2013 IEEE 13th International Conference on Data Mining Workshops

Possible Usage of Sentiment Analysis for Calculating Vectors of Felific Calculus

Rafal Rzepka and Kenji Araki
Graduate School of Information Science and Technology
Hokkaido University
Kita-ku, Kita 14 Nishi 8, Sapporo, Japan
Email: kabura,araki@media.eng.hokudai.ac.jp

***Abstract*—In this paper we introduce an algorithm for affective reasoning based on Bentham's Felific Calculus known also as the hedonic calculus. Knowledge recquired for the task is retrived from a blog corpus by means of sentiment analysis on sentences containing an action or state input. This approach allows a machine to gather information on how usually other people feel when something happens, why people did it and what could happen after the act. Such knowledge is important for understanding actions of others, and for acquiring emphatic skills by a machine. In addition to emotion categorization of Nakamura, we introduce two lexicons based on McDougall's instinct classification and Kohlbergian stages of moral development, then show some basic efficiency of the retrieved knowledge.**

I. Introduction

Sentiment analysis from the beginning of its history was meant to discover people's opinions. However, our claim is that it has much bigger potential for the Artificial Intelligence field. It can be used for retrieving wider range of knowledge about human and for creating an agent equipped with empathy, a machine that understands why people perform particular actions and what their consequences might be – not only in their emotional but also instinctual and social realms. Machines text understanding capabilities in the era of Big Data and improving machine translation research could also lead to instant comparison of human behaviors in different cultures. Human beings are different in terms of race, beliefs, values and cultural backgrounds, however, we are all similar in terms of biological reactions and internal need of pursuing well-being [1]. These seem to be identical for *homo sapiens* species. We do not claim here that everyone reacts the same in similar situations. Actually, we agree with Friedberg [2] who wrote that emotions are "contextual", and that they depend on the circumstances and on types of cognition. Therefore we concentrate on thorough context analysis which is currently out of AI systems' reach and process emotions depending on actors, patients, places, duration, etc. These features were already proposed by Jeremy Bentham for calculating utility [3] and we believe that to realize his idea we need to combine sentiment analysis techniques with context processing algorithms. This paper introduces a basic set of ideas for an artificial empathy agent based on experiences of crowds, a program that in a long span could probably help not only machines to reason about humans but also help when human imagination fails while judging others.

II. Comparison with Other Approaches

The main contribution of our research in this early stage is development of small lexicons allowing existing sentiment analysis algorithms to retrieve wide range of common sense knowledge. It must be clearly stated that these lexicons are not meant to replace existing emotional ones as WordNet-Affect [4], SentiWordNet [5] or SenticNet [6] but to automatically enrich ontologies like ConceptNet [7]. Our lexicons are constructed for retrieving specific types of knowledge, and when it comes to emotions we are using an existing classification of Nakamura [8] which we find more suitable for Japanese language than classical Western approaches [9]. Our idea is that the *methods* for retrieving affect can be utilized for acquiring usual *reasons* and *consequences*. For instance, an usual affect analysis system recognizes that beer is described as pleasant X times in texts and Y times as unpleasant, but our system is supposed to deduce what instincts make us drink, why we decide to do it, what senses are used, what are short (intoxication) and long distance consequences (lost of driving license) for a user or environment, how many people will be affected, etc.

III. Retrieval Methods

Before we introduce the ideas for implementing Bentham's idea, we need to explain our method for dividing pleasant and painful statements, positive and negative instincts (reasons) and consequences of actions. All proposed lexicons, used NLP tools and blog corpus are in Japanese language as we currently perform experiments limiting search span to only one culture. The below explained sets of phrases (lexicons) are later used for observing quantitative relations between causes and effects and for recognizing their polarity.

A. Emotional Consequences Lexicon

As mentioned above, this lexicon contains a set of words expressing emotional states and is borrowed from Nakamura [8]. It contains adjectives: *ureshii* (happy), or *sabishii* (sad); nouns: *aijō* (love), *kyofu* (fear); verbs: *yorokobu* (to feel happy), *ai suru* (to love); fixed phrases/idioms: *mushizu-ga hashiru* (give one the creeps [of hate]), *kokoro ga odoru* (ones heart is dancing [of joy]); proverbs: *dohatsuten wo tsuku* (be in a towering rage), *ashi wo fumu tokoro wo shirazu* (be with ones heart up the sky [of happiness]); or metaphors/similes: *itai hodo kanashii* (sadness like a [physical] pain). Originally lexicon contained 1677 items (words and phrases) describing

978-0-7695-5109-8/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDMW.2013.70

emotional states but we created a shorten version for faster matching and noise elimination (the dictionary is based on Japanese literature and contains many archaic expressions). Nakamura determined in his research 10 emotions classes and in our research we follow his classification. The breakdown (with number of items per emotion type) is as follows: joy (224), anger (199), gloom (232), fear (147), shame (65), fondness (197), dislike (532), excitement (269), relief (106), surprise (129).

B. Instinctual Causes Lexicon

We performed a research survey on human instincts and found William McDougall's classification [10] most appealing and fitting our purposes. His idea is that our instincts consist of three following components: *perception* – human beings pay attention to stimuli relevant to our instinctual purposes; *behavior* – human beings perform actions that satisfy their instinctual purposes; *emotion* – instincts have associated negative and positive emotions. What was different from classic stimulus-response based behaviorism in his case is purposiveness of instincts meaning that they are goal-directed. Below we show particular McDougall's instincts with the technics we use to retrieve their values (co-occurrences in text resources).

- **Escape**: words associated with fear were collected, for example *scary*, *scared*, *fearful*, *terrifying*, *run away*, *horrifying* or *hair-raising* (21 phrases in total).
- **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).

TABLE I. CATEGORIES AND NUMBERS OF ITEMS IN SOCIAL CONSEQUENCES LEXICON.

POSITIVE	NEGATIVE
Praises (18)	Reprimands (33)
Awards (25)	Penalties (15)
Society approval (8)	Society disapproval (8)
Legal (8)	Illegal (8)
Forgivable (6)	Unforgivable (5)

- **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).

C. Social Consequences Lexicon

For social consequences retrieval we have created a lexicon inspired by Kohlberg's theory on moral stages development [11]. In short, it divides our lives in particular developmental steps, where in the first we are oriented toward obedience and punishment and think how we can avoid punishment. Then we turn to a self-interest orientation asking ourselves what are the benefits of our acts. In the second stage, we start caring about an interpersonal accord and conformity (social norms). Next, an authority and social-order maintaining becomes important and we achieve "law and order morality". The third level includes social contract orientation and universal ethical principles - we acquire so called "principled conscience". These stages inspired us to create a polarized lexicon which mirrors first developmental steps. The items in the lexicon were distributed as shown in Table 1. Phrases in particular categories were written in different styles (kana / kanji), cases and tenses, often stemmed for broader matching coverage. Most of the words were taken from Japanese thesauri, so the *awards* category has many synonyms of prizes, and the *punishment* category consists also of words and phrases describing imprisonment, fines, etc.

D. Web-mining Process

We developed a simple technique for extracting associations from the Web. It takes a short action description as an input and counts how many times the input query occurs with phrases from above introduced lexicons. The technique is composed of four steps: a) accepting any input phrase simply describing human action (object - particle - verb); b) modification of the phrase with causality morphemes (conditional and continuative forms, 9 in total); c) searching for the modified phrase in the corpus (Apache Solr's exact match feature); d) matching words from predetermined lexicon and extracting associations; e) creating a ranking of top causes and effects. The input phrases ending with a verb or an adjective are modified grammatically by the addition of 9 above mentioned causality morphemes, which correspond to causality markers like *because* or *since* in English. Finally, the modified phrases are queried in the blog corpus (5.6 billion words in 350 million sentences) made by indexing ameba.jp, popular Japanese blog site [12]. All matching sentences are extracted from the corpus and cross-referenced with the expressions contained in lexicons described below. The higher hit-rate of an expression in retrieved consequences, the stronger the association of a given act (or state) to the consequence type becomes. Blog entries where input phrase was found are divided by semantic analysis tool ASA[1] into chunks. The system is set to search for instinct phrases only on the left side, and emotional and social consequences only on a right side of input phrase.

IV. Bentham's Felific Calculus

Probably the first attempt to explain human behavior algorithmically was made 214 years ago by one of the fathers of utilitarianism, Jeremy Bentham, who saw our lives as a never ending struggle for maximizing *pleasure* while minimizing *pain* [3]. Depending on action's intensity, duration, possible outcome, etc., the calculus can also measure amount of negative and positive loads of the action, and help a machine to explain a doer's motivations. To the authors' best knowledge no computer science research on implementing this famous notion was made, probably because implementation of all vectors proposed by Bentham requires several complicated modules and each of them seems very difficult to compute, mostly due to the idea's high level of abstraction. However, we believe that we can (to some extent) skip the step of creating sophisticated algorithms and use semantic analysis techniques to retrieve reasonable output for practically any action as an input by borrowing the Wisdom of Crowd. Below we introduce all 7 vectors of hedonic calculus with our proposed methods. Although the ideas are for Japanese language, we believe that these methods can be easily recreated for any language with a basic NLP toolset, phrases for lexicons and Internet resources – the bigger the better.

A. Intensity

While estimating level of pleasure and pain, the following question must be answered: "how intense was an act leading to positive or negative consequences?". To perform this task the algorithm needs to recognize not only emotive words but also adverbs that intensify them and estimate their strengths. We prepared a set consisted of 24 intensifiers as "very" or "a lot" and 6 deintensifiers as "comparatively" or "a bit". Currently program adds 0.5 to a hit when a intensifier proceeds lexicon phrase and subtracts 0.5 from it in case of detected deintensifier.

B. Duration

Bentham's idea for the importance of time in estimating pleasure is easy to understand, but not so easy to implement. The question that needs to be answered here is "for how long the pleasure (or pain) would last?". For instance if a party lasts for *30 minutes* it is presumably not so much fun as a *few hours* party and if somebody's headache lasts for *days*, not *hours*, the feelings of the sufferer will vary. To measure the time, we created an algorithm to calculate the duration of a given action by ourselves because most of the research on recognizing time span was conducted for English language. The program utilizes a temporal expressions database created as a result of a blog corpus analysis with the use of time tag from Juman dictionary[2]. A set of rules describing time points or duration periods corresponding to particular temporal expressions was manually added to the database which allows our system to look for such expressions. Depending on the type of temporal expression or the input sentence's structure, a predefined duration value is output or calculation on the basis of two given time points is performed. The duration value is given in a number of days. So for a sentence "I am going to be busy for whole year starting tomorrow" the output is 365.2425 (the length of an average year in days), and for "It was raining from morning till evening" it is 0.4582 (approximation of 11 hours).

C. Certainty – Uncertainty

These vectors show the probability that current state will be changed. To estimate the likelihood of people's statements more accurately, we have developed a set of words (mostly adverbs) studying linguists achievements in the field of lexical and grammatical marking of evidentiality. For example words like "probably" and "certainly" or expressions as "I believe that", "I am sure that" or "I hope" and "maybe / perhaps" are capable to subtract or add 0.5 points to the found sentence score.

D. Propinquity – Remotness

This pair needs to predict how soon a positive consequence will occur. We plan to use the *Duration* estimation algorithm to search for usual time periods between acts and consequences. In case of low coverage we are going to use WordNet [13] and ConceptNet [7] to broaden the search. Currently we have prepared words divided into two subsets: "soon" (from the act time and the end of the same day) and "later" (from the next day to infinity), however they are not yet used as the authors have not yet reached any agreement on how this vector should influence the score.

E. Fecundity

Probability that an act will preserve the current state are calculated from the usual web search for $cause$ – $state$ – $change of state$ (social + emotional consequences) triplet. For example *eating* preserves good state of *being satisfied* and it is relatively easy to retrieve under the condition we have sufficient number of possible states.

F. Purity

Probability that an act will not cause an opposite consequence type is estimated with the same set of tools as *Fecundity*.

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

[2] Juman System, a User-Extensible Morphological Analyzer for Japanese. Version 7.0: http://nlp.ist.i.kyoto-u.ac.jp/index.php?Juman

G. Extent

This vector, added by Bentham later, needs to retrieve information on how many people usually are influenced by acts similar to the input one that is currently being analyzed. Although looking trivial, there are many cases where the number of people is not exactly specified (like "crowd" or "few") and heavily depends on context. We plan to use similar approach to both Extent and Duration vectors where usual adjectives as "long" or "many" will change their values with every entry. If context information is insufficient, the system will have to perform additional search or ask users for details to fill the contextual gap.

V. Evaluation Experiments

Here we present results of preliminary tests for efficiency of the algorithm and the used lexicons. A list of 127 action phrases as "(to) eat a hamburger" or "(to) kill a cow" was used for inputs. It was an extended set of inputs used previously for recognizing ethically problematic acts. Everyday life actions like "keeping a pet" were added to the original set of 100 phrases.

A. Categorization Efficiency

First we evaluated instincts as possible reasons for an action and took only one top instinct for an action input. Proposed system achieved precision of 75.0% with the recall of 53.54%. After a closer look we have noticed that majority of state describing inputs (like "man is alive") are hard to be evaluated as ones having instinctual motivations and we excluded them (11 phrases, 9% of all phrases) together with 3 phrases input in erroneous Japanese. For the second test we checked all the retrievals (553 hits), not only the top-scoring categories. The first author performed both preliminary evaluations and this more thorough judgement process showed that 77.78% of category assignments were correct. The recall dropped to 49.61% but it was unavoidable as state phrases with high hit-rates were ignored. In the second step we have examined quality of retrieved consequences. The system achieved precision of 78.20% and 0.52 f-score for emotional, but slightly lower results for social consequences retrieval module: 70.50% with 0.48 f-score. However, if we assume that it was a sentiment analysis-like processing task, the results can be treated as relatively high because of usually low human agreement when sentiment evaluation is made.

B. Results Analysis and Conclusions

Restrictions we set caused low recall but we needed them in order to avoid noisy retrievals and in the end only 490 sentences in average were processed for one input. Problematic expressions in the lexicon (e.g. *crying* can signalize more than one instinct working) and classifications philosophically difficult to evaluate are another problems. For instance *cheating on partner* was classified as an effect of "Parental" instincts category and *going by plane* was in "Repulsion". However, these ambiguous cases should be eliminated in the next step where we will concentrate on specific contexts, because for example flying a plane can be caused by repulsion in some specific cases like terror attacks. Wider error analysis are difficult to be described in full within a short paper therefore we plan to elaborate about the system performance in a separate paper.

VI. Conclusions and Future Work

The main purpose of this paper was to introduce the possibilities of lexicons based on social sciences for calculating vectors of Bentham's hedonic calculus. We performed small preliminary experiments, which showed that the proposed techniques can lead to determining reasons and consequences of human acts as the efficiency is quite high even before applying any context processing like negation recognition. We believe that equipped with tools like ours not only machines could benefit from reasoning about human actions – researchers of sociology or psychology and lay people could be provided with broader interpretations when exploring behaviors and become less biased thanks to wider data. We believe that humankind has reached new era of empathy and rationality, also due to more global knowledge, and with our paper we want to provoke a discussion about an answer to following question – will future machines improve our tolerance of diversity or compassion for our fellows? We think that semantic analysis methods can be applied not only to opinions and can lead to more user-friendly machines, and even tools that can help their user to be a better person. We also believe that Bentham's ideas can give useful hints for deeper, contextual semantic analysis itself and for that reason we decided to share our idea with the community in such an early stage of development.

References

[1] S. Harris, *Moral Landscape, How Science Can Determine Human Values*. Free Press, 2010.

[2] F. Friedberg, *Caveat Homo Sapiens, The Furtive Mind*. University Press of America, 2000.

[3] J. Bentham, *An Introduction to the Principles and Morals of Legislation*. London: T. Payne, 1789.

[4] C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet," in *Proceedings of LREC, vol. 4*, 2004, pp. 1083–1086.

[5] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," in *Proceedings of based sentiment classification, ACL HLT*, 2011, p. 132.

[6] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis," in *Proceedings of AAAI FLAIRS*, 2012, pp. 202–207.

[7] R. Speer and C. Havasi, "Representing general relational knowledge in conceptnet 5," in *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 2012.

[8] A. Nakamura, *Kanjo hyogen jiten [Dictionary of Emotive Expressions]*. Tokyodo Publishing, 1993.

[9] C. M. Whissell, *The Dictionary of Affect in Language. Emotion: Theory, Research and Experience: vol. 4, The Measurement of Emotions*, R. Plutchik and H. Kellerman, Eds. New York: Academic, 1989.

[10] W. McDougall, *Outline of Psychology*. Methuen And Company Limited, 1923.

[11] L. Kohlberg, *The Philosophy of Moral Development*, 1st ed. Harper and Row, 1981.

[12] M. Ptaszynski, R. Rzepka, K. Araki, and Y. Momouchi, "Annotating syntactic information on 5 billion word corpus of japanese blogs," in *In Proceedings of The Eighteenth Annual Meeting of The Association for Natural Language Processing (NLP-2012)*, vol. 14-16, 2012, pp. 385–388.

[13] G. A. Miller, "Wordnet: A lexical database for english," *Communications of the ACM*, vol. 38, pp. 39–41, 1995.