XX ENMC
ENCONTRO NACIONAL DE MODELAGEM COMPUTACIONAL
VIII ECTM
ENCONTRO DE CIÊNCIA E TECNOLOGIA DE MATERIAIS
16 a 19 de Outubro de 2017
Instituto Politécnico - Universidade do Estado de Rio de Janeiro
Nova Friburgo - RJ

TEXT MINING DESCRIPTIONS OF DREAMS: AESTHETIC AND THERAPEUTIC EFFORTS

Renato Fabbri[1] - renato.fabbri@gmail.com
Fabiane M. Borges[2] - catadores@gmail.com
[1]University of São Paulo, Institute of Mathematical and Computer Sciences - São Carlos, SP, Brazil
[2]Federal University of Rio de Janeiro, School of Fine Arts - Rio de Janeiro, RJ, Brazil

Abstract. *Dreams are highly valued in both Freudian psychoanalysis and less conservative therapeutic traditions. Text mining enables the extraction of meaning from writings in powerful and unexpected ways. In this work, we report methods, uses and results obtained by mining descriptions of dreams. The texts were collected as part of clinical practices from dozens of volunteers. They were subsequently mined using various techniques for the achievement of poems and summaries, which were then used in schizoanalysis (therapeutic) sessions by means of music and declamation. The results were found aesthetically appealing and effective to engage the audience of patients. The expansion of the corpus, mining methods and strategies for using the derivatives for art and therapy are considered for future work.*

Keywords: *Dreams, Text mining, Poetry, Art, Schizoanalysis*

INTRODUCTION

Although dreams are described in texts that range from ancient sacred **?** to recent medical **?**, there is no consensus of what dreams are. We can exemplify the diverse theories with three simple cases:

- Dreams are often regarded by the dreamers as accessing spiritual helms or other realities.
- Many scientists regard dreams as by-products of the sleeping process: arbitrary interpretations given by the conscious mind to noisy signals without substantial meaning.
- Freudian and Jungian psychoanalysis traditions understand dreams as symbolic constructs output by the unconscious mind.

We can, however, state some facts about dreams that make them very interesting material for therapy and for art. First, dreams are often very rich in impacting and symbolic images. Second, they are told by the person who dreamed in a very attentive manner, as being very significant to the dreamer. In fact, most of us should be able to remind of a number of situations where

someone (perhaps ourselves) was describing a dream in a rapid, almost euphoric, succession of words. Dreams are so effective in yielding artistic materials that surrealism is an aesthetic explicitly inspired by dreams and symbolism is an example of artistic movement heavily influenced by dreams.

Text mining is data mining applied to textual data. There are many models for the text mining pipeline, but it can be summarized as: data collection and preparation, pattern recognition, evaluation of the output and reporting. This work addresses text mining of descriptions of dreams with aesthetics and therapeutic purposes.

Section **??** describes the corpus and methods. Section **??** is dedicated to presentation and discussion of results. Section **??** holds conclusions and further work considerations.

MATERIALS AND METHODS

Corpus

The description of dreams we used are **?** all in Brazilian Portuguese, collected as part of clinical practices in the year of 2014. Volunteers described the dreams and sent them to the second author or this paper. Thereafter, another collection of dreams **?** was gathered in the same way by the second author, with the purpose of expanding the analysis and synthesis of texts performed with the previous corpus, but was not used until now. It is a larger corpus, also in Brazilian Portuguese. Interestingly, both corpus contains description of dreams by women only. Both corpus are summarized in Table **??** in numbers of dreams, paragraphs, and letters.

Analysis and derivation methods

The texts were analyzed to support the extraction of meaning from the dreams and for the creation of artistic texts. We strived to keep the methods very simple in order to avoid puzzling the involved parties. We considered three lists of tokens:

- punctuations !”#$%&’()*+,-./:;¡=¿?@[]\^_‘{}—˜. Obtained through the command `string.punctuation` of Python’s string (standard) library.
- Portuguese *stopwords*[1] obtained through NLTK **?** by the command `nltk.corpus.stopwords.words("portuguese")`.
- Tokens in the texts which were not punctuations nor stopwords. These were regarded as the most meaningful words in the text and were used in their order of appearance.

This list of most meaningful words was used as the core material for the achievement of more interesting constructions for art and therapy through filtering and ordering. Most significantly, the ordering could be based on the alphabet, the size of tokens in number of letters, or the count of incidences of the words, or any combination of these. Filtering could be performed by restricting the vowels, consonants (e.g. fricatives), word size, frequency, or collocations.

[1]The exact definition and list of stopwords are not consensual. Anyway, one can regard them as words with lesser meaning and which are very frequent, such as conjunctions and prepositions.

RESULTS AND DISCUSSION

The list of most meaningful words (described in the previous section) was filtered and ordered in many ways to yield diverse sequences of interest. After an inspection of the results, these criteria were selected to compose a final document:

- Ordering by: incidence (most frequent words first), alphabetic, size in characters, with and without repetitions. These were considered the most raw sequences and used subsequently to derive other sequences with such variations of ordering and repetition.
- Words with only one vowel (repeated any number of times).
- Only words with fricatives or plosives or some combination of them (e.g. plosives and m and vowels 'a' and 'e').
- Words that start and end sentences.
- Collocations (pair of words which are frequent).

Such final document and other files are available online and exposed in Table **??**. An example of these texts is in Table 1 with a translation from Portuguese to English. These texts were used for aesthetic appreciation and also in a schizoanalysis group session in 2014, Casa Nuvem, Rio de Janeiro, RJ, Brazil. The artist Giuliano Obici used the texts to make live electroacoustic music, which was accompanied by declamations. The groups was constituted by the volunteers who described the dreams and the description of the episode is somewhat impressive: the members had strong impressions, some of them cried and entered a quasi-shock state.

Table 1: Example of artistic text obtained from description of dreams. This text was obtained through picking only the first and last words of each sentence.

Escorregava glandes	Slipping glands
Numa assustavam	At once, they scared
Eu suada	I sweated
As cavalos	The horses
Não acabou	It's not over
Barras mim	Bars me
Andei construtores	I walked builders
Pessoas)	People)
Sonhei formei	I dreamed I formed
Estava menino	It was boy
Depois boa	Then good
Esse meu	This mine
Sonhos descendo	Dreams coming down
O irmão	The brother
Meu punição	My punishment
Começa irmão	Begins brother
Meu ele	My him
Meu demonstração	My demonstration
Depois ”	After ”
Eu parede	I wall
Sinto dele	I'm fell him
A importência	The “importence”
O buraco	The hole
Acordei ofegante	Woke up breathless
Sensação NÃO	Feeling NO
Já rumo	I'm on my way
Estava perseguido	Was persecuted
Quando percebeu	When realized
A tempo	In time
Até porta	Up to door
O disso	The this
Eu sobreviver	I survive

CONCLUSIONS AND FUTURE WORK

We understand that the results are compelling for both art and therapy. Only the first corpus was used, which is smaller and made easier the selection of the resulting texts. The methods applied are very simple, favoring the communication between the parties, and are promptly deepened and expanded into more complex processes. This work seems unique in the sense of using text mining of dreams for art and therapy, which, in our opinion, benefit the appreciation of it as a multidisciplinary scientific contribution in computer science, art and psychology (schizoanalysis).

In further efforts, we might use for the corpus:

- descriptions of dreams in the literature (e.g. Bible);
- other languages;
- an expansion of current corpus;
- dreams from specific groups, e.g. again gender related or of a specific age span, professional or educational background, etc.

About the text mining methods, we might:

- Use specific routines for classification (e.g. clusterization) of texts or their features.
- Expand the methods of selection of words to better encompass meter (e.g. number of syllables).
- Expand methods of selection of words and phrases by their sonorities: sequence of vowels or consonants, mute consonants, paroxytones, etc.
- Use Wordnet in order to relate terms through semantic links (e.g. hypernymy, meronymy, synonymy).

The exploration of the results in therapeutic sessions and for the achievement of compelling selections of artistic texts should be kept as the core purposes.

Acknowledgements

The authors thank the volunteers who supplied the descriptions of dreams; the subjects with attended to the schizoanalysis sessions; the open source software developers, especially those who enabled this work by developing the Python language and the NLTK.

REFERENCES

Aznar, M., Pessoa, F.L.P. and Silva Telles, A. (1994), "Vapor-Liquid Equilibria of Mixed Solvent-Salt Systems using a MHV2 Model with the Wilson Equation", *X Congresso Brasileiro de Engenharia Química*, São Paulo, vol 1, 38-43.

de Freitas, G.C.S.; Peixoto, F.C.; Vianna Jr, A.S. (2008), Simulation of a thermal battery using Phoenics. Journal of Power Sources, 179, 424-429.

Duarte, C.S.A. (1999), *Equilíbrio Líquido-Líquido em Sistemas contendo Polímero e Eletrólito: Água/ Polietileno-glicol/Fosfato*, Laboratório de Equilíbrio de Fases, FEQ/UNICAMP, Campinas.

Fredenslund A. e Sorensen, J.M. (1993), "Group Contribution Estimation Methods", in *Models for Thermodynamic and Phase Equilibria Calculations*, S.I. Sandler (ed.), Marcel Dekker, Inc., New York.

Silva, L.F. (2005), "*Predição de Pontos Críticos de Misturas Termodinâmicas*", Tese de Doutorado, IPRJ/UERJ, Nova Friburgo.

Van Krevelen, D.W. (1990),"*Properties of Polymers. Their Correlation with Chemical Structure, Their Numerical Estimation and Prediction from Additive Group Contribuition*", 3º ed., Elsevier, Amsterdam.

APPENDIX A

This section, if necessary, must be included here.

For Papers written in Portuguese or Spanish a translation of title, abstract and keywords into English must be provided after the Appendix using the same format for title, abstract and keywords presented in the first page of the template.