motivation to seek initiative and several others.

2.7 Conclusion

The motivation behind this work is to address the need for an efficient algorithm that could accurately predict Integrative Complexity from text. This could eliminate the time, expense and laboriousness required in the manual scoring process. The two week long workshop required for manual raters can be eliminated with the efficient adoption of algorithms that closely approach manual rater efficiency. This was the first study that attempted to apply machine learning algorithms for determining IC from text. The results indicate that the logistic regression model and the multi-class classifier can be used in

predicting IC from text. In light of the fact that a manual rater usually achieves an accuracy of 80%, our achievement of an accuracy of 78% in this three-class classification problem is significant, mostly because of the dearth of more efficient algorithms for solving the problem in current literature. This work should not be considered as the final conclusive step in solving the problem, but as a necessary starting point in solving the problem.

CHAPTER 3

AUTOMATED SCORING OF INTEGRATIVE COMPLEXITY THROUGH NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING[3]

[3] Aardra Kannan Ambili, Khaled Rasheed. Submitted under review to the 28th International FLAIRS Conference, 11/14/2014.

3.1 Introduction

The earlier work focused on solving the problem through a pure machine learning approach. In most real-world problems, the task of text categorization can be easily and efficiently solved; given that (i) the features that distinguish an instance as a member of a class is retrievable through available feature extraction methods (ii) and there's enough labeled data for an equitable labeled distribution. The problem I described earlier on in this body of work didn't really adhere to either of these constraints. Therefore, in that and more, the problem of automating the process of scoring integrative complexity is non-trivial and fascinating at the same time. It asks the age-old question of how to make a computer understand the intricacies of language enough to address two competing perspectives in text, or to make a computer recognize a text sample that connects between interweaving connections and reach a rational conclusion from them?

The automation of scoring of Integrative Complexity is a particularly tough problem, as the problem of detection of differentiation and integration is virtually unsolved in current literature. The problem of detecting differentiation in text boils down to the detection of differentiated statements on a particular issue. The concept of differentiation makes sense to a human reader because of the ability of humans to comprehend the thesis of the issue at hand and the 'extent of differentiation' expressed by the subsequent/preceding statements in the text. Intuitively we can see that a deeper semantic understanding of the text at the discourse/sentence level could produce much better results for the computer at what differentiation means. Without the involvement of a semantic understanding from the computer, the contribution that will be played by automation will still be at the periphery. On the other hand, the problem of detecting integration could be

solved in the following way: an integrating statement could be seen as comprising of semantic information from either its competing perspectives or from its thesis. This chapter focuses on engaging the knowledge engineering approach with the machine learning approach to solve this problem more efficiently.

The inculcation of semantic information from the competing perspectives to act as a predictor for differentiation and integration is implemented through the introduction of a semantic predictor in this paper. This newly designed feature has been aptly named Semantic Paragraph Coherence.

3.2 Understanding Semantic Paragraph Coherence

The level of Integrative Complexity in a text is contingent upon the connections or links present in the text. It is the presence of the connecting or competing perspectives that predict differentiation. However, without a minimum level of differentiation in a text sample, it is not possible for integration to be present in the sample.

For e.g. consider the following sample:

"It is cloudy. Therefore it might rain today."

The sample would be scored a minimum IC score of 1. Since, there is no explicit or implicit differentiation. And since there are no differentiating statements, there cannot be any integration present.

Therefore, it seems reasonably intuitive to design a feature that would attempt at predicting differentiation, and thereby predict integration. The combination of this newly designed feature with a bag of words approach could optimistically lead to better performance. For example, consider the below example. (Taken from Peter Suedfeld's integrative complexity training workshop: Suedfeld et al. 1992)

'One form of self-expression is influenced by our interpersonal relationships and experiences. My relations with my parents and friends have made me value honesty and intimacy. Another child's upbringing may have made independence a central concern. Unfortunately some children's social environment fosters mistrust and fear of rejection. By adulthood if not earlier we have all created a style of expressing ourselves each subtly different because of our varying backgrounds which alter the paths we follow through life.'

The thesis for the sample text considered is the statement that: '*One form of self-expression is influenced by our interpersonal relationships and experiences*.' This is followed subsequently by a series of differing perspectives on the issue. All these three perspectives are integrated with the final declarative statement that: '*By adulthood if not earlier we have all created a style of expressing ourselves each subtly different because of our varying backgrounds which alter the paths we follow through life.'* And therefore, the thesis receives further support from the final integrating statement. The considerable amount of differentiation and explicit integration earn the instance high integrative complexity.

For the text sample to be qualified as having high integrative complexity, numerous differentiations have to be made, subsequently followed by explicit and well-articulated integration that draws from the differentiated perspectives. It could be inferred that differentiating statements often relate to each other with a non-zero amount of semantic similarity. Often it is the case that, most differentiating and integrating statements would intuitively be semantically similar in content to an extent.

In this particular example, the differentiating statements do have some semantic similarity. The semantic content of the reference made in the thesis sentence of the text to "*self-expression is influenced by our interpersonal relationships and experiences*", is referred to semantically in content in the subsequent differentiating statement as "*my parents and*

friends" and "*children's social environment".* In the final integrating conclusion, we can determine a semantic similarity to "*a style of expressing ourselves*". It is this property that is exhibited by the complex instances that could be exploited in the prediction of levels of integrative complexity.

Paragraph coherence as a feature.

The problem of measuring semantic similarity between sentences could be translated into measuring the semantic similarity of words that carry the most information in these sentences. A primary assumption that we make for the development of this feature is that most often the semantic content in sentences come from the nouns, verbs and adjectives and to a lesser degree from adverbs, prepositions and the rest. Semantic similarity between sentences could be limited to calculating the semantic similarity between words that are common with the two sentences (Meadow, Boyce & Craft, 2000). This worked reasonably well in texts of longer lengths. The probability of co-occurrence of words is higher in longer texts. However, for shorter texts, a method which focused on the semantic meaning of the words rather than the word itself was required.

Semantic similarity between words.

The method for calculating semantic similarity between words is based on Li, Bandar & McLean's work in 2003 (Li, Bandar & McLean, 2003). The similarity of two words is calculated using a hierarchical semantic knowledge base e.g. WordNet (Fellbaum, 1998 ed; Fellbaum, 2010; Miller, 1995). The work presented in this paper calculated semantic similarity as a function of path length in WordNet and depth in WordNet. Path length in WordNet is the minimum number of words lying between the considered words in the

hierarchical knowledge base, and the depth word is the depth of the subsumer[4] in the hierarchy. The derived function from Li et al.'s work in 2003 is a function of path length (l) and depth (h).The parameters α and β are that are used to scale the contributions of path length and depth respectively. Let the semantic similarity between two words w_1 and w_2 be noted by $S(w_1, w_2)$. Then according to (Li et al. 2003):

$$S(w_1, w_2) = f(l).f(h) \quad --(1)$$

In equation (1) $\alpha \geq 0$ and $\beta > 0$. The paper (Li et al. 2003) also proposed optimal values of $\alpha = 0.2$, and $\beta = 0.6$ as the recommended parameter values for close correlation with human understanding.

3.3. Implementation details for Semantic Paragraph Coherence

This section describes the method for calculating semantic similarity between words. This rendition is just for reference. For further explanation, please refer to the original paper (Li et al. 2003). The semantic similarity method was coded in SWI Prolog, since WordNet version 3.0 was also available in Prolog. (Fellbaum, 1998 ed; Fellbaum, 2010; Miller, 1995).

Contribution of path length

The path length between two words in a hierarchical knowledge base can vary between 0 to very large numbers. Hence the function should be designed so that it will have values ranging from 0 to 1 (Li et al. 2003).

This function will depend on three cases: In the first case, $f(l) = 1$; if w_1 and w_2 belong to the same concept. In WordNet (Fellbaum, 1998 ed; Fellbaum, 2010; Miller, 1995), two words belong to the same concept if

[4] A subsumer is a concept in a lexical taxonomy. It is a word that is less general than the concept which subsumes it. For e.g. Dog is subsumed by Animal. Therefore, Dog is a subsumer.

i. If w_1 and w_2 are the same words

ii. if w_1 is an instance of w_2;

iii. if w_1 and w_2 are verbs and if $w_1$entails w2 ;

iv. if w_1 and w_2 are adjectives; and they mean the same thing

v. if $w_1$is a memeber meronym of the first synset;

vi. if w_2 is a substance meronym of w_1;

vii. if w_2 is part meronym of w_1;

viii. if $w_1$has been classified as a member of the second word;

ix. if there exists a reflexive lexical morphosemantic between the two words representing derivational morphology

x. if there exists a first-order hypernym relation between w_1 and w_2.

In the case that the two words do not belong to the same concept, but have the same word linking them, their semantic similarity is calculated as:

$$f(l) = \ e^{-\alpha l} \qquad --(2)$$

Contribution of depth

A subsumer of two words is the first common hypernym between them. Depth, h is the depth of the subsumer in the hierarchical semantic net. For example, for the words 'boy' and 'girl', the path is 'boy-male-person-female-girl', then the synset for 'person' (first common hypernym) is the subsumer for 'boy' and 'girl'. The depth is calculated by counting the levels from the subsumer level to the top of the lexical hierarchy in wordnet. In case of

polysemous words, the subsumer of the shortest path is considered in deriving the depth of the subsumer.

$$f(h) = \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} \qquad - (3)$$

Calculation of semantic similarity between words

The semantic similarity between two words is a function of path length (l) and depth (h), According to Li et al. (2003). Then the semantic similarity between two words w_1 and w_2 be noted by $S(w_1, w_2)$ (i.e a product of equations (1) and (2)),

$$S(w_1, w_2) = f(l).f(h) = e^{-\alpha l}.\frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} \quad - (4)$$

Calculation of Paragraph Coherence

The designed intelligent feature measures the semantic cohesiveness of a paragraph. The hypothesis that Paragraph coherence can be used as a feature to predict the level of Integrative Complexity is tested in this paper. As has been mentioned before, the scoring of Integrative complexity involves the measurement of Differentiation and Integration in text. A text sample that has been characterized as having low Integrative Complexity, could be seen as having low differentiation, or in other words as comprising of unconnected discourse. It could also be seen as a paragraph that focuses on a single thesis with descriptive statements. Admittedly, paragraph coherence would fail as a feature in the scoring of Integrative Complexity for examples where the thesis was being referred to repeatedly in the subsequent statements without any differentiation. In the wake of the machine learning task, these examples could be seen as outliers that deviate from 'normal'

predicting behavior, and could be accounted for by machine learning algorithms. Another perspective to this issue is that traditionally such text samples would be rare. These text samples wouldn't necessarily have references to the semantic information present in the thesis subsequently in the paragraph. It is this assumption that is behind the development of the proposed feature. Therefore that particular concern could be discounted.

The proposed method calculates paragraph coherence by calculating the semantic similarity between the first sentence in a sample text (considered the thesis statement) and the rest of the sentences that follow the thesis. The calculation of semantic similarity between the sentences is limited to nouns and verbs. The assumption behind this choice is that nouns and verbs carry the most semantic information. Additionally, this also allows the number of calculations to be restricted to a smaller number, thereby reducing computational complexity.

The calculation of Paragraph coherence is a two-step process. Initially, the calculation of all the semantic similarities of the words in the first sentence with every other word in the rest of the sentences in the paragraph is performed. This step itself is composed of two steps. For each word, w_j present in the first sentence (otherwise named as the topic sentence), the semantic similarity between itself and every relevant word, w_i in the rest of the paragraph is calculated. Let this value be s_{ij}. Here m is the maximum value of i, *i.e.* the total number of relevant words present in the paragraph (with the exception of the topic sentence). Therefore for a word w_j present in the topic sentence, the associated semantic similarities with the rest of the paragraph is formulated as below. Let $g(w_j)$ be this measure. Then:

$$g(w_j) = \frac{1}{m}\sum_{i}^{m} s_{ij} \qquad -(5)$$

Let n be the total number of relevant words in the topic sentence. Then the total associated semantic similarity value of the paragraph could be treated as, sum.

$$sum = \sum_{j}^{n} g(w_j) \qquad -(6)$$

We would want to keep the value of paragraph coherence between 0 and 1. Paragraph coherence, P could be calculated as:

$$P = e^{-1*sum} \qquad -(7)$$

For example, consider the below statement, taken from Dr. Suedfeld's Integrative complexity training workshop page (Suedfeld et al. 1992):

> *"As for myself, I do not fear death, nor do I look forward to it. There is no appropriate time for death; if one conceives life as a dialectic, one realizes that issues are never settled once and for all. When every item on my list is completed a new list of items is generated. Relationships are never fulfilled: the deeper a relationship becomes the more nurturance and care it generates. In fact I am not exactly in agreement with the choice points of Eriksons last stage - integrity versus despair. While despair is certainly the negative outcome, I am uncertain that integrity - or acceptance of one's life as "good" - is the desirable resolution. For me, death simply means that the eternal struggle has ended."*

The sample text given above scored high on Integrative Complexity. Our proposed method scored a P (Paragraph Coherence) value of 0.4729.

> *'The group of men whom you got together in April in New York for Zionist work have, in the main, been rather disappointing in performance. I am glad you are coming east soon, and I trust that you will be able to impress them with the sanctity of a promise, and secure performance. Very truly yours,"*

Whereas the text given above (also taken from Suedfeld et al. (1992)) scores low on Integrative complexity. And has a *P* score of 0.310. Therefore it could be seen that Paragraph coherence could act as a predictor for scoring Integrative complexity.

3.4 Methodology

The data for these experiments were taken from Suedfeld's Integrative Complexity training workshop page (Suedfeld et al. 1992). They consisted of 83 text samples along with their original scores, scored by trained coders. Each instance has been scored on a 1-7 scale. All of the instances were binned into three classes. This pre-processing step was deemed essential in order to simplify the machine learning task from being a 7- class classification problem into a 3-class classification problem. Instances that belonged to IC scores of 1 or 2, were classified as having low IC and therefore given a class label of 'low'. Subsequently, instances that have been given IC scores of 3, 4 or 5 were classified as having moderate levels of IC, and were given a class label of 'mid'. Instances that were scored IC scores of 6, 7 were classified as having high IC and were given class label of 'high'. This was considered reasonable since 2, 4 and 6 are basically transition scores. Transition scores are given to a text sample if a scorer faces ambiguity while scoring a text sample. See Suedfeld et al. (1992) for further explanation.

The method for determining the value for the Paragraph Coherence feature was coded in SWI Prolog (Wielemaker, Schrijvers, Triska & Lager, 2012). The code made use of WordNet 3.0 (Fellbaum, 1998 ed; Fellbaum, 2010; Miller, 1995) written in Prolog. Then the code was run on each instance to calculate the paragraph coherence of each instance. The code for calculating the length of a paragraph (in words) was also written in Prolog. This was then followed by cleaning of the data and then converted into an ARFF (Attribute

Relation File Format) file format for use in Weka. (Hall, Frank, Holmes, Pfahringer, Reutemann, & Witten, 2009). String to Word Vector filter in Weka (Hall et al., 2009) was used for converting the string in the text attribute of each instance is converted to a set of attributes representing word occurrences. Feature selection methods played a huge role in this text-classification problem. Paragraph Coherence, Length of Paragraph were used along with a bag of words approach for experimentation. The number of attributes were greatly reduced using Attribute Selection methods.

For classification purposes, machine learning algorithms were employed. the open source machine learning software, Weka (Hall et al., 2009) was used for experimenting with the data. After intensive experimentation with numerous machine learning algorithms, the ones which have reported some of the best performances are mentioned in the paper. They are Bagging, the Multinomial Logistic Regression model, Multi-layer perceptron, AdaBoost.M1 and the Multi-class classifier.

A boosting algorithm, called Adaboost (short for Adaptive Boosting) was used significantly improve classifier performance. Traditionally, weak learners can predict with a rates a little better than random guessing. In boosting, weak learning algorithms are run on different parts of the distribution of the training data and then combined to form a composite classifier. (Freund & Schapire, 1996). AdaBoost.M1 is a special case of AdaBoost, where easy examples that are correctly classified by the weak learning algorithms are given less weightage than examples that get misclassified by the weak learning hypotheses. AdaBoost.M1 was found to work particularly well in classification tasks where the performances of base classifiers could be improved by boosting.

Natural Language Applications often use the Multinomial Logistic Regression Model. This is mostly due to the fact that the model does not assume statistical independence of its features, as is often the case with text. A generalization of the logistic regression model for multi-class problems, the algorithm defines the probabilities describing the outcomes of an instances as a function of its features, using a logistic function.

the Multi-class classifier- suitable for the multi-class classification problem in Weka was another meta classifier used for experimentation. Logistic regression and the multi-layer perceptron were used as base classifiers. Popular multi-classification methods like 1-against-1 and pairwise classification were also used.

The MultiLayer Perceptron Algorithm was also used for experimentation. It is composed of multiple layers of nodes arranged in a directed graph. The network uses a supervised learning algorithm known as Backpropagation for training the classifier. The activation functions used were sigmoid functions.

One of the learning algorithms used for running the experiments was Bagging (also known as Bootstrap Aggregation). It is used to improve stability and accuracy of base algorithms. This algorithm grants ‘votes’ to base classifiers that are trained on different bootstrap samples. A final classifier is built from all the base classifiers trained on all the bootstrap samples, whose prediction is based on the most predicted by its base classifiers.

3.5 Evaluation and Results

Stratified 10-fold cross-validation was used to evaluate the performance of the machine learning algorithms. The performance measures used were classification accuracy and effectiveness measures such as precision, recall and F-1 measures. Effectiveness is a generic term for precision, recall and F-1 measures. Priority should be given to the

effectiveness measures since our dataset is relatively small. The validity for Effectiveness measures is higher in evaluating the performance of machine learning algorithm on smaller datasets (Manning, Raghavan, Schütze, 2008). Macro-averaged results have been computed to get a better sense of effectiveness over classes containing small numbers (< 100).

Results obtained were promising. It could be seen that the addition of Paragraph Coherence and the length of the text sample as features have influenced the performance of the algorithms in a positive manner. The combination of these features along with a bag of word approach have produced a decent performance.

Table 3.1 and Table 3.2 show the precision, recall and F-1 measures of the classifications. Table 3.3 shows the classification accuracies. The highest classification accuracy was reported by the Multinomial Logistic Regression Model with a ridge estimator-II. The same classifier also reported the highest precision and recall. Some of the classifiers have high precision (1.000) for the class high. While some of them have high recall for the class mid. Table 3.4 and Table 3.5 show the confusion matrices for those algorithms that report the highest amount of classification accuracy on the dataset.

Classification accuracies of 80% to 83% have been reported. These accuracies are at par (or slightly above) the accuracies that are reported with the human rater reliability which stands at 80%. However the relatively small size of the dataset denies us the opportunity to fully rely on the reported methods. A closer look at these methods with a larger dataset could affirm or deny the stated hypothesis.

The Confusion Matrix for the Multinomial logistic regression with a ridge estimator-II (shown in Table 3.4) shows promising results for the classes ‘mid’ and ‘high’ despite the

obvious class imbalance issues. This improvement in performance could be attributed to the contribution of the NLP-feature, Paragraph Coherence. Among the 24 instances of class 'low', 12 were misclassified as 'mid', and 12 were classified correctly. However the 49 instances of 'mid' were classified correctly. Of all the 10 instances that belonged to class 'high', 8 were classified correctly, where 2 were classified incorrectly as 'low'. It should be noted that except for 2 instances, the algorithm was able to correctly differentiate between 'mid' and 'high', and 'low' and 'high'.

Table 3.1: Effectiveness measures for Bagging, Multi-Class Classifier and Multinomial Logistic Regression with a ridge estimator -II

Class	**Bagging**			**Multi-Class Classifier**			**Multinomial logistic regression with a ridge estimator-II**		
	Precision	**Recall**	**F-1 measure**	**Precision**	**Recall**	**F-1 measure**	**Precision**	**Recall**	**F-1 measure**
low	0.800	0.500	0.615	0.923	0.500	0.649	0.857	0.500	0.632
Mid	0.790	1.000	0.883	0.762	0.980	0.857	0.803	1.000	0.891
high	1.000	0.600	0.750	0.857	0.600	0.706	1.000	0.800	0.889
Weighted. Avg.	0.818	0.807	**0.790**	**0.820**	**0.795**	**0.779**	**0.843**	**0.831**	**0.816**

Table 3.2: Effectiveness measures for AdaBoostM1- II, Multi-layer perceptron and AdaBoost M1-I.

Class	AdaBoostM1- II			Multi-layer Perceptron			AdaBoostM1- I		
	Precision	**Recall**	**F-1 measure**	**Precision**	**Recall**	**F-1 measure**	**Precision**	**Recall**	**F-1 measure**
low	0.846	0.458	0.595	1.000	0.417	0.588	0.593	0.667	0.627
mid	0.742	1.000	0.852	0.716	0.980	0.828	0.784	0.816	0.800
high	1.000	0.400	0.571	0.833	0.500	0.625	1.000	0.500	0.667
Weighted. Avg.	**0.803**	**0.771**	0.744	**0.813**	0.759	0.734	0.755	0.735	0.734

Table 3.3: Classification Accuracies

Classifier	**Specifications and comments**	**Accuracy.**
Multi-layer Perceptron		75.9%
AdaBoostM1- I	*Base classifiers and their weights:* *Random forest of 10 trees, each constructed while considering 5 random features.*	73.5%
AdaBoostM1- II	*Base classifier: SMO with Polykernel*	77%
Multi-Class Classifier	*Base classifier: Multinomial logistic regression with a ridge estimator* *Method: 1-against-all*	79.5181 %
Bagging	*Base classifier: Multinomial logistic regression with a ridge estimator*	**80.7229 %**
Multinomial logistic regression with a ridge estimator-I		**80.7229 %**
Multinomial logistic regression with a ridge estimator-II	*Uses ConjugateGradientDescent*	**83.1325%**

Table 3.4: Confusion Matrix for the Multinomial logistic regression with a ridge estimator-II

Low	Mid	High	← classified as
12	12	0	Low
0	49	0	Mid
2	0	8	High

Table 3.5.: Confusion Matrix for Bagging

Low	Mid	High	← classified as
12	12	0	Low
0	49	0	Mid
3	1	6	High

3.6 Conclusion

The proposed approach produced classification accuracies ranging from 75% to 83%, which are a first in the literature for automated scoring of integrative complexity. More experiments on a much larger dataset could establish the proposed hypothesis. It was observed that the addition of the two NLP features, Paragraph Coherence and Length of the text has predictably improved performance over the earlier model reported in Section 2. It is of significant importance to note that the second model reported performances on a smaller dataset that the earlier model. Although the Natural language processing feature did draw upon some semantic understanding of the problem, it is obvious that the feature needs to be improved upon: drawing on the other finer qualities of how differentiation and integration appear to the human reader. By imparting our natural instincts on the

understanding of the structure of thought, we could optimistically create deeper natural understanding which could yield improved performances. It is the convergence of the knowledge engineering approach and the machine learning approach that has produced a marked improvement.

CHAPTER 4

CONCLUSION & FUTURE DIRECTIONS

This thesis has addressed the need for automating the scoring of Integrative Complexity. We obtained accuracies in the automated scoring of Integrative Complexity in the range of 75% to 83% in a 3- way classification, which is a marked improvement from previously published work on automated scoring of Integrative Complexity. Manual scoring methods usually obtain acceptable classification accuracies of 80 to 85%. The premise that justifies the handling of this classification problem as a 3 class problem, instead of a 7-class problem is simply that it intuitively makes more sense to classify text on levels of low integrative complexity, moderate integrative complexity and high integrative complexity. Moreover, the 3-class problem is computationally more viable than a 7-class problem.

Integrative Complexity is used to capture the complexity of the cognitive processes, rather than the variables that influence the cognitive strategies used in formulating the passages. For instance, two passages that describe the same topic in diverse ways could receive the same IC score. Although measuring the extent of the complexity of thought on a general level is a significant issue, the construct has a drawback in that it limits the theorizing on why a particular paragraph of speech or written text is complex (Conway et al. 2011).

In response to this limitation, Conway (2008) developed a model of complex thinking. The model was based on the hypothesis that people can be complex in two different ways

within the Integrative Complexity system. The two developed measurements for each of these two routes are: Dialectical Complexity (DC) and Elaborative Complexity (EC). Dialectical complexity involves the measurement of the implicit recognition of tension between different perspectives on an issue. Some determinants of DC involve the following: acknowledgment of competing perspectives as valid and recognition of qualifications as contributing to the issue. Elaborative Complexity on the other hand, is related to how a singular dominant theme can be developed in a complex way. Determinants of this measure include the development of clear differentiated perspectives on an issue, the presence of a dominant theme or idea for which differing sources can be used as evidence for the issue.

In other words, IC could be seen as a construct that measures the structural aspect of thoughts (envisioned in speech or text) rather than the content of the passage. In this manner, IC measurement is un-prejudiced and un-biased. A democratic method for measuring an individual's capacity/ability to gauge all known perspectives on a particular issue and then reach a rational conclusion by integrating the said perspectives. As grandiose as the promises are that are being made of Integrative Complexity and its successors, it would be a formidable feat to automate the scoring process. This has one significant far-reaching consequence: that the discovery would itself enable the creation of a computer program that could create highly differentiated or lowly integrated text or its various combinations as needed. It would indeed mark the step towards natural language understanding, a step further from natural language processing.

Future Work could focus on automating the processes for scoring Elaborative Complexity and Dialectical complexity. More work also should focus on the knowledge

engineering perspective, rather than a simplistic statistical approach. A deeper appreciation of the problem of assessing the structure of human thought could produce promising and elegant results. Such an approach could be obtained through the inter-disciplinary study of linguistics, cognitive science and psychology. A fresh approach to the problem would involve all the aforementioned elements with a training dataset that has been manually scored by expert coders. This could allow the effective implementation of an automated scores that doesn't move far from the ideal scoring system. Future work should definitely harass large amounts of data to enable automated integrative complexity to be a foreseeable reality. Semi-supervised learning algorithms could be used in the case of small amounts of labeled data.

REFERENCES

Abe, J. A. A. (2011). Changes in Alan Greenspan's language use across the economic cycle: A text-analysis of his testimonies and speeches. *Journal of Language and Social Psychology*, 30, 212-223.

Abe, J. A. A. (2012). Cognitive–affective styles associated with position on war. *Journal of Language and Social Psychology, 31*(2), 212-222.

Baker-Brown, G., Ballard, E. J., Bluck, S., de Vries, B., Suedfeld, P., & Tetlock, P. (1992). The integrative complexity coding manual. In C. Smith (Ed.), *Handbook of thematic analysis* (pp. 605-611). Cambridge, England: Cambridge University Press.

Baker-Brown, G., Ballard, E. J., Bluck, S., de Vries, B., Suedfeld, P., & Tetlock, P. (1990). Coding manual for conceptual/integrative complexity. University of British Columbia and University of California, Berkeley.

Bieri, J. (1955). Cognitive complexity-simplicity and predictive behavior. *Journal of Abnormal and Social Psychology,* 51, 263-268.

Bieri, J. (1971). Cognitive structures in personality. *Personality theory and information processing. New York: Ronald,* 178-208.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Bruch, M. A., Juster, H. R., & Kuethe, M. (1985). Conceptual complexity as a mediator of negative thoughts and affect in socially anxious individuals. *British Journal of Cognitive Psychotherapy*, 3, 59-69.

Bruch, M. A., McCann, M., & Harvey, C. (1991). Type A behavior and processing of social conflict information. *Journal of Research in Personality*, 25, 434-444.

Conway, L. G., Conway, K. R., Gornick, L. J., & Houck, S. C. (2014). Automated integrative complexity. *Political Psychology*..

Conway III, L. G., Dodds, D. P., Towgood, K. H., McClure, S., & Olson, J. M. (2011). The biological roots of complex thinking: are heritable attitudes more complex? *Journal of personality*, *79*(1), 101-134.

Conway, L. G., III. (2008). Coding dialectical complexity and elaborative complexity: A supplement to the integrative complexity coding system. Unpublished manuscript, The University of Montana. Available online at

http://psychweb.psy.umt.edu/conway/Documents/dialectical_elaborative_manual_final.doc

Conway, L.G., III, Suedfeld, P., & Tetlock, P.E. (2001). Integrative complexity and political decisions that lead to war or peace. In D.J. Christie, R.V. Wagner, & D. Winter (Eds.), *Peace, conflict, and violence: Peace psychology for the 21st century* (pp. 66–75). Englewood Cliffs, NJ: Prentice-Hall.

Coren, S., & Suedfeld, P. (1995). Personality correlates of conceptual complexity. *Journal of Social Behavior & Personality*.

Curşeu, P. L., Schruijer, S., & Boroş, S. (2007). The effects of groups' variety and disparity on groups' cognitive complexity. *Group Dynamics: Theory, Research, and Practice*, 11(3), 187.

Feist, G. J. (1994). Personality and working style predictors of integrative complexity: A study of scientists' thinking about research and teaching. *Journal of Personality and Social Psychology*, 67, 474-484.

Fellbaum, C. (1998, ed) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Fellbaum, C. (2010). WordNet: An electronic lexical database. 1998. *WordNet is available from http://www. cogsci. princeton. edu/wn*

Freund, Y. & Schapire, R.E. (1996). Experiments with a new boosting algorithm, in L. Saitta, ed., Machine Learning: *Proceedings of the Thirteenth National Conference*, Morgan Kaufmann, pp. 148- 156.

Gottschalk, L. A. (1995). *Content analysis of verbal behavior: New findings and clinical applications*. Lawrence Erlbaum Associates, Inc.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*, 193-202

Gruenfeld, D. H., & Hollingshead, A. B. (1993). Sociocognition in work groups: The evolution of group integrative complexity and its relation to task performance. *Small Group Research*, 24, 383-405.

Hall M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. (2009). The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Volume 11, Issue 1.

Joachims, T. (1998). *Text categorization with support vector machines: Learning with many relevant features* (pp. 137-142). Springer Berlin Heidelberg.

Joachims, T., & Sebastiani, F. (2002). Guest Editors' Introduction to the Special Issue on Automated Text Categorization. *Journal of Intelligent Information Systems*, *18*(2), 103-105.

Jordan, T. (1998). Structures of geopolitical reasoning–A constructive-developmental approach. A preliminary exploration of theoretical framework and method. *Occasional Papers*, *9*.

Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton, 2 vols.

Knight, K. 1999. Mining online text. *Communications of the ACM 42*, 11, 58–61.

Leary, M. R., & Hoyle, R. H. (Eds.). (2009). *Handbook of individual differences in social behavior*. Guilford Press

Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied statistics*, 191-201.

Lee, F., & Peterson, C. (1997). Content analysis of archival data. *Journal of Consulting and Clinical Psychology*, 65, 959-969.

Li, Y., Bandar, Z., McLean, D., & O'Shea, J. (2004). A Method for Measuring Sentence Similarity and its Application to Conversational Agents. In *FLAIRS Conference* (pp. 820-825).

Li, Y., Bandar, Z. A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *Knowledge and Data Engineering, IEEE Transactions on*, *15*(4), 871-882

Liht, J., & Savage, S. (2013). Preventing violent extremism through value complexity: Being Muslim Being British. *Journal of Strategic Security*, *6*(4), 3.

Liotti, G. (1987). The resistance to change of cognitive structures: A counterproposal to psychoanalytic metapsychology. *Journal of Cognitive Metapsychology*, 1, 87-104.

Manning, C. and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, US.

Manning, C. D., Raghavan, P., & Schütze, H. (2008) Text classification and Naive Bayes. In *Introduction to Information Retrieval* (Vol. 1, p. 6). Cambridge: Cambridge university press

Matsumoto, D., & Hwang, H. C. (2013). The Language of Political Aggression. *Journal of Language and Social Psychology*, *32*(3), 335-348.

Meadow, C.T., Boyce, B.R. and Kraft, D.H. (2000). *Text Information Retrieval Systems*. 2nd. Ed. Academic Press

Milbank, D. (2011, April 26). Obama, lost in thought. *The Washington Post*. Retrieved from: http://www.washingtonpost.com/opinions/obama-lost-in-thought/2011/04/26/AF0FrwsE_story.html

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39-41.

Oliver, E. J., Markland, D., Hardy, J., & Petherick, C. M. (2008). The effects of autonomy-supportive versus controlling environments on self-talk. *Motivation and Emotion, 32,* 200-212.

Pazienza, M. T. Ed. 1997. Information extraction. Number 1299 in *Lecture Notes in Computer Science*. Springer, Heidelberg, DE.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC [Computer software]. Austin, TX: LIWC.net

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality, 36,* 271-282.

Peterson, R. S., Owens, P. D., Tetlock, P. E., Fan, E. T., & Martorana, P. V. (1998). Group dynamics in top management teams: Groupthink, vigilance, and alternative models of organizational failure and success. *Organizational Behavior and Human Decision Processes,* 73, 272-305.

Platt, J. C. (1999). Fast training of support vector machines using sequential minimal optimization. In *Advances in kernel methods* (pp. 185-208). MIT press.

Porter, C. A., & Suedfeld, P. (1981). Integrative complexity in the correspondence of literary figures: Effects of personal and societal stress. *Journal of Personality and Social Psychology*, 40, 321-330.

Raz-Duvshani, A. (1986). Cognitive structure changes with psychotherapy in neurosis. *British Journal of Medical Psychology*, 59, 341-350

Schroder, H. M., & Suedfeld, P. (Eds.). (1971). *Personality theory and information processing*. New York: Ronald

Schroder, H. M., Driver, M. J. And Streufert, S. (1967) *Human Information Processing. Individuals and Groups Functioning in Complex Social Situations,* New York: Holt, Rinehart and Winston, Inc.

Scott, W. A., Osgood, D. W., Peterson, C., & Scott, R. (1979). *Cognitive structure, theory and measurement of individual differences*. Washington, DC: VH Winston.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR), 34*(1), 1-47.

Smith, C. P. (Ed.). (1992). *Motivation and personality: Handbook of thematic content analysis*. Cambridge University Press.

Strohmer, D. C., Biggs, D. A., Haase, R. F., & Bruch, M. A. (1983). Cognitive style and synchrony in measures of anxiety. *Measurement and Evaluation in Guidance*, 16, 13-17.

Suedfeld, P. (2010). The cognitive processing of politics and politicians: Archival studies of conceptual and integrative complexity. *Journal of personality*, *78*(6), 1669-1702.

Suedfeld, P., & Bluck, S. (1988). Changes in integrative complexity prior to surprise attacks. *Journal of Conflict Resolution*, 626-635.

Suedfeld, P., & Bluck, S. (1993). Changes in integrative complexity accompanying significant life events: Historical evidence. *Journal of Personality and Social Psychology*, 64, 124-130.

Suedfeld, P. & Eichhorn, D. (2013). General Format. Retrieved from http://www2.psych.ubc.ca/~psuedfeld/Download.html#articles.

Suedfeld, P., & Leighton, D. C. (2002). Early communications in the war against terrorism: An integrative complexity analysis. *Political Psychology*, 23, 585–599.

Suedfeld, P., & Piedrahita, L. E. (1984). Intimations of mortality: Integrative simplification as a precursor of death. *Journal of Personality and Social Psychology, 47*(4), 848.

Suedfeld, P., & Tetlock, P. (1977). Integrative complexity of communications in international crises. *Journal of conflict resolution*, *21*(1), 169-184.

Suedfeld, P., Leighton, D. C., & Conway, L. G., III. (2006). Integrative complexity and decision-making in international confrontations. In M. Fitzduff & C. E. Stout (Eds.), *The psychology of resolving global conflicts: Nature vs. nurture* (Vol. 1, pp. 211–237). Westport, CT: Praeger.

Suedfeld, P., Tetlock, P. E., & Streufert, S. (1992*).* Conceptual/integrative complexity. In C. P. Smith, J. W. Atkinson, D. C. McClelland, and J. Veroff (Eds.), *Motivation and personality: Handbook of thematic content analysis*, 393-400. New York: Cambridge University Press

Suedfeld, P., Tetlock, P., & Ramirez, C. (1977). War, peace, and integrative complexity. *Journal of Conflict Resolution*, 21, 427-442.

Suedfeld, P., Wallace, M. D., & Thachuk, K. L. (1993). Changes in integrative complexity among Middle East leaders during the Persian Gulf crisis. *Journal of Social Issues*, 49, 183–199.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 22, 24-54.

Tetlock, P. E. (1979). Identifying victims of groupthink from public statements of decision makers. *Journal of Personality and Social Psychology*, *37*(8), 1314.

Tetlock, P. E. (1983). Cognitive style and political ideology. *Journal of Personality and Social Psychology*, 45, 118-126.

Tetlock, P. E. (1984). Cognitive style and political belief systems in the British House of Commons. *Journal of Personality and Social Psychology*, 46, 365-375.

Tetlock, P. E. (1985). Integrative complexity of American and Soviet foreign policy rhetoric: A time-series analysis. *Journal of Personality and Social Psychology*, *49*(6), 1565.

Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology*, 50, 819-827.

Tetlock, P. E., & Boettger, R. (1989). Cognitive and rhetorical styles of traditionalist and reformist Soviet politicians: A content analysis study. *Political Psychology*, 10, 209-232.

Tetlock, P. E., & Manstead, A. S. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review*, *92*(1), 59.

Tetlock, P. E., Hannum, K. A., & Micheletti, P. M. (1984). Stability and change in the complexity of senatorial debate: Testing the cognitive versus rhetorical style hypotheses. *Journal of Personality and Social Psychology*, 46, 979-990.

Tetlock, P. E., Peterson, R. S., & Berry, J. M. (1993). Flattering and unflattering personality portraits of integratively simple and complex managers. *Journal of Personality and Social Psychology*, 64, 500-511.

Walker, S. G., & Watson, G. L. (1994). Integrative complexity and British decisions

during the Munich and Polish crises. *Journal of Conflict Resolution*, 38, 3-23.

Wallace, M. D., & Suedfeld, P. (1988). Leadership performance in crisis: The longevity-complexity link. *International Studies Quarterly*, 32, 439–451.

Wallace, M. D., Suedfeld, P., & Thachuk, K. (1993). Political rhetoric of leaders under stress in the Gulf crisis. *Journal of Conflict Resolution*, 37, 94–107.

WEKA 3- Data Mining with open source Machine Learning software in Java. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/. [Accessed: April 11, 2013].

Weber, R. P. (1990). Basic content analysis (2nd ed.). Newbury Park, CA: Sage.

Wielemaker, J., Schrijvers, T., Triska, M., & Lager, T. (2012). Swi-prolog. *Theory and Practice of Logic Programming*, *12*(1-2), 67-96.

Winter, D. A. (1993). Slot rattling from law enforcement to law-breaking: A personal construct theory exploration of police stress. *International Journal of Personal Construct Psychology,* 6, 253-267.

Winter, D. G. (1996*). Personality: Analysis and interpretation of lives*. New York: McGraw-Hill.

Winter, D. G. (2007). The role of motivation, responsibility, and integrative complexity in crisis escalation: comparative studies of war and peace crises. *Journal of Personality and Social Psychology*, *92*(5), 920.