A Hierarchical Ontology for Dialogue Acts in Psychiatric Interviews

Conference Paper · June 2021

DOI: 10.1145/3453892.3461349

6 authors, including:

Maria Trigka
University of Patras

Dimitris Kosmopoulos
University of Patras

A Hierarchical Ontology for Dialogue Acts in Psychiatric Interviews

Aristeidis Bifis
University of Patras
Patras, Greece
bifis@ceid.upatras.gr

Maria Trigka
University of Patras
Patras, Greece
trigka@ceid.upatras.gr

Sofia Dedegkika
Health Center of Chaidari, 2nd YPE
Athens, Greece
sdedegkika13@yahoo.gr

Panagiota Goula
University of Patras
Patras, Greece
ygoula@upatras.gr

Constantinos Constantinopoulos
University of Patras
Patras, Greece
kkonstantino@upatras.gr

Dimitrios Kosmopoulos
University of Patras
Patras, Greece
dkosmo@upatras.gr

ABSTRACT

We present our work on modeling the context of an interview during diagnostic sessions for patients with mental health problems. The results are to be exploited by translation system for telehealth services. More specifically, we plan to use the context of the psychiatric interview in order to set informative priors over the vocabulary of the speaker. Therefore we have modelled the context with a hierarchical ontology, and we use it to classify the current state of the interview. The state is extracted after the doctor asks a question, and allow us to select a non-uniform prior regarding the vocabulary of the patient.

CCS CONCEPTS

• **Computing methodologies → Discourse, dialogue and pragmatics**; *Supervised learning by classification*; *Instance-based learning*; Machine translation.

KEYWORDS

dialogue context, stress, depression, hierarchical classification

ACM Reference Format:
Aristeidis Bifis, Maria Trigka, Sofia Dedegkika, Panagiota Goula, Constantinos Constantinopoulos, and Dimitrios Kosmopoulos. 2021. A Hierarchical Ontology for Dialogue Acts in Psychiatric Interviews. In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29-July 2, 2021, Corfu, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3453892.3461349

PETRA 2021, June 29-July 2, 2021, Corfu, Greece

ACM ISBN 978-1-4503-8792-7/21/06. . . $15.00
https://doi.org/10.1145/3453892.3461349

1 INTRODUCTION

Anxiety disorders, stress and depression are quite common in the general population. They are associated to the modern way of life and often cause significant reduction in the individuals' functionality and typically result into notable burdens on health systems. The close relationship and coexistence of anxiety and depressive disorders with physical ailments (either as a cause or as a consequence) are noteworthy. Thus, the individual's ability to access mental health services and the provision of appropriate psychiatric treatment, are crucial factors in the control and prognosis of anxiety and depressive disorders. A prerequisite for the proper treatment of each individual, is the collection of a detailed record of the patient through the psychiatric interview, which will lead the specialist to the diagnosis and then to the selection of the appropriate treatment [2, 6, 8].

In this paper we aim to model the context of a psychiatric interview, in order to assist the translation of the patients' responses within a translation system (e.g., sign language) the system itself will not be presented here. We will present the context modeling part, based on data from realistic scenarios. We are not aware of any similar research focusing on psychiatric interviews.

Our motivation is to create a system that is able to do interpretations in real time and facilitate doctor-patient communication. In order to achieve this goal, we model the context of the aforementioned dialogues based on a hierarchy of Dialogue Acts (DA) [3], we predict the expected vocabulary of the patient's response, and optimize the SL-to-text translation process. The task of SL-to-text translation is very challenging, and typically requires computations over large vocabularies. Our approach aims to increase the quality of the translation by assigning more weight on certain parts over the vocabulary, given the current context of the interview.

In this paper, we present the main parts of the doctor-patient dialogue modeling and prediction of the patient vocabulary based on that model. We intend to use this process in order to guide the vocabulary retrieval by enforcing a prior on the vocabulary items. The prior is the result of this paper.

In the first step we compiled a corpus of realistic scenarios for psychiatric interviews. We analyzed them and came up with an ontology, which we exploit in later phases, as part of our classification and vocabulary configuration schemes.

Next we created our classification scheme, based on the previous ontology. The purpose of the classifier is to help us predict the expected vocabulary of the patient's response to doctor queries. Our classifier accepts as input a doctor's query and predicts the class that the query belongs to, and in extend, the prior probability density function (pdf) of the vocabulary of patient's response, given that class. Based solely on the classifier's output, we defined the vocabulary prior, and merged some of the initial classes that could not be easily separated.

The paper is structured as follows. In the next section we discuss the diagnosis methodology, used by psychiatrists. In section 3 we present the context modeling. In section 4 we present the context classification. In section 5 we present the experimental results of the classification. I section 5 we describe the resulting vocabulary distributions, and in section 6 we discuss our conclusions.

2 THE DIAGNOSIS METHODOLOGY

The techniques of the psychiatric interview are adapted each time to the particularities of each patient and his/her psychopathology [2, 6, 8].

These difficulties sparked the idea to use ICT technologies to address the specific needs of patients. In this context, an attempt was made to develop a standardized form of psychiatric interview in order to collect adequately information and provide targeted psychiatric support immediately.

The psychiatrist's tools for the diagnostic process of anxiety disorders and depression include the clinical assessment of the patient, the psychometric scales, while the cornerstone is the psychiatric interview [2, 6, 8].

The proposed system uses psychiatrist-patient dialogues from common cases of patients with anxiety disorders and/or depression, based on the clinical experience of an outpatient clinic. Furthermore, considering the DSM diagnostic criteria for these disorders (DSM: American Psychiatric Association Diagnostic System) [8], an attempt was made to explore using the dialogues, through the psychiatrist's questions, the patient's identity, current mental state and history, as well as the patient's pathological and other psychiatric history [2, 6, 8].

The system's goal is to facilitate the communication by doing translating (e.g., from Sign Language or other languages). Based on this possibility, the psychiatrist can receive information that is crucial according to the international diagnostic criteria (DSM) and give a more accurate diagnosis and further support to the patient.

3 DIALOGUE CONTEXT

The task of assigning context to the parts of a dialogue is known as Dialogue Acts (DA) classification, e.g. [5, 10, 11]. The dialogue acts are labels which denote the act that the speaker is performing, e.g. asking some question, refusing a statement or giving a directive. Considering the structure of dialogues, we assume a set $\mathcal{C}$ of N dialogues, i.e. $\mathcal{C} = \{C_1, C_2, \ldots, C_N\}$, where each dialogue C_i consists of a sequence of N_i utterances $C_i = \{x_1, x_2, \ldots, x_{N_i}\}$. The utterances are the actual sentences that exchange the interlocutors. Additionally, there is a set of M dialogue acts $Y = \{y_1, y_2, \ldots, y_M\}$, where y_j is a nominal label describing the context of each sentence. The goal of the DA classification is to assign a label to each utterance.

An example of DA classification appears in Table 1, where we show an annotated excerpt form our corpus of dialogues. A psychiatric interview has a rather strict structure, because it is guided by the questions of the psychiatrist. Therefore we restrict our focus on them, assuming that the DA of the patient's response is determined by the preceding question. Thus we avoid classifying explicitly the parts of patient's speech.

We were interested in describing the context of an interview in more detail than the typical cases in literature. Thus a set of generic DAs, that could be used for everyday dialogues, wouldn't suffice. After the careful examination of a corpus with interview scripts, we propose a hierarchical ontology for the DAs, the one that is depicted in Figure 1. We believe that this is the best way to describe the innate structure of the interview.

The proposed ontology is a Directed Acyclic Graph, with *stress & depression* at its root[1]. The children DAs correspond to the main sections of each interview: *opening*, *probing* and *closing*. The *probing* in branching out to: *purpose of visit*, *psychiatric record*, *non-psychiatric record*, *social life record* and *family record*. The fully expanded graph has 30 terminal nodes, that correspond to the most detailed DAs (see Table 2 for a list).

After the completion of the DA graph, we assigned one leaf label to each question of the psychiatrist in our corpus. In the following section we discuss how we tackled the task of classifying new questions. Then we demonstrate that we can exploit the predicted DA, in order to restrict the expected vocabulary of the patient's answer. The key idea can be grasped better with an example: if the query is of type *symptoms*, then it is more probable that the response contains words like *fear* and *insomnia*, instead of *wife* or *office*.

4 DIALOGUE ACTS CLASSIFICATION

In this section, the methodology for classification and evaluation is discussed. In particular, our methodology consists of four stages: data collection, data pre-processing, feature extraction, classification, and evaluation. These stages will be explained below.

[1] Let us note that we knowingly abuse the nomenclature by using terms appropriate only for the nodes of a Tree. But we hope that the plain topological ordering of the proposed graph protects us from misinterpretations.

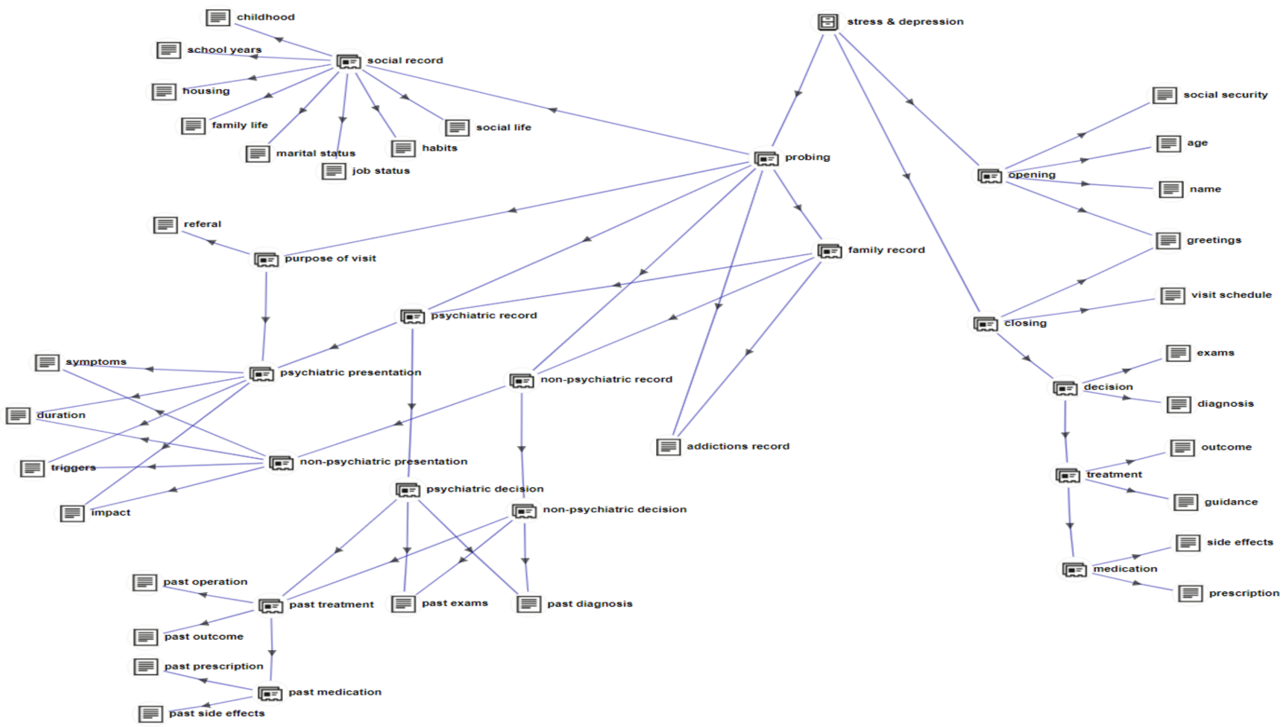

Figure 1: The proposed hierarchical ontology for labeling the parts of a psychiatric interview.

Table 1: An example of annotated interview between doctor (D) and patient (P). The original dialogue is in Greek, and it has been translated by software, for illustrative purposes.

Speaker	Dialogue Act	Utterance (original in Greek)	Utterance (translation)
D	symptoms	Πώς είναι ο ύπνος σας;	How is your sleep?
P		Τώρα με το χάπι είναι καλός.	Now with the pill it is good.
P		Ξυπνάω ξεκούραστη.	I wake up relaxed.
P		Πριν όμως να πάρω το χάπι, ξυπνούσα πολλές φορές μέσα στη νύχτα.	But before I took the pill, I woke up several times during the night.
D	past diagnosis	Προβλήματα υγείας γνωστά υπάρχουν;	Are there known health problems?
P		Μόνο χοληστερίνη έχω ανεβασμένη.	I only have high cholesterol.
P		Παίρνω φάρμακο.	I take a medicine.
D	past diagnosis	Γνωρίζετε αν συγγενείς σας πρώτου βαθμού είχαν προβλήματα με το άγχος ή με άλλες ψυχικές παθήσεις;	Do you know if your first-degree relatives had problems with stress or other mental illnesses?
P		Μόνο η μητέρα μου ήταν αγχώδης ακριβώς σαν κι εμένα.	Only my mother was anxious just like me.

4.1 Data Preprocessing

The main preprocessing steps on the available interviews are the following. First, we organized the sentences (namely, the utterances u_i) of all scripts into two types of DAs, i.e., doctor queries and patient responses. In addition, we simplified the complex structure of the patients' responses, by converting the lengthy sentences into shorter ones. All sentences were originally recorded in Greek language, and then translated in English using machine translation software.

Then, we annotated all sentences (both the queries and the responses), with a label that best describes the context of the corresponding DA. The labels are the ones shown in the interview graph, in Fig 1. By annotating all the query-response DAs, several groups of sentences for each DA are derived. Such knowledge will give us an insight on the per class vocabulary prior and will be exploited in the SL-to-text translation process.

4.2 Feature Extraction

Following the preprocessing step, the doctor's sentences were transformed, in order for them to fall in line with the classification goal. All the sentences have a varying length of words and domain specific terms which may be used more or less frequently.

A first approach would be to use a a term frequency–inverse document frequency ($tf - idf$) feature vector representation for each sentence [4]. $tf - idf$ is the formal measure of how concentrated into relatively few documents are the occurrences of a given word and is computed as follows. At first we calculate the term frequency of term i in sentence j (f_{ij}), normalized by dividing it by the maximum number of occurrences of any term ($max_k f_{kj}$) in the same sentence.

$$TF_{ij} = \frac{f_{ij}}{max_k f_{kj}}$$

The inverse document frequency for a term is defined as follows. Suppose term i appears in n_i of the N available sentences. Then $IDF_i = \log_2 \frac{N}{n_i}$. The ($tf - idf$) score for term i in sentence j is then defined to be $TF_{ij} \times IDF_i$.

Following this scheme in the specific corpus, we would fall short in terms of the classification accuracy, due to the rather small size of the dataset, the length of the sentences and the peculiarity of the vocabulary. Instead, we exploited the power of pretrained neural networks in the task of the sentence representation. We used Sentence-BERT (SBERT), a modification of the pretrained BERT network [1], that uses siamese and triplet network structures to derive semantically meaningful sentence embeddings [7], and specifically, the stsb-bert-base model from the SentenceTransformers framework based on PyTorch and Transformers. With the help of that model, we were able to represent each varying-length sentence to a vector of size 768.

4.3 Classification

The classification associates a particular sentence to one or more classes, among a set of predefined classes according to its attributes. We experimented with flat and hierarchical classification schemes.

In the flat classification scheme, there is no hierarchy. First, each sentence of the training dataset is represented as an SBERT feature vector. For a new query, we find its SBERT representation and try to classify the resulting vector to one of the labels we have at our disposal. Due to the limited number of samples for training, we retreat to the k-Nearest Neighbors classifier. In this approach there is no utilization whatsoever of the interview structure.

Hierarchical classification [9], on the other hand, organizes the classes into levels, creating DAG (Directed Acyclic Graph) of categories, exploiting the information on relationships among them. In the context of this study, the DAG structure is considered and the classes of each level are presented in Figure 1 and explained in the following:

- 1st Level classes: Opening, Probing, Closing
- 2nd Level classes: Social record, Purpose of visit, Family record, Decision
- 3rd Level classes: Psychiatric record, Non-psychiatric record, Addictions record, Treatment
- 4th Level classes: Psychiatric presentation, Non- psychiatric presentation, Psychiatric decision, Non-psychiatric decision, Medication
- 5th Level classes: Past treatment, Past exams
- 6th Level classes (leaf nodes): Greetings, Name, Age, Social security, School years, Housing, Family life, Marital status, Job status, Habits, Social life, Referral, Symptoms, Duration, Trigger, Impact, Past treatment, Past exams, Past diagnosis, Addictions record, Visit schedule, Exams, Diagnosis, Outcome, Guidance, Side effects, Prescription.

We trained one classifier per class following a top-down approach, where a given decision led us down a different classification path. To better understand how hierarchical classification operates, it is necessary to think of a hierarchical classifier as a tree. In this tree, every node, except from the leaves, is a standalone classifier, which classifies a query to one of its child nodes. To train each node we need to split the training data into sets based on the node's children. From all the training data, each time, we select the subset that contains all the sentences that belong to class-labels (leaves) that are reachable from the particular node-classifier. From the resulting subset, we additionally generate one subset per child node, that contains all the sentences that belong to leaves that are reachable from the particular child node. By acquiring those subsets, we can train each classifier, following the flat classification approach. Doing this for all the nodes, results in a system that can classify hierarchically a query. Each query starts at the root and follows a classification trail on our tree all the way down to a leaf. The classification process of a new sentence-query is the following:

- The new query starts from the root.
- Calculate the given query's distance from all the root's children nodes based on the generated subsets.
- Decide, with a confidence rate, to which child node the query is classified.
- The selected child node is considered as the new root, and the child node's subtree is considered as the new tree.
- We repeat the above steps until we end up on a leaf node.

Although a node may be reached from different paths, the goal of the classifier is to output the correct class label irrespective of the path followed, in the specific problem.

5 EXPERIMENTAL RESULTS

In this section, the classification techniques, described in the previous sections, are evaluated in terms of accuracy in the derived dataset. Although the dataset consists of both doctor's and patient' sentences, as discussed above, here, we focus on doctor sentences. To avoid the classification accuracy being biased, we considered only the unique sentences in the classification schemes. Indeed, many of the doctor's queries are repeated among the conversation scripts with an occurrence of more than one. Specifically, the dataset comprises 430 unique doctor sentences distributed into 25 classes. Since the training data size is relatively small and each class has a varying non-balanced number of sentences, we adopted a Leave-One-Out Cross Validation (LOOCV) strategy.

The chance level estimated at 27% since the class *symptoms*, which is the largest class in the dataset, contains 116 out of 430 unique sentences (27% of our dataset). Both of the classification schemes are way above chance level. The flat-classification scheme achieves an accuracy of 54.4%, while the accuracy of the hierarchical classification amounts to 60.9%. The experimental results reveal the superiority, by 6.5%, of the hierarchical classification against the flat one. However, the fact is that there are sentences with similar vocabulary, and the classifiers misclassify them to neighboring classes, as illustrated by the Confusion Matrices in Figures 2, 3. Thus their accuracy is bounded to the above values. Nonetheless, in the hierarchical scheme, these errors indicate us the classes whose vocabulary can be merged contributing to better vocabulary modelling. This aspect is discussed in the following section.

6 THE DISTRIBUTION OF THE PATIENT VOCABULARY

In our bilingual corpus, the Greek vocabulary V of the patient contains $N_V = 3211$ words. For each DA y which is a terminal node of the graph, there is a subset of words which appear in that specific context. These words form a conditional vocabulary V_y with respect to y. The original size of each V_y appears in the second column of Table 2.

A naive model for the prior over the words would be to define a uniform conditional distribution. Namely, for a given DA y, all the words belonging to V_y would have the same probability mass, and the rest of the words nil (or a very small value $\epsilon > 0$). However this is not optimal, as we can't always predict the DA of a sentence correctly.

Therefore, we examine the most common errors that the hierarchical classifier makes. These errors are summarized in Table 3, where we present the wrong predictions and the corresponding true DAs. Focusing on the interviews' ontology and comparing it with the occurred errors, we observe that six DAs share a common path in hierarchy and the classification error occurs at the bottom of the hierarchy (leaf nodes). Since our goal is to model the vocabulary of the patient, and not to create a perfect classifier for the doctor queries, we choose to expand each conditional vocabulary in the way the hierarchical classifier dictates.

As shown in Figure 1, these pairs of DAs share common paths up to a specific level. For example, the pair of *symptoms* and *triggers* share the same paths on the hierarchy until the *psychiatric presentation* and *non psychiatric presentation* nodes which notably share exactly the same vocabulary. A straightforward solution to optimize the process of the dialogue modeling, is to expand some conditional vocabularies with the vocabulary of their parent DAs. For example,

Table 2: The size of patient vocabulary given a DA, counting the unique words after removing stopwords

dialog act	**conditional vocabulary**	
	original size	**expanded size**
addictions record	69	69
age	26	26
childhood	31	1674
diagnosis	22	139
duration	147	1774
exams	3	139
family life	622	1674
greetings	25	139
guidance	32	139
habits	1109	1674
housing	63	1674
impact	773	1774
job status	84	1674
marital status	34	1674
name	17	17
outcome	12	139
past diagnosis	281	629
past exams	170	629
past operation	18	629
past outcome	156	629
past prescription	137	629
past side effects	80	629
prescription	53	139
referral	68	68
school years	22	1674
side effects	28	139
social life	251	1674
symptoms	1122	1774
triggers	460	1774
visit schedule	26	139
full vocabulary	3211	

Table 3: The most common cases of errors. The third row shows the percentage of sentences that were misclassified, over the complete dataset.

true DA	**predicted DA**	**error impact**
symptoms	triggers	3%
prescription	visit schedule	1.9%
symptoms	duration	1.6%
prescription	diagnosis	1.2%
symptoms	past diagnosis	1.2%
past diagnosis	past prescription	1.2%
habits	social life	1.2%

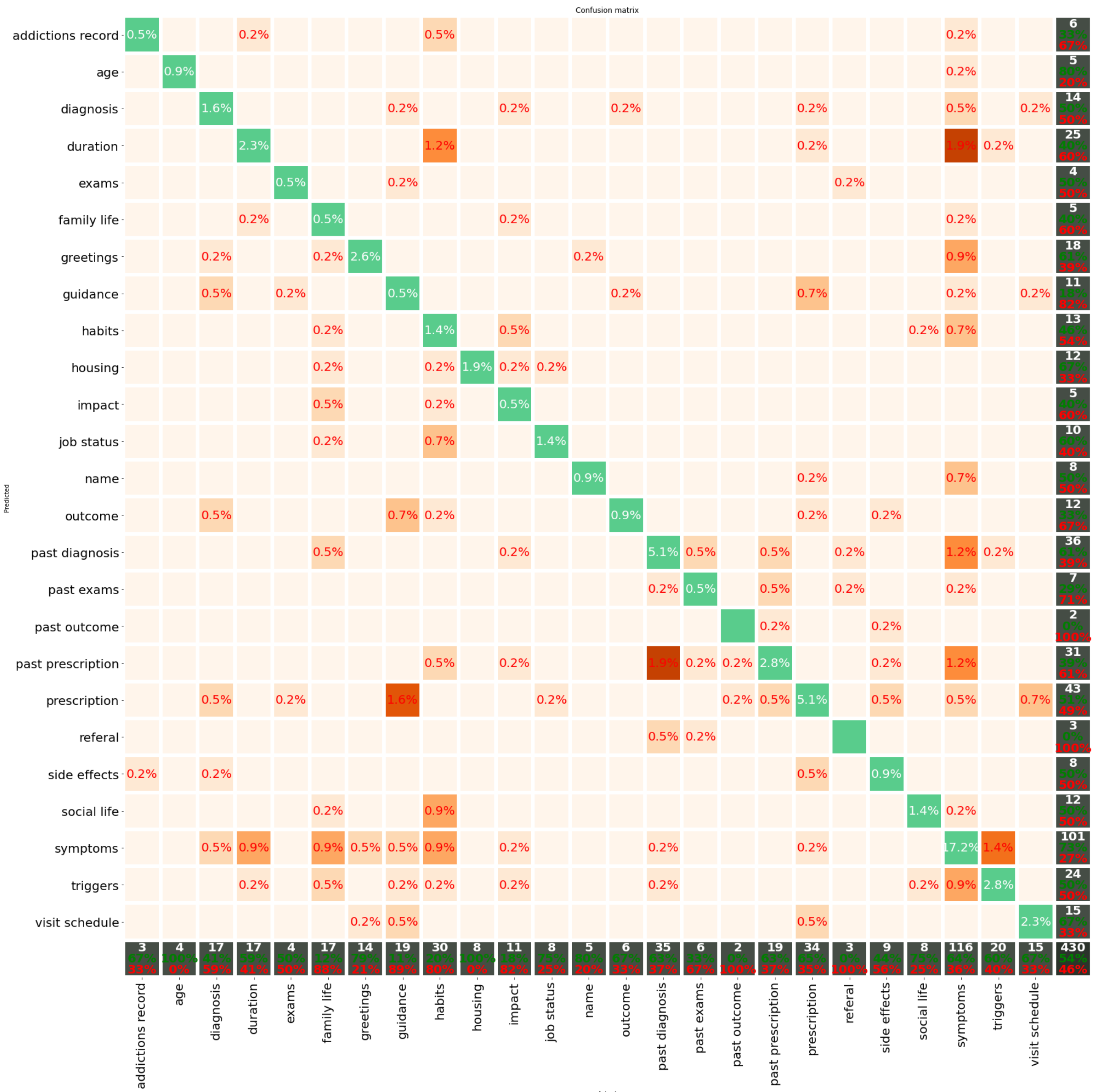

Figure 2: Confusion Matrix of the flat classifier

let's consider the nodes *psychiatric presentation* and *non-psychiatric presentation*. The child leaves are: *symptoms*, *duration*, *triggers* and *impact* DAs. According to the proposed approach, the new vocabulary for each DA is the vocabulary of its parent nodes (the combined vocabulary of the child nodes). In this case, when the true DA of a sentence is *symptoms*, but it is classified as *triggers*, we can still model the dialogue correctly, since the expanded vocabulary of the *triggers* node, contains the vocabulary of *symptoms*.

The nodes that pass their vocabularies to their child nodes are: *'psychiatric presentation'* & *'non psychiatric presentation'*, *'closing'*, *'psychiatric decision'* & *'non psychiatric decision'* and *'social record'*. The size of the expanded conditional vocabularies appear at the third column of Table 2.

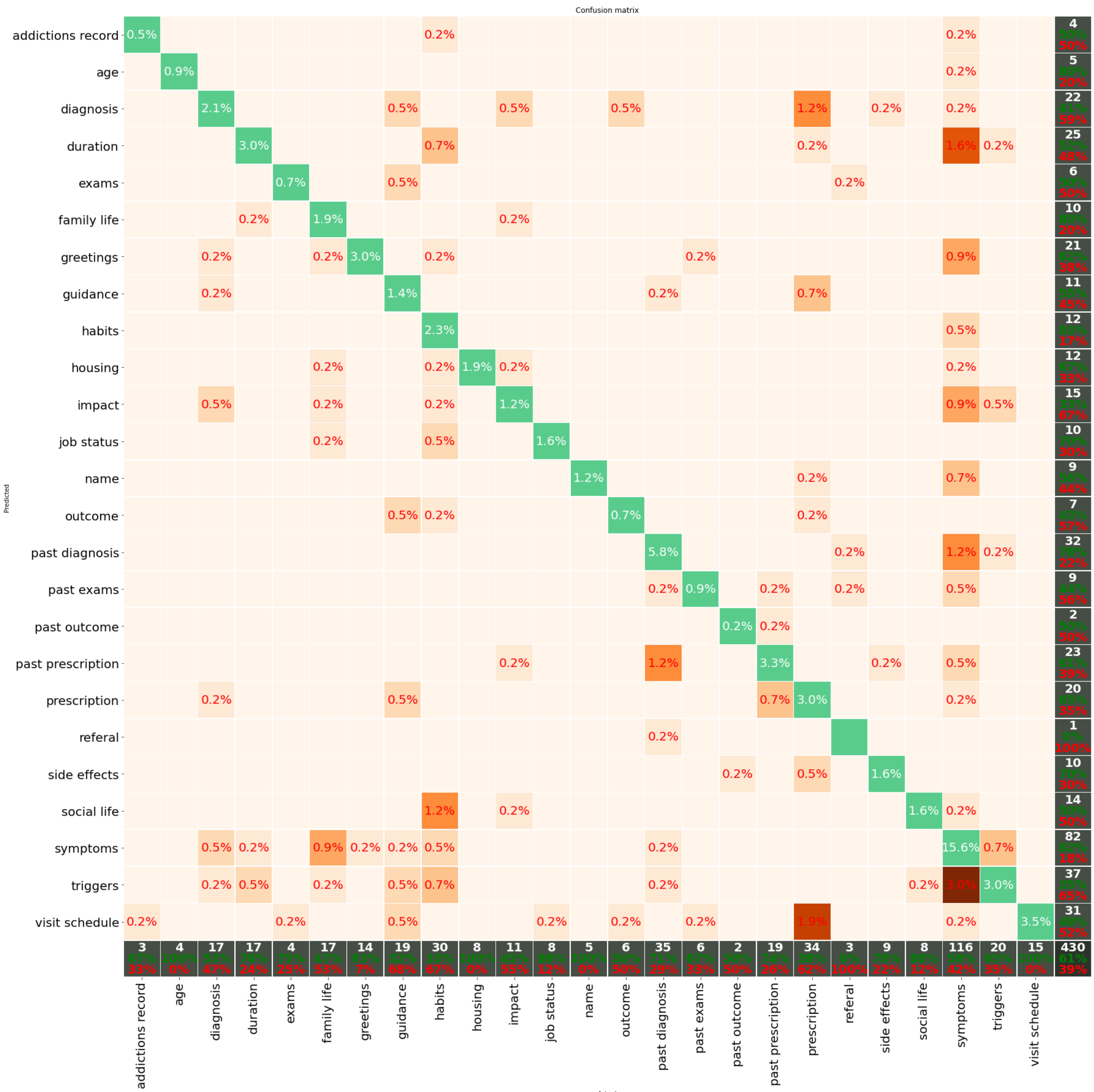

Figure 3: Confusion Matrix of the hierarchical classifier

7 CONCLUSIONS

We have presented our work on modeling the context during diagnostic sessions for patients with mental health problems. We examined a classification scheme using a flat and a hierarchical structure and demonstrate the superiority of the latter method. We demonstrated that the vocabulary of the patient's answer can be segmented, and the relevant subset can be predicted by the doctor's question. Additionally, we defined a conditional prior over all words, given that prediction.

ACKNOWLEDGMENTS

This work was supported by the $T1E\Delta K$-01299 HealthSign project, which is implemented within the framework of "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) Operational Programme 2014-2020, funded by the EU and national funds (www.healthsign.gr).

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT 2019*. 4171–4186.

[2] Michael Gelder, Nancy Andreasen, Juan Lopez-Ibor, and John Geddes (Eds.). 2012. *New Oxford Textbook of Psychiatry.* Oxford University Press.

[3] Harshit Kumar, Arvind Agarwal, Riddhiman Dasgupta, and Sachindra Joshi. 2018. Dialogue act sequence labeling using hierarchical encoder with CRF. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, Vol. 32. AAAI Press, 3440–3447. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16706

[4] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. 2014. *Mining of Massive Datasets* (2nd ed.). Cambridge University Press, USA.

[5] Yang Liu, Kun Han, Zhao Tan, and Yun Lei. 2017. Using context information for dialog act classification in DNN framework. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 2170–2178.

[6] G. N. Papadimitirou, J. A. Liappas, and E. Likouras. 2013. *Modern Psychiatry.* BETA Medical Publications, Athens.

[7] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/1908.10084

[8] Benjamin J. Sadock, Virginia A. Sadock, and Pedro Ruiz. 2017. *Kaplan & Sadock's comprehensive textbook of psychiatry. 10th Edition.* Wolters Kluwer, Philadelphia.

[9] Carlos N. Silla and Alex A. Freitas. 2011. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery* 22, 1 (01 Jan 2011), 31–72. https://doi.org/10.1007/s10618-010-0175-9

[10] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics* 26, 3 (2000), 339–373.

[11] Jason D Williams, Antoine Raux, and Matthew Henderson. 2016. The dialog state tracking challenge series: A review. *Dialogue & Discourse* 7, 3 (2016), 4–33.