

TrendMiner: Large-Scale Analysis of Political Attitudes in Public Facebook Messages

Márton Miháltz, Tamás Váradi

Research Institute for Linguistics, Hungarian Academy of Sciences
Benczúr utca 33, H-1068, Budapest, Hungary

Abstract—This paper presents the methods and results of a project that collects and analyses public comments written in response to political posts on Facebook using natural language processing and social psychological methods in order to explore emotional attitudes and social behavior.

I. INTRODUCTION

The tools developed in the FP7-financed *TrendMiner* project¹ focus on the large-scale collection, analysis and social psychological evaluation of public comments written by Hungarian Facebook users in response to posts published on the pages of Hungarian politicians and political organizations. The collected text messages are analyzed using natural language processing (NLP) methods adapted to identify psychological and social phenomena in language. Our research aims to support the investigations of emotional and social attitudes, with possible applications to model political behavior to complement traditional political opinion polls and surveys.

With respect to Cognitive Infocommunications ([1]), our work hopes to extend the side of Affective Computing that deals with the artificial understanding of intra-cognitive communication in the textual modality.

II. PROJECT OUTLINE

We collected about 140,000 public posts and 2 million associated public comments with the help of Facebook Graph API during the 12 months of the project from over 1,300 Facebook pages which belong to political parties registered in Hungary, their member organizations, elected and nominated representatives (with respect to the general elections and the European Parliament elections in Hungary in 2014).

The collected messages are processed using a pipeline of standard NLP tools: segmenting into sentences and tokens (using the *huntoken* tool), morphological analysis and disambiguation to identify word lemmas and inflections (*hunmorph*, *hunpos* and our own tools). These tools had to be adapted to the special challenges presented by social media language (slangs, neologisms, non-standard spelling, emoticons etc.) This is followed by the identification of relevant named entities (the names of persons and organizations and their party affiliations) and then finally content analysis using finite state automata developed with the help of the *NooJ* tool.

¹ <http://www.trendminer-project.eu/>

Our first special content analysis stage is *sentiment analysis*, the annotation of emotional polarity (neutral, positive or negative) expressed by the text, which is a widely used measure in social media analysis. This is achieved by custom lexicons and simple rules to account for the modifying effects of context (e.g. negation).

Two further content analysis modules relying on lexicons and rules identify social value judgements. *Agency* describes an individual in terms of the efficiency of his or her individual goal-oriented behavior (motivation, competence and control). On the other hand, *communion* describes the moral and emotional aspects of group relations (cooperation, social benefit, honesty, self-sacrifice, affection, friendship, respect, love etc.) Both indicators have positive and negative ranges.

The content analysis indicator for *optimism/pessimism* relies on the identification of verb tenses (past, present, future). This is based on the assumption that optimistic people tend to talk more about the future (open possibilities) and less about the past (unchangeable events), while pessimists talk more about the past and less about the present. Finally, we use another indicator to measure the amount of *individualism/collectivism* expressed in comment messages, reflecting the importance of the category of the self when talking about world events. This is measured by the presence of personal pronouns and verbal and noun personal inflections in the sentences.

We evaluated the custom tools used for the identification of these social psychological phenomena against human judgements ([2]). We also used this system to correlate the indicators measured on Facebook comments against traditional political opinion polls, and showed correlation in several cases ([3]). All the tools and the fully annotated data used in the project are freely available².

References

- [1] P. Baranyi, A. Csapo, "Definition and synergies of cognitive infocommunications" *Acta Polytechnica Hungarica* vol 9 pp 67-83, 2012.
- [2] M. Miháltz, T. Váradi, "TrendMiner: politikai témájú Facebook-üzenetek feldolgozása és szociálpszichológiai elemzése," XI. Magyar Számítógépes Nyelvészeti Konferencia, pp. 195-197, 2015.
- [3] T. Pólya, I. Csertő, É. Fülöp, P. Kóvágó, M. Miháltz, T. Váradi, "A véleményváltozás azonosítása politikai témájú közösségi médiában megjelenő szövegekben," XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015), pp. 198-209, 2015.

² <http://corpus.nyttud.hu/trendminer>