Automatic Dream Sentiment Analysis

David Nadeau[1, 3], Catherine Sabourin[2], Joseph De Koninck[2], Stan Matwin[1] and Peter D. Turney[3]

[1]School of Information Technology and Engineering
[2]School of Psychology
University of Ottawa
Ottawa, Canada
[3]Institute for Information Technology
National Research Council Canada
Gatineau and Ottawa, Canada
{david.nadeau, peter.turney}@nrc-cnrc.gc.ca, {jdekonin, stan, csabo069}@uottawa.ca

Abstract

In this position paper, we propose a first step toward automatic analysis of sentiments in dreams. 100 dreams were sampled from a dream bank created for a normative study of dreams. Two human judges assigned a score to describe dream sentiments. We ran four baseline algorithms in an attempt to automate the rating of sentiments in dreams. Particularly, we compared the General Inquirer (GI) tool, the Linguistic Inquiry and Word Count (LIWC), a weighted version of the GI lexicon and of the HM lexicon and a standard bag-of-words. We show that machine learning allows automating the human judgment with accuracy superior to majority class choice.

Introduction

Research in psychology shows that emotion is a prominent feature of dreams [2], [6], [11]. 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. In this work, we show how to automatically obtain equivalent measures. We used a value from 0 to 3 to estimate both the positive and negative content of dreams, as applied by independent judges and we compared it to an automatic analysis.

The granularity of our scale (4 levels) was chosen to reflect the variety of sentiment experience and to maintain simplicity. One envisioned application of this measurement is the assessment of the stress experienced by the dreamer. Previous work aiming at drawing a link between negative sentiments in dreams and dreamer stress relied on content analysis of written dreams [1].

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 individual of different age, social status, etc.

From a machine learning perspective, the task of dream sentiment analysis is expressed as a classification problem with labels {0, 1, 2, 3}. The goal of this work is to create a system that can reliably replace human in analyzing sentiments in dreams.

The next three sections go as follow: first, the dream corpus is detailed, then our experiments in automatic dream sentiment analysis are presented and, finally, related works are discussed.

Dream Bank

Dreams were gathered from a dream bank created during a normative study conducted at the Sleep Research Laboratory of the University of Ottawa (UofO). The ethics committee of UofO has approved this normative study as well as the use of the dream bank for future studies. Volunteers were informed that their dreams could be used in other studies on dreams and they all gave their consent. Their participation mainly consist of completing a brief dream diary at home during a maximum of three weeks, and to write down all the dreams they remembered when waking up, until a maximum of four dreams. A sample of 100 dreams, from 29 individuals of varied age and sex, was used in this study.

Manual Sentiment Analysis

The second author of this paper annotated the 100 dreams with two scores ranging from 0-3. One score is for the positive orientation of the dream and the other one is for its negative orientation. The third author of this paper independently annotated 26 dreams. With this second annotation, we calculated the inter-judge agreement shown in Table 1. We also report the mean squared error (MSE) on the agreement. MSE is presented and discussed in the result section. Judges based their rating on example dream passages like in Table 2.

Scale	Inter-judge agreement	MSE
Positive	57.7%	0.54
Negative	80.8%	0.19

Table 1: Inter-judge agreement on 26 dreams.

At this point, we dropped the positive scale. The reason is twofold. First, the agreement between annotators is too low to extract any meaningful results. As a matter of comparison, a majority class rule would have performed at the same level (56% of positive examples were rated ‘0’ on our scale). Second, works in dream analysis often concentrate on the negative sentiments in dreams since they are typically more present and differentiated than positive sentiments [3], [4]. The negative scale can therefore be useful in isolation.

Negative orientation		
Level	**Description**	**Sample passage**
0	Neutral	“I was back in Halifax with some of my high school friends and we were just waking around.”
1	Lightly negative	“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.”
2	Moderately negative	“I ran to the car and it wouldn’t 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.”
3	Highly 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.”

Table 2: Description of the negative scales.

Automatic Dream Analysis

The algorithmic framework presented in this section make use of the online version[1] of the General Inquirer (GI) [10], the online version[2] of the Linguistic Inquiry and Word Count (LIWC) [9], the weighted GI and HM lexicons introduced by Turney and Littman [13], and a bag-of-words approach making use of the Balie[3] text pre-processing software. Results are computed using the Weka machine learning toolkit [15].

The General Inquirer

The first analysis is performed using the General Inquirer [10]. This resource contains 3,600 words labeled “positive” or “negative” (respectively “Pos” and “Neg” tags in GI). Moreover, each word is paired with disambiguation rules that allow identifying if a specific occurrence refers to the sentiment or not. For instance, if the word “kind” is used as an adjective, it means “benevolent, charitable” and has a positive orientation. In the case the word “kind” is a noun, it has no specific orientation. For a particular dream, for example, we obtain “Neg” = 1,6%, meaning that 1,6% of the words has an unambiguous negative orientation (e.g., ANGRY, DISTURB, …) From a machine learning point of view, we create a dataset with the following the features. Note that even if these features are used to score the negative content of dreams, we still use the positive cues that may be useful.

1. the number of positive words in GI
2. the number of negative words in GI
3. the percentage of positive words in GI
4. the percentage of negative words in GI
5. the difference 1-2
6. the log ratio 1/2
7. the difference 3-4
8. the log ratio 3/4
9. the negative orientation level {0,1,2,3}

Features 1 to 4 are taken directly from GI output. The features 5 and 7 give the difference, which is the “remaining” positive or negative strength of a dream. The features 6 and 8 give the log ratio, a value related to the difference but that is less sensitive to the magnitude of the compared features.

The Linguistic Inquiry and Word Count

The second resource we analyzed is the Linguistic Inquiry and Word Count [9] software. The LIWC offers measures of the percentage of positive and negative words in texts. The LIWC dictionary is composed of 2290 words and word stems. In contrasts with the GI, this resource makes no use of disambiguation rule; it relies on simple word count. The richness of LIWC is its scrupulous choice of words made by multiple experts that came to near perfect agreement. We used the following features:

1. the percentage of positive words in LIWC
2. the percentage of negative words in LIWC
3. the difference 1-2
4. the log ratio 1/2
5. the negative orientation level {0,1,2,3}

Again, we use a feature for the difference of percentage scores and a feature for the log ratio.

The Weighted GI and HM

A third strategy is to use the weighted GI and HM lexicons as described in Turney and Littman [13]. The HM lexicon originates from work by Hatzivassiloglou and McKeown [5] that evaluates the semantic orientation of 1600 adjectives. The GI lexicon is derived from the General Inquirer used in the previous section. In both resources, words have a weight that represents their orientation and strength, in the general case. For instance, in the weighted GI, the word “kind” has a weight of +0.056. The sign ‘+’ means the orientation is positive and the absolute value that is near 0 means the word is almost neutral (maybe because of its meaning as a noun). For the matter of comparison, an unambiguous word such as “outstanding” has a weight of +13.41 while “broken-hearted” has a weight of -14.29.

[1] http://www.webuse.umd.edu:9090/
[2] http://www.liwc.net/liwcresearch.php
[3] http://balie.sourceforge.net

We parsed each dream using Balie and count each time the token canonic version exactly match an entry of the GI lexicon or the HM lexicon. We choose to use the following features for both lexicon (GI and HM):

1. the sum of positive weights
2. the sum of negative weights
3. the average of positive weights
4. the average of negative weights
5. the maximal positive weight
6. the maximal negative weight
7. the negative orientation level {0,1,2,3}

For each pair (1-2, 3-4 and 5-6), we also add a feature for the difference and a feature for the log ratio.

The Bag-of-Words

We experiment a Bag-of-words (Bow) approach as a fourth strategy to classify dreams. The Bow approach consists in using as feature every unique word appearing in any dream. Our dream sample is composed of 2758 unique tokens that turns out to be 2758 features. A particular dream (a textual document) is represented by a Boolean vector of length 2758 for which the value of element *j* is 1 if the token *j* appears in the document, and 0 otherwise. This technique is often used in text classification. It allows linking a class (ex.: 1, on the negative scale) to some specific words (ex.: dark, cold, night, etc.)

Results

Two metrics are required in our experiments. First, we calculate classifiers accuracy – the sum of correct guesses over the total number of guesses – i.e. their performance at exactly finding the right label (e.g., human rates 3, machine guess 3). 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 near the human (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).

In Table 3, we report the accuracy percentage (ACC) and mean squared error (MSE) of every strategy. Results are for stratified 10-fold cross-validations.

	Linear regression with GI	Linear regression with LIWC	Linear regression weighted GI & HM	Naive Bayes with BOW
ACC	50%	48%	35%	38%
MSE	0.577	0.608	0.865	1.392

Table 3: Accuracy and mean squared error of various strategies on analysis of dream negative sentiments.

The baseline accuracy is given by a classifier that always guesses the majority class. In our dataset, 33% of dreams were rated with label “2” and this is the majority class. Guessing always “2” results in 33% accuracy. The baseline mean squared error is given by a classifier that always guesses the average of classes. The average of all classes is 1.37 in our dataset. It results in a mean squared error of 0.993.

Features from the General Inquirer outperform other strategies accuracy (highest number of correct guesses) and mean squared error (lowest difference with human judgment when incorrectly guessing). LIWC is considered as good as GI since there is no statistically significant difference between both resources.

We tried many different supervised learning algorithms, but the best result was linear regression. Standard classification algorithms have the downside of resulting in bad mean squared errors. In Table 3, the last column (BOW) is for a Naïve Bayes algorithm known to perform well in text classification. Even if the accuracy is not the lowest, the mean squared error is the worst.

Discussion

The best features to automate dream sentiment analysis are from the GI tool [10] and the LIWC tool [9]. We believe this constitutes a significant first step in this field. Even if 50% of accuracy may appear to be a poor score, it is statistically better than the baseline accuracy (majority class guessing) with 95% confidence.

The MSE of 0.577 for an accuracy of 50% means that most errors have a difference of 1 on the scale (e.g.: human rates 3, machine guesses 2). As a matter of comparison, if every error was for a difference of 1, it would result in a MSE of 0.5 (50 dreams out of 100 with an error of 1 or -1 = 50 time a squared error of 1 out of 100 = mean squared error of 0.5). If every error was for a difference of 2, the MSE would be 2.

Related Works

Dream Analysis in Psychology

Sentiment analysis is an important component for the studies of dreams since emotions are considered by many as responsible for structuring the content of dreams [4], [8]. Recent findings from brain imaging studies have shown an increased activation of limbic and paralimbic areas during Rapid-Eye Movement (REM) sleep [7]. Dreams being strongly associated with this sleep phase, this may account for the emotional intensity of dreams [2]. 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 [3], which are considered as being the most detailed and complete coding system available for scoring dreams [2]. 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 system is time consuming and depends on the rater’s judgment. It is of greatest interest to develop objective means of scoring dreams that are independent of a human judgment and that can be reproduced across laboratories. 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.

Sentiment Analysis in AI

In this work, we classify whole texts using 4-level scales. In most related literature, texts are analyzed at the sentence-level. This representation would be an interesting alternative for our work but, unfortunately, the UofO dream bank is not annotated at the sentence-level at this time. Moreover, many works (e.g., Turney [12]) formulate the problem as classifying texts as positive or negative (binary classification). This formulation differs from our 4-level scale that we motivate by the need of fine grain analysis of sentiment strength for further processing (e.g., analyzing stress level of dreamers). We believe our problem formulation is more difficult than the binary classification but gives more flexibility.

The most severe limitation of our work is the rather limited use of context. In [14], negations and modalities handling is added to a model making use of the GI and the HM lexicons. It allows recognizing when the context changes the polarity of a word (for instance the passage "is not kind" means the opposite of benevolent, charitable.) This improvement is reported to future work items.

Conclusion and Future Work

In this paper, we show how to automate dream sentiment analysis. We specifically experimented with techniques aiming at rating a dream on a 4-level negative scale. We reached accuracy of 50% with a mean squared error of 0.577, a statistically significant improvement over the majority class guessing. We found that the GI and LIWC resources offer the best features from an automatic dream sentiment analysis point of view.

In our future work, we will first extend our dataset. We expect that this will significantly improve our results, given that we have a 4-class problem and only a very limited set of labeled instances. We will also improve the handling of negations and modalities that can completely change the polarity of words in our current framework. The long-term research goal would be to support further processing in the dream analysis field such as stress analysis.

Acknowledgement

We would like to thank Sonia Matwin who introduced us to the LIWC resource.

References

[1] Delorme, M.-A., Lortie-Lussier, M., and De Koninck, J. (2002). Stress and Coping in the Waking and Dreaming States during an Examination Period. *Dreaming*, 12(4), pp. 171-183.

[2] Domhoff, G. W. (2003). *The Scientific Study of Dreams: Neural Networks, Cognitive Development, and Content Analysis*. American Psychological Association, New York.

[3] Hall, C. S. and Van de Castle, R. L. (1966). *The Content Analysis of Dreams.* Meredith Publishing Company, New York.

[4] Hartmann, E. (1998). *Dreams and Nightmares: The New Theory on the Origin and Meaning of Dreams.* Plenum, New York,

[5] Hatzivassiloglou, V. and McKeown, K. R. (1997) Predicting the Semantic Orientation of Adjectives. *Proc. Conference of the European Chapter of the ACL.*

[6] Hobson, J. A., Stickgold, R. and Pace-Schott, E. F. (1998). The Neuropsychology of REM Sleep Dreaming. *NeuroReport*, Vol. 9, pp. R1-R14.

[7] Maquet, P., Péters, J. M., Aerts, J., Delfiore, G., Degueldre, C, Luxen, A. and Franck, G. (1996). Functional Neuroanatomy of Human Rapid-Eye-Movement Sleep and Dreaming, *Nature*, Vol. 383, pp.163-166

[8] Nielsen, T. A. and Strenstrom, P. (2005). What are the Memory Sources of Dreaming? *Nature*, Vol. 437, pp.1286-1289.

[9] Pennebaker, J. W., Francis, M. E. and Booth, R.J. (2001) *Linguistic Inquiry and Word Count LIWC2001*. Erlbaum Publishers, Mahwah, NJ.

[10] Stone, P. J., Dunphy, D. C., Smith, M. S., Ogilvie, D. M. (1966) *The General Inquirer: A Computer Approach to Content Analysis*, The MIT Press.

[11] St-Onge, M., Lortie-Lussier, M., Mercier, P., Grenier, J. and De Koninck, J. (2005). Emotions in the Diary and REM Dreams of Young and Late Adulthood Women and Their Relation to Life Satisfaction. *Dreaming*, Vol. 15, pp.116-128.

[12] Turney, P. (2002) Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proc. 40th Annual Meeting of the Association for Computational Linguistics*.

[13] Turney, P. and Littman, M. (2003) Measuring Praise and Criticism: Inference of Semantic Orientation from *Association. ACM Transactions on Information Systems* (TOIS). 21(4). pp. 315-346.

[14] Wilson, T., Wiebe, J. and Hoffman, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of HLT/EMNLP.*

[15] Witten, I. H. and Frank, E. (2005) *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco.