

Monitoring Social Attitudes Using Rectitude Gains

Aleksander Wawer

Institute of Computer Science, Polish Academy of Science,
ul. J.K. Ordonia 21, 01-237 Warszawa, Poland
axw@ipipan.waw.pl

Abstract. The article describes a prototype system aimed at monitoring attitudes toward any social group. The approach involves web mining and content analysis based on Rectitude Gain category from the Laswell dictionary of political values, extended into shallow predicate rules. The system requires no lexical sentiment resources and no training corpora. It has been designed, implemented and tested in Polish.

Keywords: content analysis, partial parsing, extremism detection.

1 Introduction: Extreme Movements and Security

Knowledge about attitudes that potentially give rise to radical extremist activity becomes all the more important. In the United States, for well-known reasons, the focus is on movements grounded in radical religious ideology. In Eastern and Central Europe, long-term threats to public safety seem to be more related to progressing nationalist radicalization¹. This demands active monitoring of emerging nationalist ideas and their frequency in user generated content.

The goal of such monitoring is not only to identify movements with potential for radicalization and danger, which enables policy makers to undertake preventive steps, but also to measure the impact of their messages on internet audience. Arguably, the ideal moment when such formations should be tracked is their early age, when a group or movement has not yet reached a violent stage.

Identifying extremist ideas at their early stage is not an easy task. The main reason of this difficulty is that extreme and thus potentially dangerous ideas span over a wide spectrum of social movements and groups, emerging from various religious, political and social backgrounds². Because it is not possible to identify

¹ Perhaps amplified by economic downturn, nationalist ideology is gaining momentum as it became evident in the last Europarlament Elections. Examples of extreme nationalist movements include Hungarian Guard, Slovakia's SNS, German NPD and many more.

² Domestic Extremism Lexicon (<http://www.fas.org/irp/eprint/lexicon.pdf>) compiled by the Strategic Analysis Group and the Extremism and Radicalization Branch, U.S. Department of Homeland Security, includes (among many others) anarchist, animal rights, antiabortion, antitechnology, environmental, Jewish and neo-Nazi movements.

one ideology or belief system which defines extreme formations, early detection of extreme views should go beyond keyword recognition. Such diversity indicates also that corpora-trained systems may be ineffective at recognizing emerging, new extreme groups.

2 Previous Work

Examples of related works include research on text categorization using machine learning, applied to automated hate speech detection as in [4]. Authors experiment with bag of words and POS feature vectors to train SVM classifiers on racist text corpora. Another relevant research is [2], which mixes social network analysis with sentiment analysis (scores calculated using SentiWordNet-based algorithm) to detect online jihadist radicalisation on YouTube. Both mentioned works rely either on hate speech corpora or existing sentiment lexicons.

3 Content Analysis and Rectitude

3.1 Laswell Rectitude Gains

In the Laswell dictionary [5] of political values³, one category is of specific interest: RECTITUDE. Among the five RECTITUDE subcategories, two are substantive (ETHICS and RELIGIOUS) and three refer to transitions (GAINS, LOSSES, ENDS). Substantive categories address the problem of rectitude sources, either established as *the social order and its demands as a justifying ground* (ETHICS) or *transcendental, mystical or supernatural grounds* (RELIGIOUS)⁴.

The subcategory around which the proposed system has been designed is RECTITUDE GAIN, a very abstract category which has to be carefully interpreted. Specifically, RECTITUDE GAIN must not be confused with perhaps too obvious situations like court sentences or apologies. Instead, RECTITUDE GAIN is better explained as states or situations when unethical, immoral or bad conduct **is revealed** (*blame, breach, double cross, lying etc.*) and states or situations referring to **actions aimed at rectitude growth** (*root out, condone, forgive, justify, reparation, restitution, vindicate*).

Detection of extremism, as proposed in this article, is not based on direct attacks or expressions of hate, as in hate speech, but instead turns toward somehow less offensive statements — if offensive at all, expressing (the need of) rectitude gain related to a social group, like nation, race, ethnical, sexual or religious minority. The need of rectitude growth is the driver or rationale of actions against

³ Integrated with the General Inquirer Harvard IV dictionary [7]. Keeping the convention, references to rectitude category names in the General Inquirer are typed upper-case.

⁴ It is remarkable that Laswell dictionary does not mention nation among substantive RECTITUDE subcategories. It seems that several 20th century extreme European social movements, including the cruelest ones, were grounded on nationalism rather than ethics or religion.

this social group, explanation of the need for acting as the group does not conform to rules of moral conduct, historical sense of justice, transcendental or spiritual rules, or any other rectitude grounding substance. Perhaps rectitude gains, as used in the sense of this paper, can be compared to a hypothetical zero step on the Allport scale [1].

3.2 Rectitude Gains in Polish

Translating senses in RECTITUDE GAIN category into Polish has been done manually with special attention to ensure maximum quality. Out of 30 entries in the English Laswell dictionary we selected 6 ones, presented with their Polish translation and sentiment in Table 1, which are likely to appear in the context of social group names.

Table 1. Selected Rectitude Gain entries and translation

General Inq.	POS	Meaning	Polish
BLAME#1	noun	Reproof, culpability	są winni (-)
BLAME#2	verb	To hold responsible for failure, to censure	obwiniać (-)
CLEAR#6	verb	To absolve from blame	rozgrzeszać (+)
CONDONE	verb	To forgive, to excuse	przebacząć (+)
EVEN#3	idiom verb	Get even; to have one's revenge	wyrównać
ROOT#4	idiom verb	To root out, to extirpate	wyplenić (-)

The RECTITUDE GAIN category contains entries marked as negative, as well as positive on the Osgood evaluative dimension. Taking this into consideration, we distinguished between good (+) and bad (-) rectitude entries in Table 1 and the remaining part of the paper ⁵.

4 Web Mining

Mining the social web and meanings it generates for monitoring extreme ideas requires as wide access to the influx of user generated content as possible. In Poland, where this research has been carried out, Google has dominant market position in the search engine market, and because of this fact alone it is more than likely that no other engine indexes as much web content as Google. Thus, we have decided to collect the data by regularly submitting queries to Google and analyzing obtained results. The queries are Polish translations of Laswell entries as in Table 1, wrapped in double quotes for exact phrase matching.

⁵ "good/bad" was chosen over "positive/negative" to avoid confusion with Osgood's categories.

5 Spejd Rules

The corpus of 4747 web pages has been analyzed using Spejd⁶, originally a tool for partial parsing and rule-based morphosyntactic disambiguation, used in the context of this work to find names of social groups — subjects of Laswell rectitude gains. 36 Spejd rules were manually constructed by extending translated terms into subject-predicate structure, to capture subject appearances. The recall has yet to be verified, but due to Google queries, they are likely to be exhaustive in covering the most common syntactical structures involved.

Full description of Spejd formalism is out of the scope of this paper. As an example, we describe 6 rules, the translation of BLAME#1 extended into subject-predicate structure. Rules take into account inverted syntax and possible insertions of modifiers. Extracted subjects are base forms (`base~"word"`) of plural noun forms (`pos~"subst" && number~"pl"`), sometimes restricted to specific grammatical case (for example, `case~"nom"` means forms with possible nominative interpretation), referred to by the last two `group` arguments in `Eval` rule part. Originally, this notation has been introduced to mark semantic and syntactic heads of a group, in our system we use it to point to appropriate token number. The first three rules (handling negation) are as follows:

```

Rule    "X, which are not to be blamed"
Match:  [pos~"subst" && case~"nom" && number~"pl"] ns? [pos~"interp"]?
        [base~"kto"]? []? [orth~"nie"] [base~"być"]
        [base~"winny" && negation!~"neg"];
Eval:   group(kt_nie_sa_winni_1, 1, 1);

Rule    "are not to be blamed X (inverted)"
Match:  [orth~"nie"] [base~"być"]? []? [base~"winny" && negation!~"neg"]
        [pos~"subst" && case~"nom" && number~"pl"];
Eval:   group(nie_sa_winni_1, 5, 5);

Rule    "X are not to be blamed (negated adjective form)"
Match:  [pos~"subst" && case~"nom" && number~"pl"] [base~"być"]?
        [base~"winny" && negation~"neg"];
Eval:   group(sa_niewinni_1, 1, 1);

```

Another three rules for BLAME#1 are analogous to the three presented above, but do not involve negation:

```

Rule    "X, which are to be blamed"
Match:  [pos~"subst" && case~"nom" && number~"pl"] ns? [base~"kto"]? []?
        [base~"być"] [pos~"adv"]? [base~"winny" && negation!~"neg"];

```

⁶ <http://nlp.ipipan.waw.pl/Spejd/> [6]

```

Eval:  group(kt_sa_winni_0, 1, 1);

Rule   "are to be blamed X (inverted)"
Match: [base~"być"? []? [base~"winny"]
        [pos~"subst" && case~"nom" && number~"p1"];
Eval:  group(sa_winni_0, 4, 4);

Rule   "X are to be blamed"
Match: [pos~"subst" && case~"nom" && number~"p1"] []?
        [base~"być"? [base~"winny"];
Eval:  group(sa_winni_0, 1, 1);

```

Because Spejd is a cascade of grammars, more specific rules like those with negation have to be run prior to more general ones. Frequency counting inverts rectitude polarity when extracted by negated rules (from bad to good in the case of presented BLAME#1 rules).

6 Results

As anticipated, extraction rules yielded different types of subjects with regard to which rectitude raises, not only names of nationalities, minorities and social groups. For reporting purposes, the list has been limited to manually selected names, appearing in the initial corpus. Yet, lists of subjects have to be reviewed periodically for social group names. While this approach may seem labour-intensive, it ensures that no social group name is omitted. Unfortunately, at the moment no exhaustive list of nations and social groups in Polish appears to be publicly available. Automated methods have yet to be implemented.

Good and bad rectitude frequencies toward social groups and nations extracted from the gathered data have been presented in Table 2 below. While the interpretation of these results is out of the scope of this paper, they seem to be rather confirming the general intuition on how Poles feel about particular

Table 2. Good and bad rectitude frequencies toward nations and social groups

	Good rect.	Bad rect.
Jews	0	10
Israeli	0	8
Germans	6	9
Russians	8	9
Poles	0	9
Iraqi	0	2
Afghans	0	4
homosexuals	0	4
masons	0	1

nations and groups. Low frequencies are the consequence of not using nation and group names in web queries. As it has been argued above, querying using explicit names is less likely to produce reliable results.

7 Conclusions and Future Work

We presented a prototype system aimed at detecting attitudes toward any social group and tested it on a sample of web pages. Initial results indicate that rectitude gains extended into shallow rules can be a promising way of mining sentiment toward nations and minorities. The proposed method has advantages over existing approaches. Firstly, it does not require sentiment lexical resources like General Inquirer's evaluative categories or SentiWordNet [3] and does not rely on training corpora. Secondly, it does not require apriori knowledge of social group names and is easy to implement in any language. In the final version, queries will be issued monthly and limited to results indexed within the last month to enable continuous monitoring. Additional work has to be conducted to further evaluate the rules with regard to complex sentence structures.

References

1. Allport, G.: *The Nature of Prejudice*. Addison-Wesley, Reading (1954)
2. Bermingham, A., Conway, M., McInerney, L., O'Hare, N., Smeaton, A.F.: Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. In: *International Conference on Advances in Social Network Analysis and Mining*, pp. 231–236 (2009)
3. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: *Proceedings of LREC* (2006)
4. Greevy, E., Smeaton, A.: Text categorisation of racist texts using a support vector machine. In: *SIGIR 2004: Proceedings of the 27th Annual International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2004: Proceedings of the 27th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 468–469 (2004)
5. Namenwirth, Z.J., Weber, R.P.: *Dynamics of Culture*. Allen & Unwin, Inc. (1987)
6. Przepiórkowski, A., Buczyński, A.: spade: Shallow parsing and disambiguation engine. In: *Proceedings of the 3rd Language & Technology Conference, Poznań* (2007)
7. Stone, P.J., Dunphy, D.C., Ogilvie, D.M., Smith, M.S.: *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge (1966)