

Analyzing Human Intentions in Natural Language Text

Mark Kröll

Graz University of Technology
Inffeldgasse 21a
8010 Graz, AUSTRIA
mkroell@tugraz.at

Markus Strohmaier

Graz University of Technology and Know-Center
Inffeldgasse 21a
8010 Graz, AUSTRIA
markus.strohmaier@tugraz.at

ABSTRACT

In this paper, we introduce the idea of *Intent Analysis*, which is to create a profile of the goals and intentions present in textual content. Intent Analysis, similar to Sentiment Analysis, represents a type of document classification that differs from traditional topic categorization by focusing on classification by intent. We investigate the extent to which the automatic analysis of human intentions in text is feasible and report our preliminary results, and discuss potential applications. In addition, we present results from a study that focused on evaluating intent profiles generated from transcripts of American presidential candidate speeches in 2008.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]: Human Factors, H.3.1 [Content Analysis and Indexing]: Linguistic Processing.

General Terms

Algorithms, Experimentation, Human Factors.

INTENT ANALYSIS

In this paper, we introduce a method for analyzing the goals and intentions of humans in natural language text extracted from web documents. Traditional topic categorization attempts to classify a document according to its predominant subject matter (what the page is about, e.g. sports or politics). In this work we take a different approach and attempt to classify documents according to the intentions described within them (what goals a page is about, e.g. *Achieve Happiness* or *Maintain Good Health*). Similar to Sentiment Analysis and Opinion Mining, Intent Analysis represents an orthogonal view on topic categorization and aims at creating intent profiles of textual documents. Instead of analyzing text according to the conveyed sentiments or opinions, Intent Analysis attempts to answer *which intentions and human goals* (i.e. *future states of affairs that some agent wants to achieve*) *are referenced in a given document*. This form of analysis deals with a different temporal focus than sentiment analysis, where a *present* (emotional) state is approximated.

Using a sample of web documents we observed that people rarely state their intentions explicitly in text, which makes the task of Intent Analysis an especially challenging

endeavor. As an example, consider the human religious intention to *Achieve Salvation* (taken from [1]). Although this is an activity pursued by many, it is extremely rare to find someone who states their plan on how to accomplish this goal. However, people are quite prolific in writing about the actions and activities they participate in on a daily basis, such as “convert to Christianity”, which indirectly contribute to *Achieve Salvation*.

In this work, we explore the use of such *actions* as a proxy for inferring intentions from textual content. *Intent Analysis* now can be understood as identifying a corresponding intent category for every *action* indicative of intent in a given text. The task of Intent Analysis is to approximate the unknown function $f: S \times C^I \rightarrow \{True, False\}$, where $C^I = \{c^1, c^2 \dots c^n\}$ is a set of predefined intent categories, D is a domain of text documents and each document d_i contains a sequence of sentences $S = \{s_1, s_2 \dots s_{|S|}\}$.

AUTOMATIC APPROACH

In a first step, we search the web to build up a knowledge base that maps indicative actions to intent categories based on an existing goal taxonomy. In a second step, we scan a target text for indicative actions and look them up in the knowledge base to produce an intent profile (*ip*).

Enriching a Taxonomy of Human Intentions: We employed the social-psychological theoretical framework [1] that organizes high-level intentions of people into 135 categories. In order to further describe these categories, we attempted to find descriptive phrases by conducting brainstorming sessions. To give an example: Descriptive phrases for the category “Achieve Salvation” included “to reach spiritual enlightenment” or “to get into heaven”. Dr. S. Read, a co-authors of [1], evaluated these mappings to help us better understand the intent category distinctions.

Constructing the Knowledge Base: We identified actions associated with each of the 135 categories by searching for sentences on the web that contained both (i) one of the descriptive phrases for the category, and (ii) an action-based causal relation. To build the knowledge base, we constructed a series of query strings by concatenating each descriptive phrase with the following two causal relation phrases: “in order to” and “for the purpose of”. Then, exact phrase searches were issued to the web using the Yahoo! BOSS API. Result page sentences that contained the phrase were stored in an Apache Lucene index. Table 1 shows sample phrase queries and retrieved sentences with the respective, indicative action highlighted.

Table 1: Exemplary query strings for the category "Looking Young" and retrieved sentences containing actions.

Query string causal relation + descr. Phrase	Retrieved Sentences (Yahoo) indicative actions
"in order to look young"	In order to look young and beautiful, you need <i>to take care of your skin.</i>
"in order to avoid wrinkles"	You need <i>to moisturize inside and out,</i> in order to avoid wrinkles.

The minimum number of sentences per category was 12 ("Firm Values"), the maximum number was 7,323 ("Helping Others") and the average number was 1,249.

Intent Analysis of Documents: To automatically generate an intent profile, we first segment a given document into a set of sentences. Then, each sentence is issued as a query to the knowledge base. Using the default Lucene similarity measure, we identified the most similar entry in our knowledge base. The associated intent category is then assigned. The intent profile (*ip*) of the document is subsequently computed as a normalized frequency of category assignments.

APPLICATION AND EVALUATION

We applied our prototype to a selected set of textual documents, i.e., 44 transcripts of political speeches from the two leading American presidential candidates in 2008, John McCain and Barack Obama. The speeches can be expected to contain a broad variety of political intentions, which make them particularly suitable for our purpose. Every sentence was treated as a query for the knowledge base. The resulting intent annotations can be represented as a 135 dimensional vector per document as illustrated in Figure 1, which sort of profiles Obama's speeches given in April and June 2008. Categories such as "Charity" and "Helping Others" are prominent over the whole period. Peaks such as the category "Pursuing Ideals" can be easily detected. In Figure 2, two averaged intent profiles for Barack Obama's and John McCain's speeches are contrasted. At first glance, similarities and differences between the two candidates can be recognized, providing some sort of intentional summary of the speech contents. Both candidates conveyed messages to their audience that were often assigned to high-level intent categories such as "Leader", "Helping Others" and "Charity". Figure 2 also reveals intent categories that are more strongly connected with one of the two candidates. Categories such as "Self Esteem" were stronger linked to McCain's speeches whereas Obama's speeches seemed to emphasize other categories such as "Pursuing Ideals" and "Aspirations". We also conducted two human subject studies to evaluate (i) the quality of the knowledge base and (ii) the difficulty of the automated intent analysis task itself. In the first study, a random sample of 674 knowledge base entries was evaluated according to the plausibility of the action/category mapping. 57% of the sample sentences contained indicative actions. In the second study, we compared the top 25 intent categories that were assigned by the human annotators to the categories assigned by our automated approach. This resulted in an agreement of 12

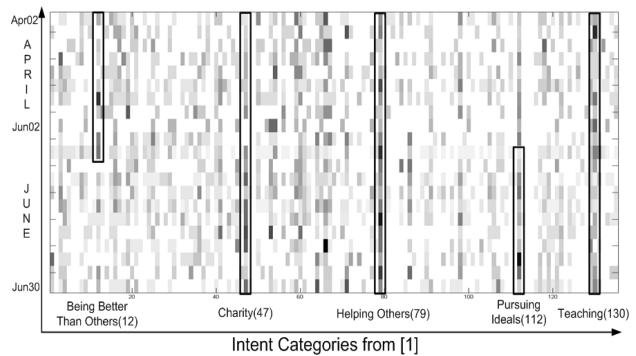


Figure 1: Intent Profiles of 21 speeches given by Barack Obama in April and June 2008.

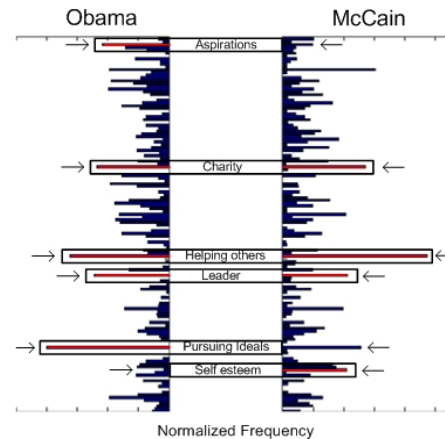


Figure 2: Intent Profiles of Obama and McCain, generated from and averaged over 44 speeches (April and June 2008).

out of 25 intent categories (48%) based on averaged intent annotations from Obama's speeches.

Intent Analysis represents a novel addition to the repertoire of textual data analysis techniques. We presented a prototypical implementation of an automated method for Intent Analysis that generates intent profiles of natural language text documents. Our results indicate the potentials of Intent Analysis as a quick, visual evaluation of natural language text from an intentional perspective.

ACKNOWLEDGMENTS

Thanks to E. Tavano, J. Liegl, H. Stern and D. Zibold for conducting the annotation tasks and to S. Read for assistance regarding the human goal taxonomy. Special thanks to A.S. Gordon and R. Swanson at the ICT, University of Southern California, for their assistance in completing this research. This work is funded by the FWF Grant P20269 *TransAgere* and the Know-Center Graz.

REFERENCE

- [1] Chulef, A. S., Read, S. J. and Walsh, D. A. A hierarchical taxonomy of human goals. In *Motivation and Emotion*, 25 (3), 191--232, 2001.