Detecting and tracking ongoing topics in psychotherapeutic conversations

Ilyas Chaoua[1,2], Diego Reforgiato Recupero[1], Sergio Consoli[2], Aki Härmä[2], and Rim Helaoui[2]

[1] University of Cagliari, Mathematics and Computer Science Department, Via Ospedale 72, 09124, Cagliari, Italy
ilyaschaoua@gmail.com, diego.reforgiato@unica.it

[2] Philips Research, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
[name.surname]@philips.com

Abstract. One of the key aspects in a psychotherapeutic conversation is the understanding of topics dynamics driving the dialogue. This may provide insights on the therapeutic strategy adopted by the counselor for the specific patient, providing the opportunity of building up artificial intelligence (AI) based methods for recommending the most appropriate therapy for a new patient. In this paper, we propose a method able to detect and track topics in real-life psychotherapeutic conversations based on Partially Labeled Dirichlet Allocation. Topics detection helps in summarizing the semantic themes used during the therapeutic conversations, and in predicting a specific topic for each talk-turn. The conversation is modeled by means of a distribution of ongoing topics propagating through each talk sequence. Tracking of topics aims at exploring the dynamics of the conversation and at offering insights into the underlying conversation logic and strategy. We present an alternative way to look at face-to-face conversations in conjunction with a new approach that combines topic modeling and transitions matrices to elicit valuable knowledge.

Keywords: Conversational AI; Psychotherapeutic conversations; Topics detection and modeling; Partially Labeled Dirichlet Allocation; Transitions matrices.

1 Introduction

The theoretical and technological advances in several disciplines of linguistics, computer science, and healthcare have made possible the recent investigation of therapeutic conversation analysis as a growing field of research [12]. Computational learning techniques have been leveraged to extract useful information from humans interactions through the identification and exploration of unusual patterns [5].

Therapeutic conversations methods such as Cognitive Behavior Therapy (CBT) refer to a range of therapies that can help treating mental health problems, emotional challenges, and some psychiatric disorders of patients by changing their ways of thinking and behave. Accordingly, these therapeutic methods create a new way of looking at severe psychological issues to help patients to move towards a solution and to gain a better understanding of themselves. The treatment is usually a face-to-face conversation where the counselor interacts directly with the patient to understand his feelings, e.g. confident, anxious, or depressed, as well as the causes of his feelings. During the conversation, counselor and patient create a sequence of spoken sentences, each characterized by a certain topic, creating in this way a thematic structure to the whole therapeutic conversation. By adopting a set of techniques and conversational strategies coming from clinical practice, the counselor aims at solving behavioral and psychological problems of the patient by properly reacting to the patient and redirecting the conversation towards certain thematics.

As a result, investigating and modeling the human-to-human dialogues in this kind of context may serve as a guide for the development of AI-based dialoguing systems able, e.g., at recommending the most appropriate therapeutic strategy to adopt by the practitioner for a new patient [5]. In this context, topic detection and tracking (TDT) has been the point of intensive studies since the beginning of natural language processing (NLP) [14] and artificial intelligence (AI) research. One aim of TDT is to identify the appearance of new topics and following their reappearance and evaluation [11].

In this paper we investigate patient-therapist interactions by modeling the propagation of the topics addressed during a given therapeutic conversation by using Partially Labeled Latent Dirichlet Allocation (PLDA) [19]. Traditional Latent Dirichlet Allocation (LDA) is one of the most popular topic model in the literature and is based on a bag-of-words approach. Since data we will deal are partially labeled, then PLDA is of interest because it is likely to produce better results than classic LDA. PLDA is often very useful for applications with a human in the loop, since induced classes correspond well with human categorization, as represented by the provided label space.

The study has been conducted over a dataset of 1729 real-life transcribed psychotherapeutic conversations, each made of different talk-turns, which we will describe further in Section 4. In our method we first identify the most common topics used within the psychology corpus. The PLDA model takes as input the given conversations and detects significant words for each topic. Secondly, the trained PLDA model is able to determine the potential topic addressed in each talk-turn. The talk-turns flow is then transformed into a sequence of potential topics within each conversation. Finally, the semi-supervised PLDA topic model is evaluated by computing its coherence over the most significant words for each topic. Our final aim is to find the quintessential patterns in therapeutic conversations and to understand the topic switches according to the adopted dialogue strategy and topics propagation dynamics. In our method we distinguish the topic changes driven by the counselor and the ones prompted by the patient, and two topic transition matrices are constructed accordingly. These matrices are able to characterize the conversation and to provide important hints towards the understanding of the topics propagation dynamics.

The remainder of the paper is organized as follows. In Section 2 we discuss the related works in the literature on automatic topic detection methods and therapeutic dialogue analysis, that help to establish the basis for the present work. In Section 3 we describe how topic modeling algorithms work, specifying their main characteristics. Afterwards we describe the data used for the experiments and the followed preprocessing steps in Section 4. Section 5 shows our approach on TDT by illustrating the developed methodology. Lastly, Section 6 discusses how to evaluate our model, while Section 7 ends the paper with conclusions and directions for future research.

2 Related Work

Current trends in therapeutic conversations research focus on the digitalization of spoken interactions and on the recommendation of the most appropriate treatments, employing computational techniques such as NLP which offer the potential to extract knowledge from consultation transcripts. Authors in [3] combined robust communication theory used in healthcare and a visualization text analytic technique called *Discursis*, to analyze the conversational behavior in consultations. Discursis [3] is a visual text analytic tool for analyzing human communications, that automatically builds an inherent language model from a given transcribed conversation, mines its conceptual content for each talk-turns, and creates a visual brief. The resulting report can identify communication patterns present during the discussion, with appropriate results of engagement between interlocutors to understand the conversation dynamics. In medical consultations, the classification of conversations suffers from critical weaknesses, including intense performance requirements, time-consuming, and non-standardized annotating systems. To overcome these shortcomings, authors in [13] built an automated annotating system employing a Labeled LDA [18] model to learn the relationships between a transcribed conversation and its associated annotations. Those annotations refer to the subjects and patient symptoms discussed during the therapeutic conversations. The resulting system automatically identifies and restricts those annotations in separate talk-turns within a given conversation. Differently, contributors in [15] examined the use of a LDA [7] topic model as an automatic annotator tool to explore topics and to predict the therapy outcomes of the conversation. The authors assumed that the automated detection of topics does not aim at predicting the symptoms, but it can be used to predict some essential factors such as patient satisfaction and ratings of therapy quality. The examinations from both approaches show that identification and tracking of topics can give useful information for clinicians, enabling them to assist better the identification of patients who may be at risk of loss of treatment. Analyzing human communications, the authors in [2] converted transcribed conversations to time series by developing a discourse visualization system, a text analysis model and a set of quantitative metrics to identify significant features. The system was able to understand the topics used by specific participants, and to generate reports within a single conversation. The method can be used to observe the structure and patterns of interactions and to reconstruct the dynamics of the communication, including the consistency levels of topics discussed between participants and the timing of topic changes. Contributors in [22] proposed a conceptual dynamic latent Dirichlet allocation (CDLDA) model for TDT in conversational text content. Differently to the traditional LDA model, which detects topics only through a bag-of-words technique, CDLDA considers essential information including speech acts, semantic concepts, and hypernym definitions in E-HowNet [4] [10]. It basically extracts the dependencies between speech acts and topics, where hypernym infor-

[3] `http://www.discursis.com/`
[4] `http://ckip.iis.sinica.edu.tw/taxonomy`

mation makes the topic structure more complete and extends the abundance of original words. Experimental results revealed that the proposed approach outperforms the conventional Dynamic Topic Models [6], LDA, and support vector machine models, to achieve excellent performance for TDT in conversations. Authors in [1] presented OntoLDA for the task of topic labeling utilizing an ontology-based topic model, along with a graph-based topic labeling method (i.e., the topic labeling method based on the ontological meaning of the concepts included in the discovered topics). The model was able to show each topics as a multinomial distribution of concepts, and each concept as a distribution over words. Comparing to the classical LDA, the model scaled up better the topic coherence score by combining ontological concepts with probabilistic topic models towards a combined framework applied to different kind of text collections. Contributors in [8] showed an approach to improve human-agent dialogs using automatic identification and tracking of dialogue topics, exploiting the basis of contextual knowledge provided by Wikipedia category system. This approach was constructed by mapping the different utterances to Wikipedia articles and by defining their relevant Wikipedia categories as a likely of topics. As a result, the detection method was able to recognize a topic without holding a priori knowledge of its belonging subject category.

3 Topic Modeling

Topic models are a family of probabilistic approaches that aim at discovering latent semantic structures in large documents. Based on the presumption that meanings are relational, they interpret topics or themes within a set of documents originally constructed from a probability distribution over words. As a result, a document is viewed as a combination of topics, while a topic is viewed as a blend of words.

One of the most widely used statistical language modeling for this end is Latent Dirichlet Allocation (LDA) introduced by Blei et al. [7]. LDA is a generative approach. It assumes that documents in a given corpus are generated by repeatedly picking up a topic, then a word from that topic according to the distribution of all observed words in the corpus given that topic. LDA aims at learning these distributions and inferring the (hidden) topics given the (observed) words of the documents [16]. Given the nature of our data which includes partial annotations, we employ the following two variants of LDA.

Labeled Latent Dirichlet Allocation (LLDA) [18] is a supervised version of LDA that constraints it by defining a one-to-one correspondence between topics and human-provided labels. This allows Labeled LDA to learn word-label correspondences. Figure 1 illustrates the probabilistic graphical model of LLDA.

Partially Labeled Latent Dirichlet Allocation(PLDA) [19] is a semi-supervised version of LDA which extends it with constraints that align some learned topics with a human-provided label. The model exploits the unsupervised learning of topic models to explore the unseen themes with each label, as well as unlabeled themes in the large collection of data. As illustrated in Figure 2, PLDA assumes that the document's words are drawn from a document-specific mixture of latent topics, where each topic is represented as

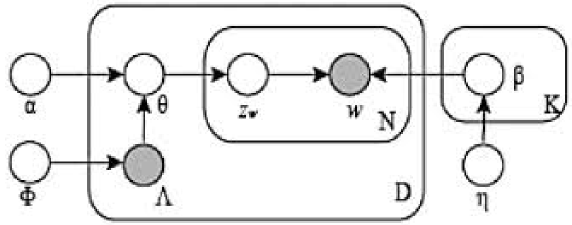

Fig. 1. Probabilistic graphical model of LLDA: unlike standard LDA, both the label set Λ as well as the topic prior α influence the topic mixture θ. [18]

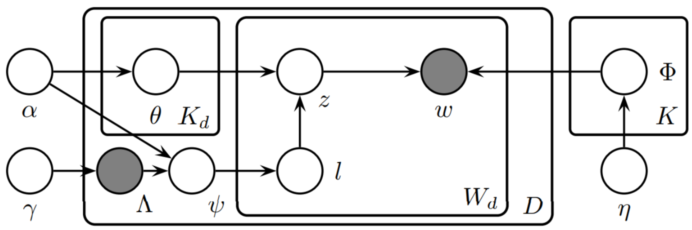

Fig. 2. Probabilistic graphical model for PLDA: each document's word ω and label Λ are observed, with the per-doc label distribution ψ, per-doc-label topic distributions θ, and per-topic word distributions Φ hidden variables. Because each document's label-set λ_d is observed, its sparse vector prior γ is unused; included for completeness. [19]

a distribution over words, and each document can use only those topics that are in a topic class associated with one or more of the document's labels. This approach enables PLDA to detect extensive patterns in language usage correlated with each label.

4 Experimental dataset

For our experiments, we use a dataset consisting of a collection of psychotherapeutic transcripts available for research. The conversations have been transcribed and collected according to the guidelines of the American Psychological Association (APA)[5]. An approval to use the collection was granted by an Internal Committee of Biomedical Experiments (ICBE) of Philips after a review of the agreements, the consent procedures, and data handling plan by legal and privacy experts.

[5] http://www.apa.org

4.1 Data description

Our dataset consists of 1729 transcripts of 1:1 conversation with a total of 340,455 talk turns, 75,732 unique terms, and more than 9 million words. Each transcript has on average 200 talk-turns and eight words for talk-turn. They are also extended with meta-data consisting of the corresponding school of psychotherapy, counselors-patients information such as gender, age range, and sexual orientation as well as a table of topics discussed during the therapeutic conversation. The table of topics contains two different kinds of information:

- *Subjects*: they are organized into three hierarchical levels. The top level is the most general whereas the other two are more precise. For example, the word *Family* could correspond to a top level topic, while *Family violence* and *Child abuse* would be associated to the second and third levels respectively. Up to 575 subjects have been used in the three levels in total.
- *Symptoms*: there are 79 symptoms (e.g. *Depression, Anger, Fear*) as defined in the DSM-IV[6] manual.

4.2 Preprocessing of the meta-data

Given the high number of items in the table of topics, we applied the following steps to merge similar topics and reduce the number of subjects and symptoms to 18 and 16 respectively:
1. Eliminate all the subjects and symptoms that occur in less than 3% of the dataset;
2. Group together all the subjects belonging to the same Wikipedia category[7] regardless their position in the given hierarchical structure.
3. Assign a label to the new subject according to the psychology topics table from APA. For example *Parent-child_relationship* and *Family* are mapped to a new subject from APA known as *Parenting*.
4. Reduce the number of symptoms by using the DSM-IV manual with the expert support of a counselor. In particular, we group symptoms with high-level correlation into a representative one. For example, *Sadness* and *Hopelessness* are merged into the symptom: *Depression*.

The final set is depicted in Figure 3.

4.3 Preprocessing the conversation text

Using the NLTK platform[8] we applied a number of NLP pre-processing steps to the dataset [21]. The resulting corpus consists of 2,849,457 tokens (14,274 unique ones) and a total of 268,478 talk turns. The performed steps include:

- tokenization, which transforms texts into a sequence of tokens

[6] `https://dsm.psychiatryonline.org`
[7] `https://en.wikipedia.org/wiki/Category:Main_topic_classifications`
[8] `http://www.nltk.org/`

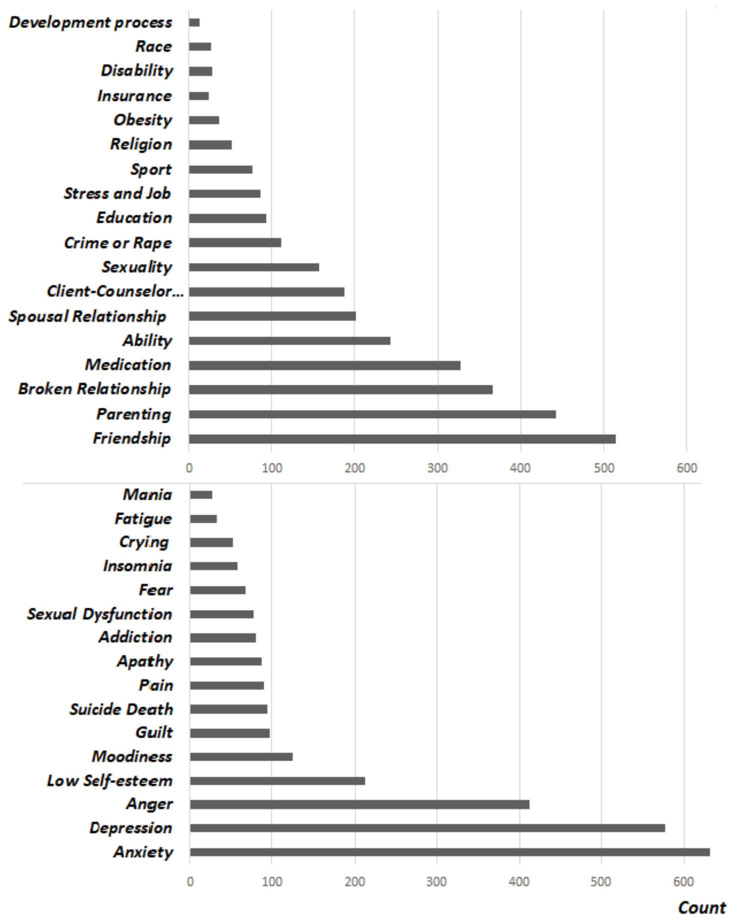

Fig. 3. The resulting 18 subjects (up) and 16 symptoms (bottom)

– removal of all punctuations, stop words, numbers, words that frequently appeared in the text with less content information (e.g., ”mm-hmm”) and words that occurred in less than five documents and keeping nouns, verbs and adjectives only. This was achieved by using unigram part-of-speech tagger [17] to identify the types of words in each talk turn

– removal of the 100 most common words as well as talk turns with one word only as well as words shorter than three characters.

5 The proposed approach for TDT

The proposed automatic TDT method has three phases; the detection of topics, assignment of topic labels to talk turns, and finally tracking of the propagation of topics over the conversation.

5.1 Detection of Topics

The topic detection was performed using a PLDA implementation based on the Stanford Topic Modeling Toolbox[9](TMT). The model requires the definition of parameters such as the number of hidden topics to discover, the hyperparameters α and η (see Figure 2), and a vast amount of short text as input for training purpose. To enlarge the number of corpora, we defined each talk-turn as a document, and we associated each document (talk-turn) with the corresponding topics from the table of topics of the corresponding transcript because PLDA is useful in general only when each document has more than one label associated to it. As a result, we obtained a broader set of documents with higher word co-occurrences. More in detail, we needed to specify how many new topics (different from those in table of topics) the model had to discover. Experimentally this number was set to 20. As a further input, we fed the PLDA with the list of 34 topics and the values of α and η were set to 0.01. In total 268,478 talk turns were obtained after the preprocessing step. Moreover, we used the CVB0 algorithm [4] where the overall number of iterations was 150. After training the model we obtained a list of topics and the associated learned words as shown in Table 1. The method also provided a *per-document topic distribution* for each talk turn. An example is illustrated in Figure 4, where the five topics with the highest likelihood in a conversation are shown (i.e. *Stress and Job*; *Suicide and Death*; *Sexuality*; *Depression*; *Fear*), each with the corresponding talk-turns.

5.2 Assignment of topics

The PLDA returns the topics and the associated terms for each talk turn (document). Based on the terms of each document, we can then determine how likely each document was associated with a topic. Table 1 shows a completion of the terms learned from the trained PLDA topic model. There we list the top ten terms for each topic. In the first column, one can see the discovered topic and its associated words whereas the second and third columns indicate two of the 34 topics already known and their related words. The terms in the same topic tend to be similar, particularly for subjects and symptoms. For example, *Parenting* carries the member of the family, such as *mom, mother, dad, etc.*. Moreover, *Addiction* includes the terms close to alcohol and drugs (*drinking, smoke, etc.*). On the other hand, the domain of the discovered topics has inherent interpretations and holds words that are not covered by the annotations in psychotherapy corpus. For example, the *Topic-5* includes similar terms but their meaning (work) is far from any annotations in APA. For this reason, for the tracking task, we only used the 34 elements present in our table of topics.

[9] `https://nlp.stanford.edu/software/tmt/tmt-0.4/`

Discovered Topic: *Topic-5*		Known Subject: *Parenting*		Known Symptom: *Addiction*	
Associated Words	Weight	Associated Words	Weight	Associated Words	Weight
pay	1125386	*mom*	1354.692	*drinking*	120.9583
month	957.7061	*mother*	1190.696	*drugs*	78.25677
working	809.5385	*dad*	1103.559	*alcohol*	60.09412
end	799.9208	*family*	996.8899	*drug*	59.39509
help	649.7758	*brother*	786.8062	*stoned*	47.30778
giving	516.6524	*parents*	745.5274	*smoking*	44.19726
months	502.375	*father*	689.4923	*marijuana*	41.16274
year	463.4732	*sister*	490.4897	*girlfriend*	37.31469
paid	421.3631	*kids*	376.6678	*smoke*	36.22964
paying	416.7115	*children*	313.5798	*uptight*	35.90694

Table 1. An illustrative example of the three kinds of topics and their most likely associated terms. *Topic-5* shows an example of the discovered topic, *Parenting* presents an example of a known subject, and *Addiction* presents an example of a known symptom.

In addition to the 54 topics (34 known and 20 discovered by PLDA) and their relevant terms, we aimed to know the likelihood of each topic in each talk-turn. As such, another output of the PLDA topic model consisted in the partition of the documents into a set of 54 topic proportions, called per-document topic distributions, where each talk-turn was represented as a combination of topics with different proportions. As already explained earlier, Figure 4 shows an example of the potential ongoing topics in each talk-turn within a therapeutic conversation from the dataset.

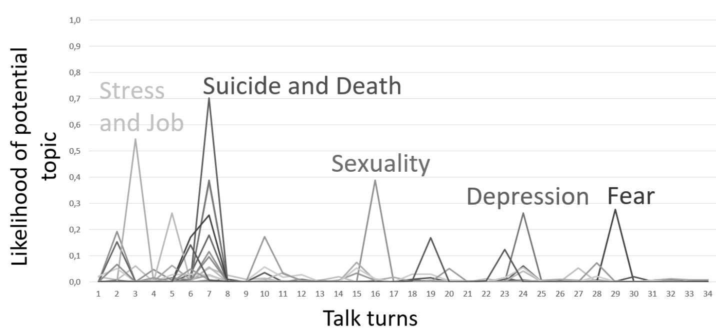

Fig. 4. Example of the *per-document topic distribution* in each talk-turn on a conversation.

5.3 Tracking Topics

The goal was to understand how the known topics propagate, are localized and change for each talker in conversations. We already used PLDA to identify a potential topic for each talk-turn from the table of topics converting the face-to-face conversation to a sequence of topics for each speaker. For the tracking topics task we added a new topic

annotated as *Meaningless talk* which we associated to talk-turns that provide poor semantics contents or language, or non-verbal communication (e.g. ”Yahh!!, Mm-hmm”). We built two topics transitions matrices (TTMs) to understand how the topics change from one talk-turn to another one. Topic changes can be seen as a dynamic mechanism that frequently occurs inside a conversation where speakers move from one topic to another and can be contested by either speaker. We recognize three kinds of changes:
1. The counselor keeps talking about the same topic to the patient from the previous talk-turn, and vice versa;
2. The counselor moves to a new topic after the talk-turn of the patient;
3. The patient moves to a new topic after the talk-turn of the counselor.

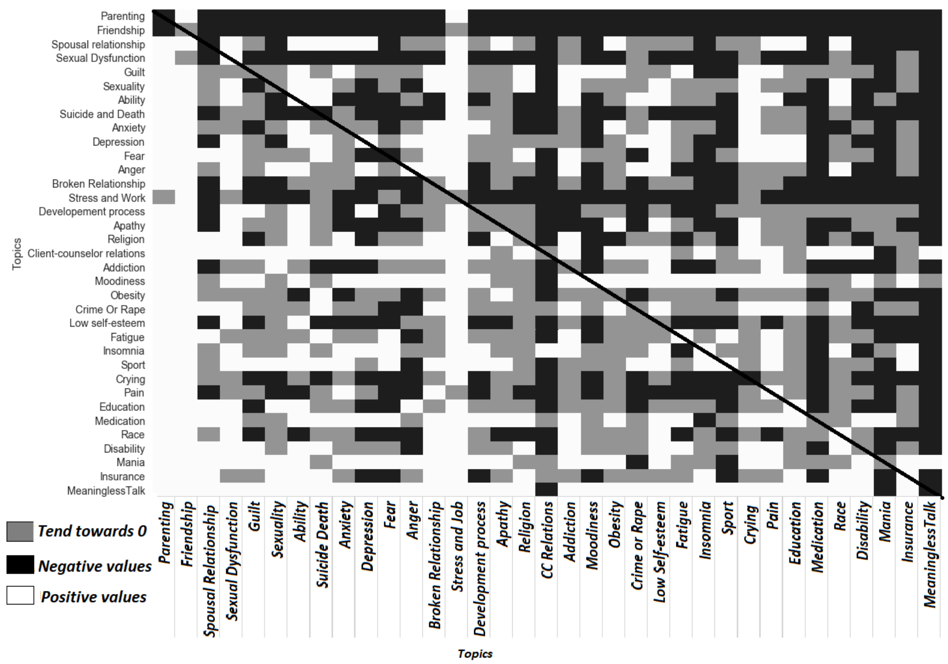

Fig. 5. The difference matrix between CP and PC.

More in detail we constructed patient-to-counselor TTM CP_k that describes all the topics changes within the conversation k. In particular, $CP_k[i, j]$ is the number of times that topic i changes into j in the conversation k. We merged the CP_k matrices together by summing up corresponding elements obtaining our final matrix CP. Similarly, we built a counselor-to-patient matrix PC by using the topics-change defined earlier by switching counselor and patient. The difference between the two matrices is illustrated in Figure 5, which, shows the engagement patterns between counselors and patients, providing a new way to describe one-to-one conversations. There are three possibilities depending on whether the resulting value is lower than -10 (black), between -10 and +10 (gray) and greater than 10 (white). The diagonal of the matrix in Figure 5 gives an idea about the first type of topics changes which corresponds to the "resistance level" on the same topic; it has 17 gray values which proves the fact that the speakers like continuing to talk on the same topic. It has also twelve black values and six white values. The former means that the counselor switches topics twice as the patient does. A possible explanation is that the counselor aims at searching for other correlated symptoms or subjects that would lead to a mental disease. The other values of the matrix describe the second and third type of topic changes; the number of white and black values are approximately equal, which means that the conversations, in general, are discussed without perceived tactics. Nevertheless, some rows and columns are mostly negative or positive, like *Parenting*, and indicate the use of some strategies. The counselor often switches the topic if the previous one was *Mania*, *Medication* or *Patient-Counselor Relations*. Instead, he/she frequently starts a new topic if the patient's talk restrains less semantic contents (*Meaningless Talk*). Contrary, the patient often switches topics if the previous topic was related to *Parenting*, *Friendship*, *Sexual dysfunction*, *Crying*, or *Stress-and-Work*. Conclusively, TTM guides to a bright understanding of how and when topics are changing thus giving important insights to the counselor for CBT.

6 Evaluation

The evaluation of the performance of a topic model is not an easy task. In most cases, topics need to be manually evaluated by humans, which may express different opinions and annotations. The most common quantitative way to assess a probabilistic model is to measure the log-likelihood of a held-out test set performing perplexity. However, the authors in [9] have shown that, surprisingly, perplexity and human judgment are often not correlated, and may infer less semantically meaningful topics. A potential solution to this problem is provided by the topic coherences, that is a typical way to assess qualitatively topic models by examining the most likely words in each topic. For such a purpose, we employed Palmetto[10], a tool to compute topic coherence of a given word set with six different methods. The one that we selected for our purposes was the C_V method [20], which uses word co-occurrences from the English Wikipedia, and that has been proven to highly correlate with human ratings. C_V is depended on a one-set segmentation of the top words and a measure that uses normalized pointwise mutual information. The one-set segmentation computes the cosine similarity between each top words vector and the amount of all top words vectors. The coherence value is then the arithmetic average of these similarities and represents an intuitive measure of the goodness of the topics produced by PLDA. In this work, we evaluated our PLDA topic model for topic detection using C_V coherence. In particular, we gave the top five terms (according to the weight of PLDA shown in Table 1) for each of the 34 topics as the input, obtaining as output a satisfactory coherence amongst all the detected topics. Indeed on average a topics coherence value larger than 50% was obtained, which is recognized in the research community already as a well-acceptable coherence score for a TDT model. This further substantiates the validity and potentials of our method.

[10] `http://aksw.org/Projects/Palmetto.html`

7 Conclusions

Constituting a crucial aspect for the analysis and modeling of counselor-patient conversations, the automatic TDT task in a psychotherapeutic conversation poses a significant challenge. In this paper we implemented a topic detector which efficiently understands therapeutic discussions in consultations. We exploited PLDA and state-of-the-art NLP techniques and topic coherence evaluation systems. Furthermore, we computed TTM to capture the dynamics of each ongoing topic in the conversations understanding how much each interlocutor is affected in the dialogue and when he/she prefers switching topics. Knowing how topics change and their propagations can be used by counselors to drive the discussion and to detect a patient emotional state during the therapeutic conversation. These aspects of interaction are critical for all mental health specialists as they are involved in patient's health concerns. We conclude that PLDA and TTM may be of benefit to the therapeutic conversational speech analysis community, by having high potentials into real-life applications in psychotherapy.

References

[1] M. Allahyari and K. Kochut. Automatic Topic Labeling Using Ontology-Based Topic Models. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 259–264, December 2015.

[2] D. Angus, A. E. Smith, and J. Wiles. Human Communication as Coupled Time Series: Quantifying Multi-Participant Recurrence. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1795–1807, August 2012.

[3] D. Angus, B. Watson, A. Smith, C. Gallois, and J. Wiles. Visualising conversation structure across time: insights into effective doctor-patient consultations. *PloS One*, 7(6):e38014, 2012.

[4] A. Asuncion, M. Welling, P. Smyth, and Y.W. Teh. On Smoothing and Inference for Topic Models. UAI '09, pages 27–34, Arlington, Virginia, United States, 2009. AUAI Press.

[5] S. Bangalore, G. Di Fabbrizio, and A. Stent. Learning the Structure of Task-Driven Human-Human Dialogs. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7):1249–1259, September 2008.

[6] D.M. Blei and J.D. Lafferty. Dynamic Topic Models. ICML '06, pages 113–120, New York, NY, USA, 2006. ACM.

[7] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

[8] A. Breuing and I. Wachsmuth. Talking topically to artificial dialog partners: Emulating humanlike topic awareness in a virtual agent. volume 358. 2013.

[9] J. Chang, J. Boyd-Graber, S. Gerrish, C. Wang, and D.M. Blei. Reading Tea Leaves: How Humans Interpret Topic Models. NIPS'09, pages 288–296, USA, 2009. Curran Associates Inc.

[10] W.T. Chen, S.C. Lin, S.L. Huang, Y.S. Chung, and K.J. Chen. E-HowNet and Automatic Construction of a Lexical Ontology. COLING '10, pages 45–48, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[11] Y. Chen and L. Liu. Development and research of Topic Detection and Tracking. In *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 170–173, August 2016.

[12] P. Drew, J. Chatwin, and S. Collins. Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy*, 4(1):58–70, March 2001.

[13] G. Gaut, M. Steyvers, Z. E. Imel, D. C. Atkins, and P. Smyth. Content Coding of Psychotherapy Transcripts Using Labeled Topic Models. *IEEE Journal of Biomedical and Health Informatics*, 21(2):476–487, March 2017.

[14] A. Gelbukh. Natural language processing. In *Fifth International Conference on Hybrid Intelligent Systems (HIS'05)*, pages 1 pp.–, November 2005.

[15] C. Howes, M. Purver, and R. McCabe. Investigating Topic Modelling for Therapy Dialogue Analysis. March 2013.

[16] W. Mohr J and P. Bogdanov. Introduction—topic models: What they are and why they matter. *Poetics*, 41(6):545 – 569, 2013. Topic Models and the Cultural Sciences.

[17] N.P.P. Khin and T.N. Aung. Analyzing Tagging Accuracy of Part-of-Speech Taggers. Advances in Intelligent Systems and Computing, pages 347–354. Springer, Cham, August 2015.

[18] D. Ramage, D. Hall, R. Nallapati, and C.D. Manning. Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora. EMNLP '09, pages 248–256, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[19] D. Ramage, C.D. Manning, and S. Dumais. Partially Labeled Topic Models for Interpretable Text Mining. KDD '11, pages 457–465, New York, NY, USA, 2011. ACM.

[20] M. Röder, A. Both, and A. Hinneburg. Exploring the Space of Topic Coherence Measures. WSDM '15, pages 399–408, New York, NY, USA, 2015. ACM.

[21] A.K. Uysal and S. Gunal. The Impact of Preprocessing on Text Classification. *Inf. Process. Manage.*, 50(1):104–112, January 2014.

[22] J.F Yeh, Y.S. Tan, and C.H. Lee. Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet allocation. *Neurocomputing*, 216(Supplement C):310–318, December 2016.