2009 International Conference on Information Technology and Computer Science

An Algorithm for Detection of Cognitive Intentionality in Text Analysis

Yi Guo
Department of Computer Science and Engineering
East China University of Science and Technology
Shanghai, 200237, China
yguo1110@ecust.edu.cn

Nan Hua
The Telecommunication Engineering Institute
The Air Force Engineering University
Xi'an, 710077, China
ehn@163.com

Zhiqing Shao
Department of Computer Science and Engineering
East China University of Science and Technology
Shanghai, 200237, China
zshao@ecust.edu.cn

Abstract—**Intentionality is a cognitive concept applied by many researchers under different circumstances or backgrounds to explain how human beings engage the world so as to understand it well. In recent a couple years, research works of automatic text analysis have changed focus to more fine-grained analysis tasks with emphasis on text elements for subjective information. Intentionality detection is another creative and promising research direction in text analysis.**

This paper presents an innovative research effort focuses on cognitive intentionality, which belongs to the category of subjective information, and complement the current research works in automatic recognition and analysis of subjective information. The cognitive intentionality detection algorithm, CIDA, is composed with six key modules, Event Identification, Syntactic Parsing, Elementary Action Extraction, Lexical-Semantic Analysis, P-Tree Construction, and Intentionality Extraction. The experimental results have testified the efficiency and accuracy of CIDA.

Keywords- Lexical-Semantic Analysis; Intentionality; cognitive; Event Identification

I. Introduction

In recent years, research works relevant with automatic text analysis have shifted concentration from a pure broad-topic classification to more fine-grained analysis tasks with emphasis on text elements for subjective information, including subjectivity identification [24], opinion detection [5][8], semantic polarity analysis [23][7] certainty identification [16], and perspective identification [12].

Intentionality is a cognitive concept applied by many academic scholars under different aims or backgrounds to explain how human beings engage the world so as to understand it well. Intentionality is a debated notion in philosophy.

Intentionality detection, another creative and promising research direction in text analysis, is fundamental to our ability of understanding both the general laws that govern events and the particular principle of how and why a specific event actually happened, in everyday life.

While previous research works have explored and investigated the automatic recognition and analysis of a subset of subjective information, namely opinions, subjectivity, certainty, perspective and sentiments, this paper presents a research effort focuses on cognitive intentionality (e.g. plans, promises and suggestions), which also falls in the category of subjective information, and whose automatic detection has largely been neglected so far.

The rest of paper is organized as follows: section 2 explains several important concepts closely connected with intentionality; section 3 describes our cognitive intentionality detection algorithm (short as CIDA in following parts); section 4 introduces implementation of CIDA and experimental results; and section 5 concludes this paper with appropriate discussion.

II. Intentionality Relevant Concepts

In philosophy, an ***event*** is an object in time or an instantiation of properties in an object. However, a definite definition has not been reached. In this paper, "event" refers to a cognitive psychological concept, and can be either a story or a sentence in microstructure. As Zwaan and Radvansky [25] treated every single sentence as an event, "event" and "simple sentence (clause)" are rendered as equal concepts in our research work as well.

Temporality is relevant to human cognition. When successive events are causally independent, people's attributions tend to show a recent effect [14]. Temporality represents the time or temporal information of events influencing the situation changing or updating. Presumably, events are narrated in their chronological order by default, as this is the experience of events in everyday life.

Hart and Honoré [2] investigated the case for an interaction between ***location*** and intentionality. They argued that when events are of equivalent status, either two voluntary actions or two physical events, the most recent event is preferred, but a voluntary action is always preferred to a physical cause, irrespective of location.

Causality indicates a relation judgment about how some concepts bias the occurrence of other concept [15]. The causality deals with events that happen in every day for some reasons. Shaver [21] provided a comprehensive theory of responsibility, which describes "causality" as the protagonist's causal contribution to the production of the

978-0-7695-3688-0/09 $25.00 © 2009 IEEE
DOI 10.1109/ITCS.2009.176

effect, and "intentionality". Shaver construed a proximate and generative relation between action and outcome. Researchers have been driven for decades to explore causal explanations for how and why things happen [2][3][4][6][22]. However, determining causes is complex in practice, which is especially true when the appreciation of different perspectives may be required, and the success or failure of actions may be judged differently according to original motivations. In most everyday situations there will be numerous events or factors that could be cited as causes of a particular outcome.

In Shaver's theory [21], intentionality and ***foreseeability*** play a double role; they are separate dimensions for attributions of responsibility, and are both sub-components in the causality dimension. Shaver [21] argues, in line with Heider [3], that the more a protagonist's anticipates the negative consequences of their action, the more they should be considered a cause of that outcome. When an obstacle has to be overcome, the causal force of an action is augmented. In this case, the obstacle is the protagonist's foresight about negative consequences of his action.

III. Algorithm Description

Section II has explained five important concepts closely connected with intentionality and the following detection algorithm. As illustrated in figure 1, CIDA is composed with six interlinked modules, which will be discussed in detail.

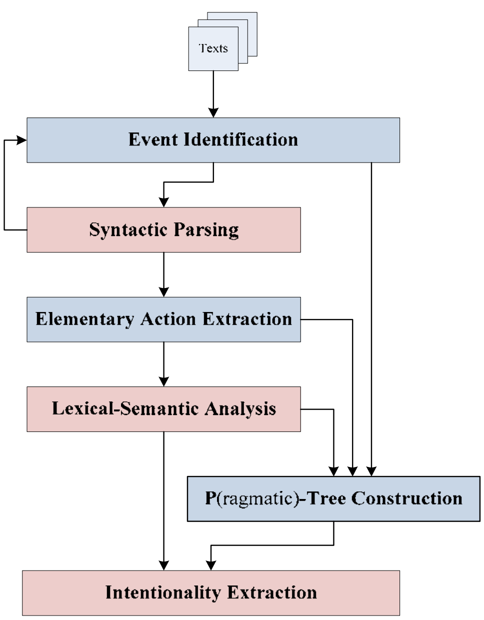

Figure 1. Structure of CIDA

A. Event Identification

Since event can be represented with a sentence in microstructure, this module takes the task to extract simple sentences and clauses from complex sentences as events. This task is executed in two levels, simple and complex sentence levels.

In simple sentence level, each original single sentence is identified as an event. While, in complex sentence level, complex sentences will be forwarded to the Syntactic Parsing module and then feedback (as shown in figure 1) with proper syntactic attributes to identify underling clauses and relations among them.

All extracted events, together with necessary syntactic attributes, are delivered to "P(ragmatic)-Tree Construction" module for further processing.

B. Syntactic Parsing

Link Grammar Parser (LGP) [1] can parse any sentence to a syntactic structure, which consists of a set of labeled links connecting pairs of words, and a constituent-tree, which shows conventional constituents (e.g., noun phrases, verb phrases, and prepositional phrases).

Among the verbs taking to-infinitives, a high number of verbs, for example, the verbs 'intend' and 'hope', with intentional meaning can be identified.

Meanwhile, six different types of realizations of a clausal complement can be distinguished in English: (1) to-infinitive in present tense, (2) to-infinitive in past tense, (3) that-clause, (4) dependent declarative clause, (5) wh-clause, and (6) whether-clause.

C. Elementary Action Extraction

In this module, elementary actions are extracted from constituent-trees generated by the syntactic parser, LGP. Elementary actions normally refer to verb phrases and other constituents that take functions of predicates in sentences or clauses.

Here, temporality is considered to construct continuity of elementary actions, as the successive elementary actions correspond to a extreme branches that make up a high level action abstract in P(ragmatic)-Trees [19].

D. Lexical-Semantic Analysis

In our research work, the automatic identification of intentionality is based on the use of lexical clues. A major part of these clues constitutes lexical-semantic classes, which are fully explored to capture close relationships between syntax and semantics of verbs, nouns, and adjectives.

Using lexical-semantic classes as clues for a specific text analysis task, such as intentionality detection, has two major advantages.

Firstly, several lexical-semantic classes, varying in size and coverage, have already been constructed to capture the syntax-semantics relationship; for English, there is Levin's large verb classification [10] and its extension by [9] based among other sources on Rudanko's work on classes of verbs taking clausal complements [17][18].

Secondly, and even more important, the definition of lexical-semantic classes with pure syntactic properties makes it possible to acquire these classes, in automatic manner, from corpus data, which has been shown to be feasible by several research attempts [13][20] recently made in this area.

In first step, we list common syntactic properties of classes that the verbs belonging to. These common syntactic properties provide the basis for the semi-automatic acquisition of the verb classes from corpus data.

Afterwards, we execute the process of manually filtering the lexical-semantic classes in order to get lexical clues for intentionality.

In the final step, we present taxonomies of the resulting lexical-semantic classes in order to provide a characterization of the verbs belonging to these classes. Two main classes are distinguished: (1) verbs presenting actions as intentionality, and (2) verbs presenting propositions as intentionality.

E. P(ragmatic)-Tree Construction

All the processing results of "Event Identification", "Elementary Action Extraction", and "Lexical-Semantic Analysis" are merged into this module to construct P(ragmatic)-Trees [19], which are, in the case of discourse, classifiable into a number of recursively applicable types.

P(ragmatic)-Tree, as a pragmatic structure, represents the intentionality of behavior, and constitutes a level that underlies the more superficial syntactic levels of discourse. In this module, four types (events, syntax, elementary action and lexical-semantic) of information are processed in a comprehensive manner. The construction algorithm includes eight steps, which are described below.

Algorithm (P-Tree Construction)
Input: Events, Syntax, Elementary Action, Lexical-Semantic information ***Output***: P-Tree(s)
Begin (1) Sort events according to temporality; (2) Cluster events with protagonist(s) as first index and location(s) as second index; (3) Sort elementary actions with temporality; (4) Cluster elementary actions with protagonist(s) as first index and syntactic object(s) as second index; (5) Use events and elementary actions to construct the fundamental P-Nodes; (6) Use the cluster information to construct a upper level of P-Nodes; (7) In the upper level repeat the Sort and Cluster operations to construct a higher upper level of P-Nodes; (8) Finally all the P-Nodes will converge to minimum point(s), called Root(s). ***End***

F. Intentionality Extraction

From the P-Tree Construction algorithm denoted in section 3.5, we can visualize that all the events and elementary actions finally merge into one or more trees. One tree shows that all the events and actions are connected for a common purpose or intention; different trees present various purposes contained in the original input textual information.

Thus, we conclude that, in a tree (P-Tree), each branch is the intention of all its leaves (P-Nodes) or sub-branch(s). From root(s) to the very end leaves, we can extract many different intentionality chains with various lengths, which refer to number of levels between. Intentionality extraction is carried out in clausal and elementary action (verbs) levels.

Algorithm (Intentionality Extraction)
Input: P-Tree(s) and Lexical-Semantic information ***Output***: a collection of intentionality
Begin (1) Find and Conclude the number of P-Tree(s); (2) Trace each branch and its sub-branches, sub-sub-branches, …, and end leaves from a P-Tree Root; (3) Record all the leaf-to-root chains as intentionality chains; (4) Repeat (2) and (3) for each P-Tree. ***End***

IV. IMPLEMENTATION AND EXPERIMENTS

Our CIDA is programmed and implemented with ActivePerl (v5.10) from ActiveState, C# of Microsoft Visual Studio .NET 2005, and Microsoft SQL Server 2005, in Microsoft Windows XP Professional, using a workstation computer (Dell Precision M4300, Dual-core Intel™ Core® 2 Duo 2.40GHz, 4G DDR2 SDRAM).

Our experiments are conducted using the English part (RCV1) of the new Reuters corpus [11], which consists of mostly short news stories for use in research and development of natural language processing, information retrieval, and machine learning systems. RCV1, is significantly larger than Reuters-21578 collection that has been heavily used in the text classification community.

RCV1 is organized in four hierarchical groups: CCAT (Corporate/Industrial), ECAT (Economics), GCAT (Government/Social), and MCAT (Markets). In order to examine CIDA comprehensively, 40 articles are selected randomly for each group.

Since intentionality is a type of subjective information, which is not tagged out in RCV1, 10 human judges have been invited to mark out all the intentionality chains or pairs. The processing results of CIDA are compared against the human-made standard collections of intentionality.

The traditional measures, Precision, Recall and F-

Measure, have been widely applied to analyze and evaluate the performance of CIDA. The experimental results are listed in figure 2.

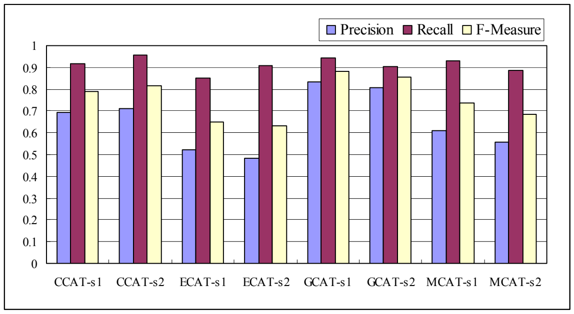

Figure 2. Experimental Results of CIDA

V. CONCLUSIONS

In recent years, research efforts of automatic text analysis have changed focus to more fine-grained analysis tasks with emphasis on subjective information collection. Intentionality detection is a creative and promising research direction in text analysis.

This paper presents an innovative research effort for cognitive intentionality detection, which complements the current research works in automatic recognition and analysis of subjective information.

The experimental results reveal that (1) the performance of CIDA relies on different domains, for example, Precisions on CCAT (Corporate/Industrial) and GCAT (Government/Social) overwhelms that of ECAT (Economics) and MCAT (Markets); (2) CIDA achieves better scores on Recall than Precision, which denotes that CIDA identifies less false positive results.

Our future works include (1) to research the domain dependency patterns of CIDA, and (2) to improve the performance of CIDA, especially on Precision.

ACKNOWLEDGEMENTS

This research work has been jointly funded with the Returned Scholar Research Funding of Chinese Ministry of Education and the Scholar Research Funding of East China University of Science and Technology.

REFERENCES

[1] D. Grinberg, J. Lafferty, and D.Sleator, “A robust parsing algorithm for link grammars”, in: Proceedings of the Fourth International Workshop on Parsing Technologies, Prague, 1995.

[2] Hart, H. L. A., and A. M.Honoré, Causation in the law (2nd ed.), Oxford University Press, 1959/1985.

[3] Heider, F., The psychology of interpersonal relations, Wiley, New York, 1958.

[4] D. J. Hilton, J. McClure, and B. Slugoski, “Counterfactuals, conditionals and causality: A social psychological perspective”, In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), The psychology of counterfactual thinking, Routledge, London, 2005, pp.44-60.

[5] M. Hurst, and K. Nigam, “Retrieving Topical sentiments from online document collections”, in: Document Recognition and Retrieval XI, Proceedings of SPIE, Vol. 5296, San Jose, CA, 2004, pp.27-34.

[6] H. H. Kelley, “Attribution theory in social psychology”, in D. Levine (Ed.), Nebraska symposium on motivation, University of Nebraska Press, Lincoln: Lincoln, 1967.

[7] S. Kim, and E. Hovy, “Determining the sentiment of opinions”, in: Proceedings of the Conference on Computational Linguistics (COLING-04), Geneva, Switzerland, 2004, pp.1367-1373.

[8] S. Kim, and E. Hovy, “Identifying and analyzing judgment opinions”, in: Proceedings of the Human Language Technology/North American Association of Computational Linguistics Conference (HLT-NAACL 2006), New York, NY, 2006.

[9] Korhonen, and T. Briscoe, “Extended lexical-semantic classification of English verbs”, in: Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics, Boston, MA, 2004, pp.38-45.

[10] Levin, B., English Verb Classes and Alternations, The University of Chicago Press, Chicago, 1993.

[11] D. D. Lewis, Y.Yang, T.Rose, and F. Li, “RCV1: A New Benchmark Collection for Text Categorization Research”, Journal of Machine Learning Research, 5, 2004, pp.361-397.

[12] W. Lin, T. Wilson, J. Wiebe, and A. Hauptmann, “Which side are you on? Identifying perspectives at the document and sentence levels”, in: Proceedings of CoNLL-2006, New York, USA, 2006, pp.109-116.

[13] P. Merlo, S. Stevenson, “Automatic verb classification based on statistical distribution of argument structure”, Computational Linguistics, 27 (3), 2001, pp.373-408.

[14] D. T. Miller, and S. Gunasegaram, “Temporal order and the perceived mutability of events: Implications for blame assignment”, Journal of Personality and Social Psychology, 59, 1990, pp.1111-1118.

[15] [19] A. Peña, H. Sossa, and A. Gutiérrez, “Causal knowledge and reasoning by cognitive maps: Pursuing a holistic approach”, Expert Systems with Applications, 35(1-2), 2008, pp.2-18.

[16] V. Rubin, L., Kando, L. Noriko, D. Elizabeth, “Certainty identification in texts: categorization model and manual tagging results”, in: J. G. Shanahan et al., (Eds.) Computing Attitude and Affect in Text, Springer, Dordrecht, The Netherlands. 2005.

[17] Rudanko, J., Complementation and Case Grammar: A Syntactic and Semantic Study of Selected Patterns of Complementation in Present-day English, State University of New York Press, Albany, NY, 1989.

[18] Rudanko, J., Prepositions and Complement Clauses: A Syntactic and Semantic Study of Verbs Governing Prepositions and Complement Clauses in Present-day English, State University of New York Press, Albany, NY, 1996.

[19] F. Scheppers, “P(ragmatic)-trees: Coherence, intentionality and cognitive content in discourse and non-verbal action”, Journal of Pragmatics, Elsevier, 35, 2003, pp.665-694.

[20] S. Schulte im Walde, and C. Brew, “Inducing german semantic verb classes from purely syntactic subcategorisation information”, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, 2002, pp.223-230.

[21] Shaver, K. G., The attribution of blame: Causality, responsibility, and blameworthiness, Springer-Verlag, New York, 1985.

[22] S. A. Sloman, and D. A. Lagnado, “Causal invariance in reasoning and learning”, in B. Ross (Ed.) The psychology of learning and motivation, Elsevier Science, San Diego, 2004, Vol. 44, pp.278-325.

[23] P. Turney, and M. Littman, “Measuring praise and criticism: inference of semantic orientation from association”, ACM Transactions on Information Systems (TOIS), 2003, 21(4), pp.315-346.

[24] J. Wiebe, T. Wilson, R. Bruce, et al., “Learning subjective language”, Computational Linguistics, 2004, 30(3), pp.277-308.

[25] R. A. Zwaan, and G. A Radvansky, “Situation Models in Language Comprehension and Memory”, Psychological Bulletin, Vol.123, No.2, Cambridge University Press, Cambridge, 1998, pp.162-185.