Society as a Life Teacher – Automatic Recognition of Instincts Underneath Human Actions by Using Blog Corpus

Rafal Rzepka and Kenji Araki

Hokkaido University, Sapporo, Kita-ku, Kita 14 Nishi 9, 060-0814, Japan
{kabura,araki}@media.eng.hokudai.ac.jp
http://arakilab.media.eng.hokudai.ac.jp

Abstract. In this paper we introduce a method for generating a set of possible reasons of an action needed by an AI program for reasoning about human behavior. We achieve this goal by using web-mining and lexicons of keywords reflecting 14 instincts categories developed by psychologist William McDougall. We describe our system, the experiment and analyze its results of 78% of correct retrievals. The paper is also meant to be a message to social scientists who might be interested in testing their theories on constantly growing group of Internet users.

Keywords: causal knowledge retrieval, human instincts, text-minig.

1 Introduction

There is more than one possible reason for our behaviors and their causes are most often multidimensional even if we tend to simplify them. Although, when asked, we are able to imagine a whole range of possible reasons for a human's action or a state experienced by us or a third person. Experiences we gather from our earliest days allow us to explain things to our children and even ourselves when we have difficulties with immediate understanding. Machines do not have this capability and their attempts to reason about human acts tempt to lack necessary depth. Usually researchers working on causal knowledge retrieval concentrate on dry and easily verifiable facts as "people dry the laundry because it is sunny weather" [1], but when it comes to emotions and deeper analysis of usual or unusual acts, a computer needs a wider variety of possible causes to perform context processing. On the other hand, when, for example, a person makes a statement that *she is drinking alcohol* to a dialog system, usually there is not enough contextual data and the program needs to "imagine" why usually people drink and if a reason is confirmed (e.g. "to celebrate"), the system can more easily assume an emotional state of the person ("happy"). Retrieving affective consequences is a widely popular topic of the sentiment analysis field, but the set of instinctual causes and emotive effects is rarely a subject of knowledge acquisition. We think that our methods may be interesting not only for the AI researchers but also people from the humanities who would like to test their theories (or theories of others – like in our case) on thousands of people who share

A. Jatowt et al. (Eds.): SocInfo 2013, LNCS 8238, pp. 370–376, 2013.

their thoughts online. In this paper we introduce our attempts to automatically categorize acts and states according to McDougall's theory of instincts.

1.1 McDougall's Categorization of Instincts

In our research on machine ethics we assume that human beings are equipped with the same instincts which build our morality. Haidt and Joseph [2] have performed a survey on common core of moral values, concerns, and issues across cultures and found three pairs: suffering/compassion, reciprocity/fairness, and hierarchy/respect. First we tried to build our system based on these pairs but we needed more sophisticated categorization and discovered works of William McDougall (1871-1938), who has been largely forgotten – until recently, with genetics and evolutionary psychology on the rise. The psychologist saw instincts as having three components. One is *perception* – human beings pay attention to stimuli relevant to our instinctual purposes, the second is *behavior* – human beings perform actions that satisfy our instinctual purposes, and the third is *emotion* – instincts have associated negative and positive emotions [3]. What is different from classic stimulus-response based behaviorism in case of McDougall's approach is purposiveness of instincts meaning that they are goal-directed. For that reason we found his theory useful for our purpose of creating a web-mining module which can "imagine" why somebody did something and what was the outcome of every possible motivation of the given act. In McDougall's opinion, all three above mentioned components work simultaneously and in concert with other instincts and we agree with him on this point – an analytical algorithm should be able to blend different instincts, not only the dominant one.

1.2 Developing a Lexicon for Blog Queries

McDougall came up with 14 types of instincts and their accompanying emotions. According to his list (in English) we created a lexicon of Japanese expressions (as we currently perform experiments limiting search span to only one culture but plan to extend the system to work with English, Chinese and Polish). We utilized phrases from Nakamura's dictionary of emotive expressions [4] and added our own query words trying to fit the explanations of instinct categories left by McDougall. The Categories are enlisted below.

- **Escape:** words associated with fear were collected, for example *scary*, *scared*, *fearful*, *terrifying*, *run away*, *horrifying* or *hair-raising* (21 phrases in total). The number of phrases in each category is not equal but we have already shown that it does not influence recall [5]
- **Combat:** words associated with anger, for example *get angry*, *furious*, *raging*, *enraged*, *outraged*, *pissed off* and *lose temper* (7 phrases in total).
- **Repulsion:** "disgust" associations (e.g. *disgusting*, *disgusted*, *disgustful*, *nauseating*, *sickening*, *can't believe* or *make one puke*) (18 phrases in total).

- **Parental** (protective)**:** words associated with love and tenderness, for example *lovely*, *attachment*, *kind*, *friendly*, *nice*, *pleasant* or *dear* (12 phrases in total).
- **Appeal** (for help)**:** words for matching distress and feeling of helplessness were added here, for example *weak*, *fragile*, *depressed*, *depressing*, *hopeless*, *powerless* or *couldn't do anything* (13 phrases in total).
- **Mating:** lust and attractiveness related words, for instance *beautiful*, *gorgeous woman*, *sexy*, *pretty*, *handsome*, *want to make out with* or *I'd marry* (10 phrases in total).
- **Curiosity:** words bearing meaning of feeling of mystery, of strangeness and of the unknown, e.g. *interesting*, *surprising*, *worth checking*, *rare*, *peculiar*, *strange* or *want to know* (8 phrases in total).
- **Submission:** words for feeling of subjection, inferiority, devotion, humility or negative self-feeling, for instance *ashamed*, *embarrassed*, *guilty*, *inferior*, *bashful*, *shy* or *blush* (10 phrases in total).
- **Assertion:** words for feeling of elation, superiority, masterfulness, pride and positive self-feeling, for example *happy*, *glad*, *easygoing*, *feeling good*, *good mood*, *satisfied* or *grin* (17 phrases in total).
- **Gregariousness:** words expressing feeling of loneliness, isolation or nostalgia – *lonely*, *crying*, *nostalgic*, *lonesome*, *tears*, *hurt*, *grieve*, etc. (16 phrases in total).
- **Food-seeking:** expressions for appetite or craving as *tasty*, *looking tasty*, *want to eat* or *wish to eat* (6 phrases in total).
- **Hoarding:** words expressing feeling of ownership and greed – *want to have*, *want to own*, *want to get*, *want to collect*, *don't want to lose*, etc. (7 phrases in total).
- **Construction:** expressions bearing meaning of feeling of creativeness, making, or productivity, for instance *would like to make*, *want to create*, *felt good to make*, *wanted to give birth*, *want to produce*, etc. (20 phrases in total).
- **Laughter:** words for amusement, carelessness, relaxation, for example *funny*, *laughed*, *feel relief*, *feel peaceful*, *peaceful* or *peace of mind* (19 phrases in total).

Above lexicon divided into 14 subsets was then used for matching process described in the next section.

2 Retrieval Process

Our system takes an input phrase consisting of a noun, a Japanese particle[1] and a verb. Then it automatically creates 9 queries by modifying the verb into conditional and continuative forms. These queries (as exact matches) are then used for retrieving sentences from 5 billion sentences blog corpus [7]. In order to avoid noisy inputs we use a hand-made facemarks database to cut strings which

[1] A suffix in Japanese grammar that immediately follow the modified noun and indicate if it is an object, topic, place, etc.

had no periods; we also set length limits to avoid too short and too long (or most probably wrongly divided into sentences) strings. The search is made by Apache Solr[2] and the retrieved sentences are automatically cleaned by removing ornate characters (notes, hearts, etc.) and passed to a semantic analysis tool ASA[3] for creating chunks that are more meaningful and are less prone to errors than usual N-grams. Finally, the left side chunk precedent to the query is matched against all the phrases from every subset of instincts lexicon. The hits are counted and a ranking is made. Some examples of input and output are shown in Table 1.

Table 1. Example of retrievals

Input	Top Instinct Category	Example Precedents
Kill someone	Repulsion (15)	disliked the society, felt grudge toward people, his/her state got worse...
Make a phone call	Escape (8)	mother was worried, feared about me, mom was shaking...
Help somebody	Appeal (15)	this year was in trouble, was on weak position, is in real trouble...
Tell a lie	Parental (2)	I don't like alcohol, I loved her/him
Go to restaurant	Food-seeking (8)	tasty at home, tastes nice at the sea, yummy seafood...

3 Experiment and Evaluation

In this section we introduce a preliminary test we performed to investigate how efficient are the algorithm and the used lexicon. By efficiency here we mean the number of natural instinct associations against unnatural ones. For example *sending kinds to school* is obvious in "Parental" category but does not fit "Mating" category. More than one natural categories are possible and many exceptions may happen (like "Repulsion" in case of very bad school environment), however we currently concentrate on commonsensical side of categorization. Having stated so, we are working toward a broader goal of context processing and this paper introduces an important part of deeper contextual analysis to enable machines to handle particular cases without falling in dangerous generalizations.

3.1 Experimental Setup

As an input we utilized 127 action phrases as "(to) call an ambulance" or "to kill a hero", etc. This list was an enhanced set we used in [8] for recognizing ethically

[2] Sorl (`http://lucene.apache.org/solr/`) is as fast as commercial engines and has no number of queries limitations, but obviously the scale of the data is very different. However, we have already proven that the higher precision of deeper search gives equal f-score to broader search in spite of the low recall [8].

[3] `http://cl.it.okayama-u.ac.jp/study/project/asa/about_asa.html`

problematic acts – this time we added more everyday life actions like "writing a book" or states like "someone in pain"[4] and removed similar entries, for example those needed for comparing *reasons for* and *reactions to* killing various kinds of animals, which in many cases have rather low hit rate in blogs.

3.2 Evaluation of the Categorization Task

First we took only the instincts with the highest score (one top instinct for an input) and the system achieved precision of 75.0% with the recall of 53.54%. After a closer look we have noticed that state describing inputs probably should not be evaluated as ones having instinctual motivations of actions[5] and we excluded them (11 phrases) together with 3 phrases written in erroneous Japanese. This time we also checked all the retrievals (553 hits), not only the top-scoring categories. Both preliminary evaluations were made by the first author and this more thorough judgement process showed that 77.78% category assignments were correct (almost 3 percentage points increase) but the recall dropped to 49.61% as we excluded state describing inputs.

3.3 Error Analysis and Possible Solutions

Rather low recall was obviously due to the restrictions we set on the retrieved data in order to avoid noisy strings for analysis. Therefore only 62,329 sentences were extracted for matching, which gave average of only 490 sentences for one input. We need to test the input against less or non-restricted data, but the low speed of semantic analysis performed by ASA using dependency parsing makes such experiments extremely time consuming.

There are obviously problematic expressions in the lexicon. For instance *liking* in "Parental" instincts category or *crying* in "Gregariousness" are too wide and can represent more than one instinct. *Worry* in "Escape" caused categorizing "make a phone call" as a fear based motivation, while this action should rather be under "Parental" instinct category. We are currently improving the lexicons by manual proofreading and automatic methods.

One Chinese character matching words should be removed as they appear in phrases with different connotations. For example there is an ideogram meaning "dislike" in a word "mood" (*go-kigen*) or "fear" character in an apology expression *kyōshuku*.

Certainly there is a need for analyzing also the right side (following chunk) of a query, because it would help disambiguate situations when such chunk can influence the meaning of the precedent (e.g. by being a negation).

[4] Many of phrases were taken from the logs of our online demo system `http://demo.media.eng.hokudai.ac.jp/?ethical`

[5] Input states were most often very broad (*being alive*, *laughing*, *enjoying*, etc.) and they were closer to being instinctual reasons for actions, not actions themselves. In future we plan to experiment with more specific state inputs.

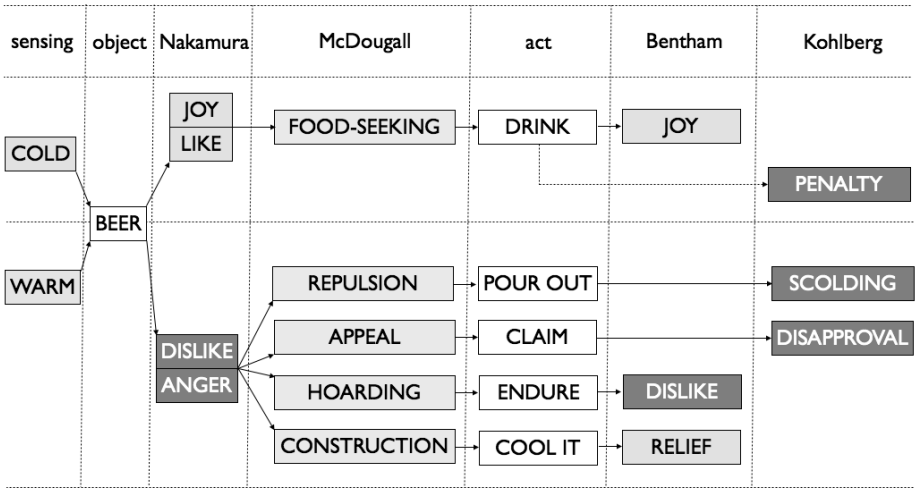

Fig. 1. Example of an act analysis using different classical theories. When an action details are input, the Web searching techniques using lexicons for each theory allow automatic retrieval of knowledge and predicting possible causes and effects. Multiple layers allow to acquire a deeper and ensure a basic understanding when one or more module fails to extract knowledge. Rules acquired in this process let the system assume what are, for example, possible outcomes when only feelings are given, or plan its own action to achieve the highest utilitarian *Pleasure* in Bentham's Felific Calculus.

4 Conclusions and Future Work

In this paper we introduced our method for automatic retrieval of possible instinctual reasons for human behaviors by using McDougall's list of instincts and web-mining and natural language processing techniques. Although our goal is to equip a machine with a capability of understanding human beings, with this research we would in addition like to suggest how interesting such approach could be also for social scientists. We have already shown usefulness of classic works of Akira Nakamura (his emotive expression dictionary [4] in sentiment analysis [6]), of Lawrence Kohlberg (his theory of moral stages development [9] in machine ethics field [10]) and currently implement Jeremy Bentham's Felific Calculus [11] into a system allowing deeper predictions of human act consequences, which we plan to use in the fields of collective behavior [12], machine self-understanding [13] and artificial empathy [14] (see Figure 1 to have a glimpse of reasoning using these theories). By this paper and above examples we wish to send social scientists a message saying that the new field of socioinformatics has plenty of opportunities for bringing back classical theories and testing them against vast number of Internet users.

As for the future work, we plan to improve our system according to the findings of above mentioned error analysis and continue developing the context recognition modules. As you can see in Figure 1, drinking a beer is usually joyful but

if the social consequence retrieval algorithm finds "driving" context, the outcome may be of the opposite nature. Another task is to cooperate with social scientists on achieving more polished and wider lexicon in order to decrease the instincts categorization error rate; the numbers of query words must be also balanced according to their occurrences in the whole corpus. Finally, the evaluation experiment must be repeated with native speaker evaluators.

References

1. Inui, T., Inui, K., Matsumoto, Y.: Acquiring causal knowledge from text using the connective marker tame. ACM Trans. Asian Lang. Inf. Process. 4(4), 435–474 (2005)
2. Haidt, J., Joseph, C.: Intuitive ethics: how innately prepared intuitions generate culturally variable virtues, Dædalus, special issue on human nature: 55-66 (2004)
3. McDougall, W.: Outline of Psychology. Methuen & Co. (1923)
4. Nakamura, A.: Kanjo hyogen jiten (Dictionary of Emotive Expressions). Tokyodo Publishing (1993)
5. Ptaszynski, M., Rzepka, R., Araki, K., Momouchi, Y.: Automatically Annotating A Five-Billion-Word Corpus of Japanese Blogs for Sentiment and Affect Analysis. Computer Speech and Language (CSL). Elsevier (2013)
6. Ptaszynski, M., Dybala, P., Mazur, M., Rzepka, R., Araki, K., Momouchi, Y.: Towards Computational Fronesis: Verifying Contextual Appropriateness of Emotions. IJDET 11(2), 16–47 (2013)
7. Ptaszynski, M., Dybala, P., Rzepka, R., Araki, K., Momouchi, Y.: YACIS: A Five-Billion-Word Corpus of Japanese Blogs Fully Annotated with Syntactic and Affective Information. In: Proceedings of The AISB/IACAP World Congress 2012 in Honour of Alan Turing, 2nd Symposium on Linguistic and Cognitive Approaches To Dialog Agents (LaCATODA 2012), pp. 40–49. University of Birmingham, Birmingham (2012)
8. Rzepka, R., Araki, K.: Polarization of consequence expressions for an automatic ethical judgment based on moral stages theory. IPSJ SIG Notes 2012-NL-207(14), 1-4 (2012)
9. Kohlberg, L.: The Philosophy of Moral Development. Harper and Row (1981)
10. Komuda, R., Rzepka, R., Araki, K.: Social Factors in Kohlberg's Theory of Stages of Moral Development the Utility of (Web) Crowd Wisdom for Machine Ethics Research. In: Proceedings of the 5th International Conference on Applied Ethics, p. 14 (2010)
11. Bentham, J.: An Introduction to the Principles and Morals of Legislation. T. Payne, London (1789)
12. Rzepka, R., Araki, K.: Consciousness of Crowds - The Internet As a Knowledge Source of Humans Conscious Behavior and Machine Self-Understanding, "AI and Consciousness: Theoretical Foundations and Current Approaches", Papers from AAAI Fall Symposium, Technical Report, pp.127-128, Arlington, USA (2007)
13. Rzepka, R., Araki, K.: Artificial Self Based on Collective Mind - Using Common Sense and Emotions Web-Mining for Ethically Correct Behaviors. In: Proceedings of Toward a Science of Consciousness Conference (2009)
14. Rzepka, R., Krawczyk, M., Araki, K.: Using Empathy of the Crowd for Simulating Mirror Neurons Behavior, To Appear in the Proceedings of the 4th Workshop on Emphatic Computing, IWEC 2013, IJCAI (2013)