Text2voronoi: An Image-driven Approach to Differential Diagnosis

Alexander Mehler
Goethe University
Institute of Computer Science
Text Technology Lab

Tolga Uslu
Goethe University
Institute of Computer Science
Text Technology Lab

Wahed Hemati
Goethe University
Institute of Computer Science
Text Technology Lab

{mehler,uslu,hemati}@em.uni-frankfurt.de

Abstract

Differential diagnosis aims at distinguishing between diseases causing similar symptoms. This is exemplified by epilepsies and dissociative disorders. Recently, it has been shown that linguistic features of physician-patient talks allow for differentiating between these two diseases. Since this method relies on trained linguists, it is not suitable for daily use. In this paper, we introduce a novel approach, called text2voronoi, for utilizing the paradigm of text visualization to reconstruct differential diagnosis as a task of text categorization. In line with current research on linguistic differential diagnosis, we explore linguistic characteristics of physician-patient talks to span our feature space. However, unlike standard approaches to categorization, we do not use linguistic feature spaces directly, but explore visual features derived from the talks' pictorial representations. That is, we provide an approach to *image-driven differential diagnosis*. By example of 24 talks of epileptics and dissociatively disordered patients, we show that our approach outperforms its counterpart based on the bag-of-words model.

1 Introduction

Physicians use medical imaging for diagnosing. Bone fractures, for example, are visualized by radiographs, pregnancies are examined by means of ultrasound scans, while neurological disorders are studied with the help of MRI scans. Our goal is to assist physicians in diagnosing mental disorders by analogy to such image-driven methods. To this end, we introduce a method for scanning physician-patient talks to get pictorial representations as input of classifiers which perform the differential diagnosis. This approach is in line with recent efforts in clinical NLP to utilize computational methods for automatically analyzing medical histories (Friedman et al., 2013). It profits from recent findings showing that linguistic features provide reliable bases for differentiating between epilepsies and dissociative disorders (Gülich, 2010; Reuber et al., 2009; Opp et al., 2015). Since the latter approach relies on trained linguists for performing the feature analysis it does not allow for daily use. The present paper aims at filling this gap. It introduces a new method for visualizing linguistic data by means of images as input to classifiers which learn from their pictorial features to arrive at the desired diagnoses. The main hypothesis of our paper (as elaborated in Section 3) runs as follows: *Linguistic features of physician-patient talks can be visualized in a way that a certain range of diagnoses can be derived from analyzing pictorial features of these visualizations*. In Section 3, we introduce so called <u>Vo</u>ronoi diagrams of <u>Te</u>xts (VoTe) to provide such expressive visualizations. VoTes are generated by our text2voronoi algorithm as described in Section 3. Unlike the classical bag-of-words model, this approach explores bags of visual features derived from the talks' image representations in terms of VoTes. To this end, we utilize the *TextImager* which automatically extracts a wide range of linguistic information from input texts to derive representational images thereof. In Section 4, we describe an experiment, which shows that our image-driven classifier can indeed differentiate between epilepsies and dissociative disorders: its F-score outperforms its classical counterpart based on the bag-of-words model. Note that we do not claim that VoTes allow for differentiating between whatever mental diseases. Rather, we start with epilep-

80

sies and dissociative disorders as two initial examples and will extend our approach by including related diseases in future work (cf. Section 5).

2 Related Work

Recent studies have shown that a linguistic examination of physician-patient talks based on *Conversation Analysis* (CA) (Drew et al., 2001) allows for distinguishing between epileptic and non-epileptic seizures (Reuber et al., 2009; Plug et al., 2009; Plug et al., 2010; Gülich, 2010; Opp et al., 2015). Reuber et al. (2009) describe a CA-inspired experiment where two linguists blinded to medical data attempted to predict the diagnosis on the basis of qualitative linguistic assessments. Using these assessments, the linguists predicted 17 of 20 (85%) diagnoses correctly. Opp et al. (2015) found that patients with epileptic seizures try to describe their attacks as accurate as possible, whereas patients suffering from dissociative disorders avoid detailed descriptions of their seizures. As a matter of fact, such differences are mirrored by linguistic choices. However, these and related methods (Gülich, 2010) rely on the expertise of trained linguists and are, thus, not practical in terms of daily use.

Other approaches use machine learning to predict diagnoses from therapy transcripts by means of extracted linguistic features (Howes et al., 2012a). Howes et al. (2013), for example, use topics that have been derived by means of LDA. Support vector machines operating on linguistic features have also been used to predict diagnoses (Howes et al., 2012b; DeVault et al., 2013; DeVault et al., 2014). Unlike these approaches to text categorization, which rely on the bag-of-words model or some of its descendants, we use pictorial representations of linguistic features as input for our classifier. This is done by extending the UIMA-based TextImager by means of visual scans of physician-patient talks as explained in Section 3. Alternatives to the TextImager are given by the UIMA-based frameworks cTAKES (Savova et al., 2010) and EpiDEA (Cui et al., 2012). Unlike the TextImager, both tools do not provide a visualization engine and, thus, do not fit our task of text classification based on pictorial text representations.

Note that the pictorial representations of texts as introduced here rely on so called Voronoi diagrams (de Berg et al., 2000). Voronoi dia-

Label	POS
C1	Noun
C2	Verb
C3	Preposition
C4	Adjective
C5	Adverb
C6	Temporal expression

Table 1: Parts of speech and expressions explored by text2voronoi.

Label	Category	Example
G1	Case	{nominative, accusative,..}
G2	Mood	{indicative, imperative,..}
G3	Number	{singular, plural}
G4	Person	{first, second,..}
G5	Tense	{past, present,..}
G6	Gender	{feminine, masculine,..}
G7	Degree	{positive, comparative,...}

Table 2: Categories explored by text2voronoi.

grams have already been used to represent semantic structures of lexical units (Jäger, 2006). We further develop this approach in the sense of deriving Voronoi diagrams as representations of natural language texts in general.

3 The text2voronoi Model of Texts

Our goal is to generate images from physician-patient talks whose visual features can be used by classifiers to perform the desired differential diagnosis. To this end, we provide the *text2voronoi algorithm* which computes this visualization in four steps (see Figure 1):

1. extraction of linguistic features,

2. embedding the features in vector space,

3. Voronoi tesselation of this space and

4. extraction of visual features from the t004 tesselation.

In what follows, we describe each of these steps.

3.1 Linguistic Feature Extraction

Each input text is preprocessed by the TextImager which utilizes several NLP tools to tag a range of linguistic features per lexical token. This includes POS tags (e.g., pronouns, prepositions), grammatical categories (e.g., case, gender, number, tense) and temporal expressions (e.g., dates, temporal adverbs) – see Table 1 and 2 for all POS and their features considered in Step 1 combining to 180
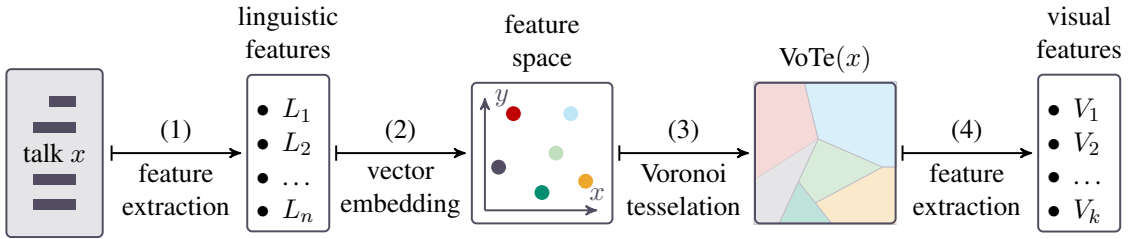
Figure 1: Workflow of the text2voronoi algorithm generating a *Voronoi diagram of* the *Text* (VoTe) x.

(a) Dissociation disorder

(b) Dissociation disorder

(c) Dissociation disorder

(d) Epilepsy
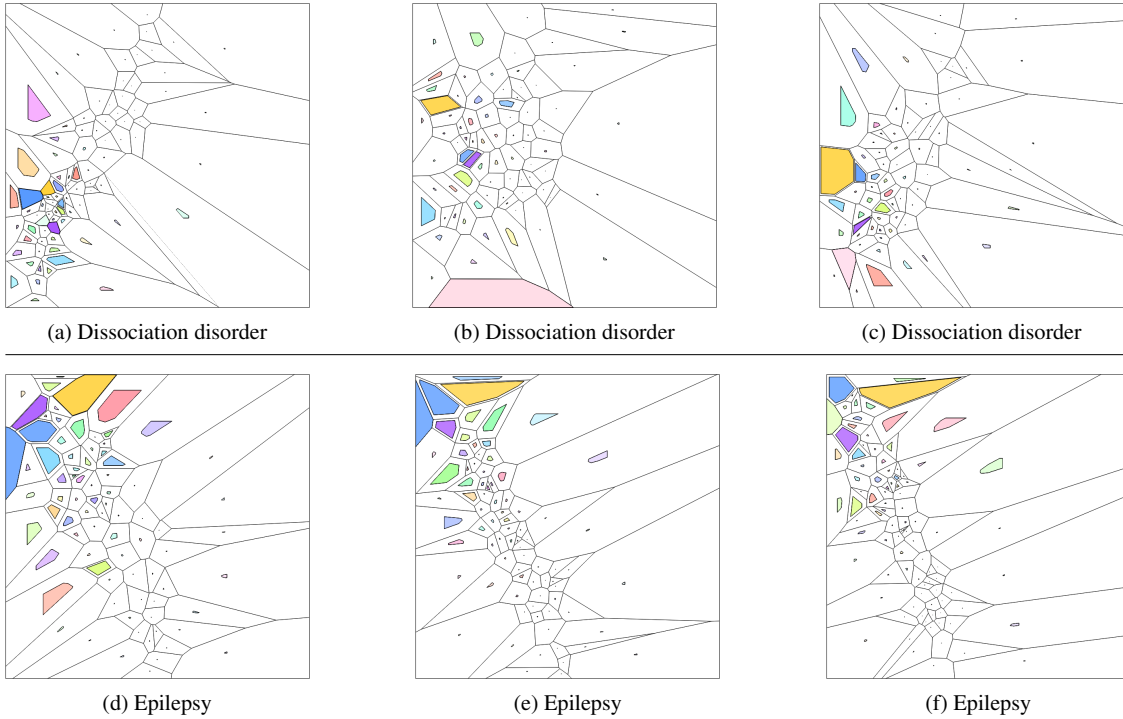
(e) Epilepsy

(f) Epilepsy

Figure 2: Visualizations (*VoTe*s) of six physician-patient talks as used in our classification experiment.

features. The reason for selecting these features is that according to (Gülich, 2010; Opp et al., 2015), patients suffering from epilepsies tend to give detailed descriptions of their seizures, while dissociatively disordered patients tend to avoid such descriptions. Thus, while the former group of patients likely uses personal pronouns in connection with prepositions (for localizing their seizures) and polarity cues (for evaluating them), the latter group will rather avoid the co-selection of such features. For tagging POS and grammatical features, we use a retrained instance (Eger et al., 2016) of *MarMoT* (Müller et al., 2013), while *HeidelTime* (Strötgen and Gertz, 2010) is used for tagging temporal expressions.

3.2 Embedding the Features in Vector Space

Since our features are tagged per token, we can transcode each sentence of the corresponding input text as a sequence of these features to make them as input to word2vec (Mikolov et al., 2013) by projecting on exactly two dimensions. The reason behind this approach is to compute feature associations in a manner that is characteristic of the input text. Thus, we do not use a (huge) reference corpus (e.g., Wikipedia) for computing "reference" associations but explore text-*specific* patterns in our two-dimensional feature space.

3.3 Voronoi Tesselation of the Feature Space

The vector embeddings span a two-dimensional space for which we compute a Voronoi decomposition (de Berg et al., 2000). Each cell of the

resulting *Voronoi diagram of* a *Text* (VoTe) corresponds to a single feature. Generally speaking, starting from a set P of distinct points in a plane we get a corresponding Voronoi diagram by coloring all points $q_1, \ldots$ of equal distance to at least two points in P (de Berg et al., 2000). The points $q_1, \ldots$ manifest the borders of the Voronoi regions that consist of all points with the same single nearest neighbor in P. To color the VoTe of a text, we additionally explore two kinds of frequency information: while the overall frequency of a feature determines how much of its cell is filled (starting from the center), the transparency of the cell depends on the feature's inverse sentence frequency: the smaller this value, the more transparent the cell. Figure 2 exemplifies the VoTes of 6 texts. Note that for each text each feature is mapped onto the same color in order to allow for comparing different texts. However, the exact position of a feature cell in a text's VoTe, its size, degree of filling, transparency and neighborhood depend on the specifics of that text. That is, they depend on the characteristics of the given physician-patient talk in terms of the co-occurrence statistics of the underlying linguistic features. Thus, our classification hypothesis is: *talks of patients suffering from the same disease induce similar VoTes.* Exploring the visual patterns of VoTes is then a way to perform the targeted classification.

3.4 Extracting Visual Features from VoTes

For the sake of the latter classification, we extract a set of visual features for each cell of the VoTes (see Table 3). The underlying hypothesis is that two VoTes are the more similar, the more of their equally colored cells share similar visual features. Each cell is characterized (1) by its *gestalt* (area, corner, filling, shape, transparency), (2) *location* (position, shape) and (3) *neighborhood* (centrality). While the first group of features informs about how a single cell looks like, the second group informs about its localization on the map, and the third group about its relations to other cells. The more of these features are shared by two equally colored cells, the more visually similar they are. For mapping neighborhood-related features, we compute the closeness centralities of the cells in the graph representation of the Voronoi diagrams. Next, for all Voronoi cells that correspond to the 180 features of Step 1, we compute 11 features (see Table 3) so that each VoTe of a

Feature	Description	#Features
Area	The surface area	1
Position	x/y coordinates of center	2
Shape	Min (x, y), max (x, y)	4
Filling	Percentage of fill coverage	1
Transparency	Degree of opacity	1
Corner	Number of corners	1
Centrality	Closeness centrality	1

Table 3: Visual features of the cells of a Voronoi tesselation (VoTe) explored by text2voronoi.

Features	Kernel	nu-SVC	C-SVC	SVM light
All	Linear	0.832	0.832	0.832
Subset	Linear	1.0	1.0	1.0
All	RBF	0.832	0.832	0.832
Subset	RBF	0.958	1.0	1.0

Table 4: F-scores of text2voronoi-based classification.

text is finally mapped onto a vector of 1980 visual features. Note that if a linguistic feature did not occur in a talk, it was mapped onto a null vector so that VoTes get also comparable for commonly absent features.

4 Experiment

This section provides experimental data on testing the text2voronoi model. To this end, we use a German corpus of 24 physician-patient talks of 12 epileptics and 12 dissociatively disordered patients. The talks were transcribed according to GAT2 (Selting et al., 2009) and annotated w.r.t. turns and seizure descriptions (Gülich, 2010; Opp et al., 2015). The corpus was further processed according to Section 3 so that each talk was mapped onto a vector of 1980 visual features. Finally, the vectors were independently made input to *SVM-light* and *LIBSVM* to compute F-scores based on a leave-one-out cross-validation. Using all features, both kernels (linear and RBF) achieve an F-score of 83.2% – see Table 4. Next, we performed an optimal feature selection for SVMs (Nguyen and De la Torre, 2010) using a genetic search on our feature space with the aim of optimizing F-scores based on the same setting of cross-validation. This optimization resulted in a perfect classification (see Table 4) regardless of the kernel and the implementation of SVMs in use. Finally, we computed a bag-of-words model based on the lexical data of all talks in our corpus. Using an RBF kernel (leave-on-out cross-validation) this model

Features	Linear kernel	RBF kernel
All	0.60	0.69
Subset	0.91	0.82

Table 5: F-scores of the bag-of-words model.

achieved an F-score of 69% (see Table 5); a search for an optimal feature subset raised this score to 91% (by means of a linear kernel).

4.1 Discussion

Obviously, our findings are independent of the kernels (linear or RBF) and the SVM implementations in use. They show that by example of our corpus data, differential diagnoses come into reach based on visual depictions of the underlying talks. Moreover, we show that an optimal feature selection for SVMs can boost the classifier enormously. This may hint at problems of overfitting (negative interpretation) or at the expressiveness of the visual features in use (positive interpretation). Evidently, our corpus data is too small to decide between these alternatives. Thus, further research is required that starts from larger corpora of physician-patient talks. As a matter of fact, such data is extremely difficult to obtain (Friedman et al., 2013) so that comparative studies have to be considered in related areas of more easily accessible data. However, as indicated by our F-scores and as exemplified by Figure 2, our VoTe representations of texts are seemingly informative enough to provide visual depictions of text that may be used by physicians as scans of neurologically disordered patients based on their medical histories. Based on our results, we may speak of a novel approach to text representation according to which symbolically coded information in texts is visually reconstructed in a way that allows for performing text operations (in our case *text classification*) indirectly by processing the resulting visual representations.

5 Conclusion

We presented a novel approach to image-driven text classification based on Voronoi tesselations of linguistic features spaces. Our method allows for high score differential diagnoses by exploring features of the pictorial representations of physician-patient talks. Our experiments show that this approach outperforms classifiers based on the bag-of-words models. In order to further test its va-

lidity, we plan to experiment with larger corpora and various tasks in text classification (e.g., authorship attribution and genre detection). A major reason to do this is to clarify whether the F-scores reached by our approach so far reflect overfitting or not. To this end, we will also experiment with data of different languages. Moreover, since a great deal of information about the correct diagnosis relates to whether a patient tends to suppress the memory of her or his seizures, polarity cues are promising candidates for extending our feature space. However, since we deal with *seizure* descriptions, such a distinction is rather challenging. The reason is that turns of patients about seizures have very likely negative connotations. An alternative is to consider simpler quantitative features (turn length, number of turns etc.) to simplify the generation of VoTes. This is needed to enable automatic differential diagnoses instantaneously during physician-patient talks, which – because of error-prone speech recognition systems – require easy to measure features. Obviously, this requirement implies a trade-off: the more easily a feature is measured, the lower its semantic specificity with respect to the target classes to be learnt. Thus, a great deal of progress may be expected by developing speech recognition systems that focus on expressive linguistic features especially of physician-patient talks. Last but not least, we may consider quantitative characteristics that are more closely related to the geometry of Voronoi diagrams (e.g., in terms of their order and size – cf. (de Berg et al., 2000)). In this way, we want to contribute to the further development of text representation models based on text visualizations.

References

Licong Cui, Samden D Lhatoo, Guo-Qiang Zhang, Satya Sanket Sahoo, and Alireza Bozorgi. 2012. EpiDEA: extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification. In *AMIA Annu Symp Proc.*, pages 1191–1200.

Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. 2000. *Computational Geometry*. Springer, Berlin/Heidelberg.

David DeVault, Kallirroi Georgila, Ron Artstein, Fabrizio Morbini, David Traum, Stefan Scherer, Albert Rizzo, and Louis-Philippe Morency. 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *Proc. of SIGDIAL*.

David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proc. of the 2014 Int. Conf. on Autonomous Agents and Multi-agent Systems (AA-MAS '14)*, pages 1061–1068.

Paul Drew, John Chatwin, and Sarah Collins. 2001. Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expectations*, 4(1):58–70.

Steffen Eger, Rüdiger Gleim, and Alexander Mehler. 2016. Lemmatization and morphological tagging in German and Latin: A comparison and a survey of the state-of-the-art. In *Proc. of LREC 2016*.

Carol Friedman, Thomas C. Rindflesch, and Milton Corn. 2013. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the national library of medicine. *Journal of Biomedical Informatics*, 46(5):765–773.

Elisabeth Gülich. 2010. Le rôle du corpus dans l'élaboration pluridisciplinaire d'un instrument de diagnostic linguistique: l'exemple de l'épilepsie. *Pratiques. Linguistique, littérature, didactique*, (147-148):173–197.

Christine Howes, Matt Purver, Rose McCabe, Patrick GT Healey, and Mary Lavelle. 2012a. Helping the medicine go down: Repair and adherence in patient-clinician dialogues. In *Proc. of the 16th SemDial Workshop on the Semantics and Pragmatics of Dialogue (SeineDial)*.

Christine Howes, Matthew Purver, Rose McCabe, Patrick GT Healey, and Mary Lavelle. 2012b. Predicting adherence to treatment for schizophrenia from dialogue transcripts. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 79–83.

Christine Howes, Matthew Purver, and Rose McCabe. 2013. Using conversation topics for predicting therapy outcomes in schizophrenia. *Biomedical informatics insights*, 6(Suppl 1):39.

Gerhard Jäger. 2006. Convex meanings and evolutionary stability. In Angelo Cangelosi, Andrew D. M. Smith, and Kenny Smith, editors, *The Evolution of Language. Proc. of the 6th International Conference (EVOLANG6)*, pages 139–144, Rome.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119.

Thomas Müller, Helmut Schmid, and Hinrich Schütze. 2013. Efficient higher-order CRFs for morphological tagging. In *Proc. of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 322–332.

Minh Hoai Nguyen and Fernando De la Torre. 2010. Optimal feature selection for support vector machines. *Pattern recognition*, 43(3):584–591.

Joachim Opp, Barbara Job, and Heike Knerich. 2015. Linguistische Analyse von Anfallsschilderungen zur Unterscheidung epileptischer und dissoziativer Anfälle. *Neuropädiatrie in Klinik und Praxis*, 14(1).

Leendert Plug, Basil Sharrack, and Markus Reuber. 2009. Seizure metaphors differ in patients accounts of epileptic and psychogenic nonepileptic seizures. *Epilepsia*, 50(5):994–1000.

Leendert Plug, Basil Sharrack, and Markus Reuber. 2010. Seizure, fit or attack? The use of diagnostic labels by patients with epileptic or non-epileptic seizures. *Applied Linguistics*, 31(1):94–114.

Markus Reuber, Chiara Monzoni, Basil Sharrack, and Leendert Plug. 2009. Using interactional and linguistic analysis to distinguish between epileptic and psychogenic nonepileptic seizures: A prospective, blinded multirater study. *Epilepsy and Behavior*, 16(1):139–144.

Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. 2010. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513.

Margret Selting, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, and Susanne Uhmann. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion*, 10:353–402.

Jannik Strötgen and Michael Gertz. 2010. Heideltime: High quality rule-based extraction and normalization of temporal expressions. In *Proc. of the 5th International Workshop on Semantic Evaluation*, pages 321–324.