ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-606-5-169

Classification of Dreams Using Machine Learning

Stan Matwin [1,2] and **Joseph De Koninck** [3] and **Amir H. Razavi** [4] and **Ray Reza Amini** [5]

Abstract. We describe a project undertaken by an interdisciplinary team of researchers in sleep and in and machine learning. The goal is sentiment extraction from a corpus containing short textual descriptions of dreams. Dreams are categorized in a four-level scale of affections. The approach is based on a novel representation, taking into account the leading themes of the dream and the sequential unfolding of associated affective feelings during the dream. The dream representation is based on three combined parts, two of which are automatically produced from the description of the dream. The first part consists of co-occurrence vectors, which — unlike the standard Bag-of-words model — capture non-local relationships between meanings of word in a corpus. The second part introduces the dynamic representation that captures the change in affections throughout the progress of the dream. The third part is the self-reported assessment of the dream by the dreamer according to eight given attributes. The three representations are subject to aggressive feature selection. Using an ensemble of classifiers and the combined 3-partite representation, we have achieved 64% accuracy, which is in the range of human experts‘ consensus in that domain.

1 INTRODUCTION

Research in psychology shows that emotion is a prominent feature of dreams [3] [5],[10], which makes dreams an interesting corpus for automatic analysis of emotional content. Recent findings from brain imaging studies have shown an increased activation of limbic and paralimbic areas during Rapid-Eye Movement (REM) sleep [6]. Because dreams are strongly associated with this sleep phase, this may account for the emotional intensity of dreams [3]. However, further studies are still needed to better understand the origin as well as the potential role of the emotionality of dreams. Typically, the level of emotions or sentiments is assessed in dreams by content analysis made by human judges using scales of various levels, or by dreamers themselves. Most of the studies on dreams have used time-consuming coding systems that depend on a ranker‘s judgment. Hence, it is of interest to develop efficient means of scoring dreams, which can be used with large data banks and reproduced across laboratories, and which can, at least to some extent, alleviate the human effort needed in the current human scoring of dream descriptions. To the best of our knowledge, our team is the first to apply machine learning and natural language processing techniques to the analysis of dream descriptions. Furthermore, quantification of qualitative data such as phenomenological reports is of great value to scientifically based psychological research. And such a tool could be used in quantifying the emotional aspect of subjective reports.We present here our work on developing a machine learning solution for the categorization of emotional contents of dreams on a 4-level scale. We used a value from 0 to 3 to estimate both the positive and the negative content of dreams, as applied by independent judges, and we compared it to the automatic analysis. The granularity of our scale (4 levels) was chosen to reflect the variety of sentiment experiences and to maintain simplicity. Previous work aiming at drawing a link between negative sentiments in dreams and dreamer‘s stress also relied on content analysis of written dreams [2]. Assessing dream descriptions on the negative scale that we present could be applied as a feature in a larger system for stress analysis. A more general application of automatically-analyzing dream sentiments would be the mining of large dream banks and discovery of unsuspected data about sentiments in dreams of individuals of different age, social status, etc.

The paper discusses the basic issues of emotions in dreams, presents the data we are working with, and discusses what we believe is the main challenge of this application — the representation used for machine learning. It then gives a brief account and discussion of our early experimental results.

[1] School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada, stan@site.uottawa.ca
[2] Institute for Computer Science, Polish Academy of Sciences, Warsaw, Poland
[3] School of Psychology, University of Ottawa, Ottawa, Canada, jdekonin@uottawa.ca
[4] School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada, araza082@site.uottawa.ca
[5] School of Psychology, University of Ottawa, Ottawa, Canada, ramin029@uottawa.ca

2 EMOTIONS IN DREAMS

Sentiment analysis is an important component for the studies of dreams since emotions are considered by many as being responsible for structuring the content of dreams [5],[10]. Recent findings from brain imaging studies have shown an increased activation of limbic and paralimbic areas during Rapid-Eye Movement (REM) sleep [6]. Because dreams are strongly associated with this sleep phase, this may account for the emotional intensity of dreams [3]. However, further studies are still needed to better understand the origin as well as the potential role of the emotionality of dreams. Until now, most of the recent studies on dreams use the classical scales of Hall and Van de Castle [17], which are considered as being the most detailed and complete coding system available for scoring dreams. It comprises various scales measuring both positive and negative content, such as the presence of friendly or aggressive interactions, emotions, good fortunes or misfortunes, and successes or failures. However, this approach is time consuming and depends on the ranker‘s judgment. Therefore a system allowing objective means of scoring dreams that are independent of a human judgment, and that can be reproduced across laboratories, is of great interest. So far, automatic analysis has not been used in studies of emotions in dreams. The development of this technology could improve our knowledge on dreams and be a major breakthrough in this research area.

3 THE DATA

Dreams were gathered from a dream bank created during a normative study conducted at the Sleep Research Laboratory of the University of Ottawa. Volunteer participants consented to the use of their dreams for this research. Their participation mainly consisted of completing a brief dream diary at home during a maximum of three weeks, and writing down all the dreams they remembered when they woke up, up to a maximum of four dreams. The dreamers were asked to rank (0-3) each of their dreams regarding the following features: " Joy, Happiness, Apprehension, Anger, Sadness, Confusion, Fear, Anxiety, Negative Affect and Positive Affect". We refer to this ranking as self-assessment.

A sample of 776 dreams, reported by 274 individuals of varied age and sex, was chosen for the dream sentiment analysis task. From those, a pure English subset of 477 tagged dream descriptions were used for training and testing the software. The dreams were categorized by a judge (an expert psychologist) according to the 4-level scale. A previous study [9], where each description was categorized by two independent judges, showed that the inter-judges agreement varied between 57.7% and 80.8%. The agreement was lower for the positive scale compared to the negative scale, and the score on the positive scale was not well differentiated from one dream to another; furthermore, works in dream analysis often concentrate on the negative sentiments in dreams since they are typically more present and differentiated than positive sentiments [5]. The negative scale can therefore be useful in isolation. Hence, we will focus on this scale in subsequent discussion.

Here is a sample dream:

> "Our family and my uncle's family were all in 1000 islands. We sat right beside the water, just above the sand. I don't recall exactly what happened , but we took our blue cooler we used to have 2 years ago. Then we all sat down to eat on the wooden table there, but a lot of people were missing. Then all of a sudden, the kids went swimming..."

and here is how example sentences from other dreams were labeled by a judge (a psychologist):

"I was back in Halifax with some of my high school friends and we were just waking around." (0: Neutral)

"I then got on the street beside a bus stop. The bus I was supposed to take past by without stopping to let me in." (1: Lightly negative)

"I ran to the car and it wouldnt start. So I ran to the bus stop. The bus finally came and I started driving it. When we got to campus, I spent 25 minutes trying to find parking." (2: Moderately negative)

"When we got there we were in the bad part of town. We asked for directions and they pulled a gun out at us." (3: Highly negative)

4 THE REPRESENTATION

In building a representation adequate for the classification of dreams with respect to their affective content, we have decided to exploit three kinds of information describing the dream from the emotional perspective: the semantics of the natural language description of the dream, the dynamic affective change in the dream, as well as the self-assessed feelings of the dreamer about her dream. This led us to a 3-partite representation in which the semantic part was built directly from the text describing the dream using lexical properties of the whole corpus, the dynamic one was using NLP techniques as well as specialized dictionaries, and the subjective one was taken directly from the data. We have selected from each representation the features most important for classification, and then performed final training of the classifier on their union. Below we describe each of the three parts of the representation and the final combined representation of dream descriptions.

4.1 Semantic Representation

Many works starting with Turney [18] and continuing with, e.g., [11] [1], address a similar problem of classifying texts as positive or negative, usually as a binary classification. In [18], emotional orientation of reviews was gauged using Pointwise Mutual Information (PMI). PMI was used to measure the association of a given phrase with a standard positive ("excellent") and negative ("poor") reference. Orientation was then aggregated for the whole review, and used to recommend (or not) and item. [11] is probably the first to use Machine Learning for classifying the sentiment of texts. Working again with reviews, they use the standard BOW representation and point out that classifying the emotional affect of a short document is much harder than categorizing such texts according to their topic. [1] are the first to point out that the BOW representation is inadequate for sentiment analysis and classification. The authors draw upon a deeper Natural Language Processing approach: compositional semantics. In compositional semantics, the semantics of complex expressions (e.g. noun phrases) is built up from the semantics of the constituents, combined then into the semantics of the whole expression using a set of fixed, given rules.The emotional polarity of words is handled by the use of the General Inquirer [4]. We use a somewhat similar approach to obtain the polarity of complex phrases, see sec. 4.2. [8] work on sentiment analysis of blogs. They combine two Naive Bayes classifiers, one classifying word sentiments and the other classifying blog sentiments. They propose a novel way of combining the two classifiers, and observe that the use of word classifiers allows them to work with a much smaller set of labeled blogs for training the blog classifier.

The related work mentioned above targets a goal which is somewhat different from ours, and works on different data. The granularity of our classification, i.e. the four classes, makes the task more challenging. This challenge cannot be addressed the standard Machine Learning solution to multi-class tasks: supplying more labeled data, as obtaining labeled data in dream research using the human judges is expensive and time consuming. We need to address this at the level of a more informative representation of dream descriptions. Moreover, our task presents an additional challenge as the texts are often in a colloquial language, and are shorter than the ones in [18]. We use words as the smallest meaningful unit of any context that plays a role in expressing meaning or intention through text. Therefore, capturing the right sense of any word in a context in the representation method is crucial. We work under the Distributional Hypothesis that says that words which occur in similar contexts tend to be similar. This posits a representation which can take into account the context (i.e. other words) and their mutual relationships as acquired from the whole corpus. The most common method of text representation, the Bag-of-Words (BOW), does not meet these requirements: texts are represented solely by the words they contain. A "first order" approach to co-occurrence by simply looking at the vectors would only tell us which words directly contributed to the contexts. However, given that dream descriptions are very short, the vectors are very sparse.

4.1.1 Second-order "soft" representation.

Schütze proposed in 1998 a powerful method [16] called second-order co-occurrence context representation. Good performance of the second-order representation of contexts was already proved in [16] and [12]. Although until now, the second order co-occurrence has been applied in variety of unsupervised purposes (e.g. disambiguation of names [13]), this is the first application of the soft augmented version of it second-order co-occurrence to a supervised text analysis task.

We tokenize the corpus, in order to build a soft co-occurrence matrix in which the closeness of co-occurring pairs is recorded. The closeness is determined by considering several configurations of any pair of words in a sentence (our window size). The strongest co-occurrence is a bigram in a sentence, then a pair of words separated by one word, more than one word, two words separated by at least one word and a comma, or a semi-colon, or a quotation. Normally, co-occurrence is considered in a specific context or in a window of a limited size such as 3 to 7 words before or after a target word, which would restrict the total context size from 7 to 15 words. We select sentences as our window size. In other words, except for the first configuration, the rest have a fraction of co-occurrence impact on the matrix (lower weight).

In the co-occurrence matrix (over the whole corpus) each row represents a word x and a column represents the word y of a co–occurrence configuration. The cell values represent the closeness $c_{x,y}$ of x and y in the corpus and are calculated as follows

$$c_{x,y} = \frac{2(w_1 \cdot df_{1_{xy}} + w_2 \cdot df_{2_{xy}} + ... + w_m \cdot df_{m_{xy}})}{df_x + df_y} \quad (1)$$

$$df_x = df_{1_x} + ... + df_{m_x}; df_y = df_{1_y} + ... + df_{m_y} \quad (2)$$

where w_i is the weight of configuration i, $df_{i_{xy}}$ is the frequency of co-occurrence of the pair x, y in configuration i in the corpus, m is the number of distinct word pair configurations, df_{i_x} is the frequency of occurrence of the word x in the configuration i with any word in the corpus. The closeness values $c_{x,y}$ computed according to (1) above are normalized to be between 0 and 1. This matrix is large (and could be very sparse if has built up over one document or small number of short texts, since most words do not co–occur with each other). There is an option to apply SVD to this co-occurrence matrix to reduce its dimensionality. Each row of the matrix is a vector that represents the given word via its co–occurrence characteristics.

In the first step of building this representation each sentence of a short text in the corpus is represented by averaging[1] the features' vectors of all words in the sentence, which are extracted from the soft co-occurrence matrix. The sentence representation vector at this stage has several times more non-zero features than the BOW representation of the same sentence. In this step, the soft co-occurrence matrix does not include stop words, hence the stop words cannot affect the creation of the representation vectors.

In the second step, we compute another vector by again averaging the representation vectors of all the sentences in the dream description (from the sentence representations obtained during the first step). Performing the aggregation (averaging) further increases the number of non-zero elements of the text representation vector. Our experiments show that almost 90% of the features are non-zero by now. The value of an element of the vector is indicative of strength of the relationship of the corresponding word to the sentence or the whole text that contains this feature. This value, however, does not show directly if the feature occurred in sentence/text or not; it globally represents the relevance level of each word to the sentence. In other words, in addition to computing the explicit participation of a given feature in a given document, we accumulate the participation of other similar features with respect to their closenesses to the given feature. This means that even if we eliminate one of the features from the feature space (after creating the soft co-occurrence matrix), we can still expect to keep its discriminatory power in the classification task, if that feature sufficiently co-occurred with other related features in the corpus.

If the number of tokens in a sentence is n, the number of pairs extracted from a sentence can be calculated as: O(n^2) AND linear in the number of sentences in a corpus. We empirically observed it took a fraction of a second to process each short text.

[1] The averaging function can be changed with another aggregation function like maximum.

4.1.2 Use of stop list

If we removed the stop words from the text prior to determining the configuration of each word pair in its context, we would have modified those configurations in which stopwords are involved. Moreover, some words that have one or more than one word in between could have been assigned a configuration that would view them as adjacent or closer than in reality, and in this way the algorithm will overestimate the degree of co-occurrence. We therefore remove from the matrix the rows/columns corresponding to stop words.

4.1.3 Contrast parameter

While the BOW features are sharply related to the presence or absence of the word they correspond to, the repeated aggregating in the process of building the semantic text representation vectors brings about smoothness of the feature space. By this we mean that in the soft semantic representation a a feature (word) related through the corpus with a given description may have a non-zero value in the representation, without that word being present in the text. Sharpness or smoothness may advantageous choices for specific tasks. In topic classification, one may want sharpness as there are keywords very closely related with a topic. In relevance ranking, we have determined in another project [7], more smoothness is required. In sentiment classification, one also wants more smoothness than in topic classification. How can one control smoothness? With the aggregating function. In sec. 4 we were smoothing the features with averaging, but other aggregating functions, e.g. min and max, are possible. That means that instead averaging the values of several vectors we take their min or max. We get largest smoothness using the min function for the aggregation, and — conversly— use of the max function results in the smallest smoothness. We have therefore decided to introduce a parameter of the semantic representation, called the contrast, that controls the degree of smoothness. If we make the values of the contrast discrete, between -9 and 9, -9 corresponds to maximum smoothness, 0 to average (and corresponds to the use of average as the aggregating function), and 9 to the most sharp features. If we desire a particular value of the contrast, e.g. α, we aggregate the features using average and scale the obtained value proportionately to the distance of the value α form the value of average (0). Our experience indicates that a good value for topic classification is 3 or 4, a good value for relevance ranking is a bit smaller, and a good value of contrast for sentiment classification is 0 or -1. Determining the contrast value for a given task is in fact a problem that could be best addressed by a wrapper or by optimization techniques. We also want to observe that varying the contrast could also be used to obtain a variety of representations of a given text for a committee of ensemble learners.

4.2 Dynamic Representation

The Linguistic Inquiry and Word Count (LIWC) project [14] offers measures of the percentage of positive and negative words in texts. The LIWC dictionary is composed of 2290 words and word stems. We have used LIWC to measure the affect of individual words. We do this by thresholding the number of times a given word occurred in a positive and negative context, and assigning it a positive or negative label. Hence, LIWC gives us the affect of the word. The CMU Link Grammar Parser helped us identify adverbs, which we use to modify the affect of the words that adverbs precede. In a process similar to compositional semantics, we modified the values based on different type of valence shifters such and negations and modals (i.e., very, extremely, etc.) in order to obtain a better representation. This allows us to recognize when the context changes the polarity of a word (for instance the phrase "$isnotkind$" means the opposite of "$kind$" and it should not be counted as positive). Consequently, the initial affective tags are modified according to the severity level of the modifier which is looked up in the modifier table in our system. The values in the table are assigned based on the severity level of the modifier. These values have been adjusted through many iterative experiments, using feedback from the results of machine learning on the obtained representation. We then further modify the assigned affect values if they are the argument of negations such as: not, non, un-, im- and so on. We empirically noticed that contrary to our expectation, when a negation modifies any type of adjective/adverb, its affective influence on the containing context is not completely reversed. For instance, if in a sentence we encounter to an affective word like "happy" with positive affective value (+1), for "not happy"(-1) is not the most appropriate tag, or when we have "so happy" with positive affective value of (+3), for "not so happy" the value (-3) is too low, this value is normally assigned to an expression like "too sad". Also, when we say "He is not so happy!" definitely it does not mean "He is too sad!". More precisely, if a term w has the affect value $a(w)$, we assign to **not** w the value $a(w) + (-1)^{opposite}$, where $opposite$ = 2 if $sign(a(w))$ = - and $opposite$ = 1 if $sign(a(w))$ = +. For instance, in the phrase "It ends at the brink of a steep hill very grassy and green not at all threatening" the word "$threatening$" will be assigned the affect value -1 by LIWC, and to interpret the negation according to the above rule we will add the value +1, obtaining the value 0 after modification.

With this interpretation of terms, we can map the description of the dream into a sequence of affect values, representing the dynamic change of affections as the dream progresses. We call this a dynamic representation, and we call its visualization an *onirogram* (the Greek name for dream is $o\nu\varepsilon\iota\varrho o$). Fig. 1 shows an example of an onirogram. We want to note that psychologists found onirograms to be a very helpful tool allowing them a quick view and analysis of the progress of emotions in a given dream.

The next step is to obtain from an onirogram attributes that can be used in attribute-value classifier induction. We do this by extracting the height and the width of individual moods during the dream, the number of positive an negative moods, the number of change of moods, etc. For most of these quantities, we then take the average, the standard deviation, the minimum and the maximum for all moods in the dream. The totals are normalized. Table 1 shows the list of those attributes extracted from the onirogram. Collectively, they constitute the dynamic representation of a dream.

4.3 Self-assessment attributes

The self-reported attributes are taken directly from a dreamer‘s self-assessment (the eight ranks described in sec. 3) Interestingly, one of the results of our research is that there is redundancy in this attributes, i.e. some of them can be almost perfectly predicted by the others (see sec. 5.

4.4 The combined representation

Prior to merging the three components of our representation described above: the semantic representation, the dynamic representation, and the self-assessment, we perform feature selection of the first two. The initial attribute sizes of each representation and the resulting number of attributes are given in Table 2. We applied the Relief Attribute Evaluator from the Weka machine learning toolkit [19]. We have experimented with different levels of feature selection, and we have found out that the aggressive selection shown in Table 2 gives good results, compared to less aggressive selection. Following the feature selection, the three reduced representation are combined into a single vector, which becomes the training set for the machine learning classifier. The labels in the training set are the labels given by the human judges.

Table 1. Description of attributes extracted from the onirogram

Height of positive affect for a mood	Width of a positive affect for a mood (number of words)
Height negative affect for a mood	Width of negative affect for a moo (number of words)
Initial mood (pos. or neg.)	Average dream affect
Number of pos. moods	Number of neg. moods
Total pos. affect (before linguistic modif)	Total neg. affect (before linguistic modif)
Total affect (before linguistic %modif)	Width of neg. affect for all moods in the dream
Width of pos. affect for all moods in the dream	

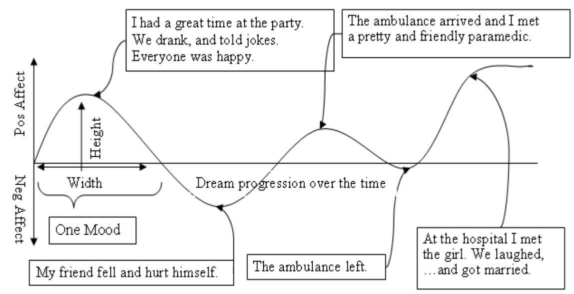

Figure 1. Affection onirogram- Illustration of the polarity and emotional tone of contextualization of dreams over the time.

5 CLASSIFICATION RESULTS AND DISCUSSION

We used two evaluation measures for our experiments. First, we calculate classifiers‘ accuracy — the sum of correct guesses over the total number of guesses — i.e. performance at exactly finding the

Table 2. Attribute selection results.

Attributes Groups	# initial attributes	# attributes after attribute selection	% of category after attribute selection
Text	4618	39	0.9%
Progression	36	21	58.3%
Demographics	2	2	100%
Dreamer Emotion	8	8	100%

right label (e.g., human rates 3, machine guesses 3 would be a correct guess). Second, we calculate the mean squared error of classifier: the average of the squares of the differences between the human labels and the machine predictions. This metric is low when a classifier guesses close to the human value (e.g., human rates 3, machine guesses 2) and becomes high if the classifier is far from human judgment (e.g., human rates 3, machine guesses 0). We report results for stratified 10-fold cross-validations. The baseline accuracy is given by a classifier that always guesses the majority class. In our dataset, 30% of the dreams were rated with label "2"; this is the majority class. Therefore, always guessing "2" results in 30% baseline accuracy. The baseline mean squared error is given by a classifier that always guesses the most probable class. After performing feature selection, we ran many simple and ensemble leaner algorithms on a variety of compositions of selected attributes, applying 10 fold cross-validations. Table 3 compares the best experimental results, based on each group of attributes individually.

Table 3. Results of our best classifiers applied on each of the attribute subsets individually.

Attributes	Number of attributes after attribute selection	Agreement with Human Judges
Text only (without affection tags)	39	55%
Emotion Prog. Analysis	21	49%
Sentiments Only	8	48%

In this step, if we compare our accuracy using the semantic representation method (55%) and accuracy of the previous work [9] (38%), we can see that the semantic method is applying the proper contrast parameter.

In the next step, we combine all the selected attributes and try to find the most discriminative classifier in order to achieve the highest agreement with our psychologists' labels. For the various machine learning models that we tried, we have calculated the accuracy of the learned classifier on the scale 0-3 and the mean-squared error (i.e., the difference with human judgment when guessing incorrectly). A voting committee of three Adaboost and two Bagging meta-classifiers[3] provided the most accurate results with the least mean squared error on the prediction of negative affection, with an Accuracy of 63%, which is significantly better than the baseline accuracy (30%), and the chance probability (25%). The mean-squared error was 0.3617, meaning that almost all errors have only a difference of 1 on the scale. With these results, we could predict 13% better than the previous work on the same task which was based only on the BOW representation method [9].
The results indicate that estimates were at most one level away from human judge scores[4] and offer a promising perspective for the automatic analysis of dream emotions, which is recognized as a primary dimension of dream construction. As for the progression of emotions along the dream reports, our model appears successful at using the estimates to provide a time-course graphical representation. Although we believe there is still room for improving the results, we can say that sentiment analysis based on the contextualized dreams can predict the four levels of Anxiety$\{1,2,3,4\}$ with 71.4% accuracy (baseline: %28.8) and Mean Square Error (MSE) 30%. We obtained also 68% accuracy (agreement with the dreamers on the same scale of 4 points) with the same rate of MSE for Fear, and for other sentiments. Results not less than 60% are useful in contextual sentiment analysis. Finally, an interesting additional experimental finding was that the self-assessment attributes are not independent. We have found out that each of the following self-assessment attributes: joy, happiness, apprehension and anger can be removed from the feature set and be predicted with accuracy of at least 97% by the eighth remaining self-assessment attributes. This finding is practically important because it allows psychologists to administer a simpler questionnaire to the dreamers.

These results offer a promising perspective for the automatic analysis of dream emotions, which is recognized as a primary dimension of dream construction. Larger databases should facilitate analysis and data mining, and emotion specific parameters may improve accuracy. To the extent that dream narrative corresponds to the time course of dream experience, graphical representations should provide a new tool to explore models of dream formation. Further development of this technology could facilitate the analysis and mining of a greater number of dreams of individuals of different age, sex, social status, thus improving our understanding of dreams.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel representation method for automatic analysis of sentiments in dreams. 477 dreams were sampled from a dream bank created for a normative study on dreams. Expert psychologists assigned scores to dream descriptions based on the expressed sentiments. We have converted textual dream descriptions to our combined sematic, dynamic, and self-reported representation method. We have then trained several chosen classifiers and compared the results. The performance of the machine learning system (64%) is close to human judging average agreement (69%). The practical value of this research is in supplying a tool to dream researchers which assists them in the task of assessing the emotional contents of dream descriptions - a process repeated daily in numerous sleep clinics around the world. The method described here alleviates the human expert (psychologist's) effort in the analysis of the emotional contents of dreams, while at the same time resulting in a more consistent assessment. Dynamic attributes contributors to the achieved performance. Word modifiers played an important role in affect extraction. There is some emotional content that is not communicated directly through words.

In the future, we are planning to add a step determining the proper context in order to optimize our window size dynamically and build the representation vectors based on its component contexts (currently our window size is based on sentences.) We can refine the dynamic attributes using known approaches to time-series analysis. We believe that a brief training of the participant dreamers in describing

[3] The simple classifiers which were used for the above classifiers were: Multinomial logistic regression and J48 decision trees.

[4] Literature shows between 57- 80% agreement among the human judgment in this area and range.

their dreams in a more structured format can improve the performance of the the system. The long-term research goal is to combine the analysis of textual dream descriptions with the data obtained from recording the functioning of brain during the dream with EEC and eventually fMRI. In future research using this technique of analysis, dream reports could be obtained orally and ideally immediately upon awakening either in a sleep laboratory or by audio recording in the home environment in order to improve the accuracy of the descriptions of the dream experience. Furthermore, subjects could be asked to narrate their dream respecting the chronology of dream events. The automatic analysis would be applied on transcripts of the narration. This would be particularly useful. For example, the ability to quantify the emotional valence level as it progresses across a dream experience (the onirogram introduced in this paper is the first step in that direction) will allow to relate it, in laboratory studies, to the underlying brain activity measured by electrophysiology or through brain imaging. This will contribute to a better understanding of the physiological substrates of dreaming. Of particular interest will be the examination of the build up of negative emotions in the intriguing phenomenology of nightmares.

The second-order soft text representation of short texts introduced in this paper is a novel technique, which can be used in a variety of other text classification tasks. It has been applied successfully in offensive language detection [15] and in assessing the relevance of medical abstracts in the Systematic Review process [7]. This novel text representation technique is also applicable elsewhere, when short (e.g. less than 50 words) texts abound, e.g. in classifying or filtering blogs and twits.

ACKNOWLEDGEMENTS

This research was supported by Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Houman Abbassian for his assistance with latex.

REFERENCES

[1] Yejin Choi and Claire Cardie, 'Learning with compositional semantics as structural inference for subsentential sentiment analysis', in *EMNLP '08: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 793–801, Morristown, NJ, USA, (2008). Association for Computational Linguistics.

[2] Monique; Delorme, Marie-Annick; Lortie-Lussier and Joseph De Koninck, 'Stress and coping in the waking and dreaming states during an examination period', *Dreaming*, **12**, 171–183, (2002).

[3] G. W. Domhoff, *The Scientific Study of Dreams: Neural Networks, Cognitive Development, and Content Analysis*, American Psychological Association, New York, 1nd edn., 2003.

[4] General Inquirer Project., *The General Inquirer; introduction to a computer-based system of content analysis*, General Inquirer Project, [Edinburgh,, 1974.

[5] M. D. Ernest Hartmann, 'Dreams and nightmare: The new theory on the origin and meaning of dreams', *Dreaming*, **10**, 247–250, (1998).

[6] Pierre Maquet, Jean M. Péters, Joël Aerts, Guy Delfiore, Christian Degueldre, and André Luxen, 'Functional neuroanatomy of human rapid-eye-movement sleep and dreaming', *Nature*, **383**(6596), 163–166, (1996).

[7] Stan Matwin, Alexandre Kouznetsov, Diana Inkpen, Oana Frunza, and Peter O'Blenis, 'A New Algorithm for Reducing the Workload of Experts in Performing Systematic Reviews', in *Journal of the Anmerican Medical Informatics Association*, to appear, (2010).

[8] Prem Melville, Wojciech Gryc, and Richard D. Lawrence, 'Sentiment analysis of blogs by combining lexical knowledge with text classification', in *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1275–1284, New York, NY, USA, (2009). ACM.

[9] David Nadeau, Catherine Sabourin, Joseph De Koninck, Stan Matwin, and Peter D. Turney. Automatic dream sentiment analysis, 2006.

[10] Tore A Nielsen and Philippe Stenstrom, 'What are the memory sources of dreaming?', *Nature*, **437**, 1286–1289, (2005).

[11] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, 'Thumbs up?: sentiment classification using machine learning techniques', in *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pp. 79–86, Morristown, NJ, USA, (2002). Association for Computational Linguistics.

[12] Ted Pedersen and Rebecca Bruce, 'Knowledge lean word-sense disambiguation', in *AAAI '98/IAAI '98: Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, pp. 800–805, Menlo Park, CA, USA, (1998). American Association for Artificial Intelligence.

[13] Ted Pedersen and Anagha Kulkarni, 'Unsupervised discrimination of person names in web contexts', in *CICLing '07: Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 299–310, Berlin, Heidelberg, (2007). Springer-Verlag.

[14] James W. Pennebaker, Martha E. Francis, and Roger J Booth. Linguistic inquiry and word count - liwc2001, 2001.

[15] Amir H. Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin, 'Offensive Language Detection Using Multi-level Classification', in *Proceedings of the 23rd Canadian Conference on Artificial Intelligence*, pp. 16–27, Berlin, Heidelberg, (2010). Springer-Verlag.

[16] Hinrich Schütze, 'Automatic word sense discrimination', *Journal of Computational Linguistics*, **24**, 97–123, (1998).

[17] Lortie-Lussier M. Mercier P. Grenier J. St-Onge, M. and J. De Koninck, 'Emotions in the diary and rem dreams of young and late adulthood women and their relation to life satisfaction', *Dreaming*, **15**, 116–128, (2005).

[18] Peter D. Turney, 'Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews', in *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417–424, Morristown, NJ, USA, (2002). Association for Computational Linguistics.

[19] Ian H. Witten and Eibe Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, CA, 2. edn., 2005.