Mixed-Initiative Real-Time Topic Modeling & Visualization for Crisis Counseling

Karthik Dinakar[†] **Jackie Chen**[†] **Henry Lieberman**[†] **Rosalind Picard**[†] **Robert Filbin**[*]
{karthik, jchen, lieber, picard}@media.mit.edu , bob@crisistextline.org
[†] Massachusetts Institute of Technology [*] Crisis Text Line

ABSTRACT
Text-based counseling and support systems have seen an increasing proliferation in the past decade. We present Fathom, a natural language interface to help crisis counselors on *Crisis Text Line*, a new 911-like crisis hotline that takes calls via text messaging rather than voice. Text messaging opens up the opportunity for software to read the messages as well as people, and to provide assistance for human counselors who give clients emotional and practical support. Crisis counseling is a tough job that requires dealing with emotionally stressed people in possibly life-critical situations, under time constraints. Fathom is a system that provides topic modeling of calls and graphical visualization of topic distributions, updated in real time. We develop a mixed-initiative paradigm to train coherent topic and word distributions and use them to power real-time visualizations aimed at reducing counselor cognitive overload. We believe Fathom to be the first real-time computational framework to assist in crisis counseling.

Author Keywords
Natural Language Interfaces, Topic Modeling, Visualizations, Mental Health Counseling

ACM Classification Keywords
H.5.2 User Interfaces: Natural Language; H.5.2 Group and Organization Interfaces: Web-based interaction

INTRODUCTION
Studies have estimated that more than one in four Americans aged eighteen and older are afflicted with a diagnostically determinable mental disorder in a given calendar year [Wang et al., 2005]. Though the distribution of mental disorders is widespread in the population, they manifest severely in a much smaller and concentrated proportion; six percent or one in seventeen Americans are thought to suffer grievously from such disorders [Wang et al., 2005]. Nearly a third of all practicing psychotherapists and counselors are expected to encounter grave crises of their clients at some point in their careers [Schwartz and Rogers, 2004]. There has been a proliferation of hotlines to help people in crises within the last decade, where a distressed individual experiencing a crisis, or the ***caller***, is connected to a crisis ***counselor*** via telephone hotlines [Gould et al., 2013] and web-based chats [Christensen et al., 2013] for crises ranging from suicide and self-harm to sexual assault. Yet there is evidence that for a distressed individual, communication via text can be preferable to voice phone calls. According a Pew report, there has been an increase of 45% in text communication since 2009 [Smith, 2012]. There is also growing evidence that distressed individuals are more comfortable sharing sensitive information using electronic-based media [Alison Bryant et al., 2006]. Text-based messaging provides opportunities for software to **(a)** read messages from distressed callers as they arrive, and **(b)** provide real-time assistance to the counselors who respond to crisis calls.

In this paper, we investigate the challenges of mental health crisis counselors on a national, 911-like hotline called *Crisis Text Line* and design analysis and visualization tools to help a counselor manage their cognitive load. We perform a contextual inquiry to identify key pain-points. We develop a mixed-initiative [Horvitz, 1999] probabilistic graphical topic model framework and a visualization framework to give real-time visual summaries to counselors of both ongoing and archived conversations. We evaluate our prototype on seven crisis counselors. An additional contribution is the integration of a mixed-initiative latent variable model like LDA with visualizations to analyze large scale call logs of crisis counseling that can potentially inform prevention science psychology.

BACKGROUND
Before going into the design of the Fathom prototype, it is worth situating it with an overview of Crisis Text Line (CTL). Crisis counseling [Hoff et al., 2011] differs from conventional psychotherapy in its brevity - the short duration of counseling constrains extended interaction between a caller and a crisis counselor. While successful psychotherapy is conditioned on a more longitudinal and detailed understanding of an individual's plight, crisis counseling can help transition a caller to longitudinal psychotherapy [Gould et al., 2012].

CTL is a text-based platform supporting a network of organizations that offer crisis counseling and support, with a particular focus on adolescents. CTL is a platform that is used by crisis specialists or counselors from different crisis support centers across the United States. A distressed individual or caller undergoing a crisis may at any time send an SMS text message to the CTL help number. Upon receiving a message from a caller, she or he is connected to a trained specialist on

IUI'15, March 29 - April 01 2015, Atlanta, GA, USA
Copyright 2015 ACM 978-1-4503-3306-1/15/03$15.00
http://dx.doi.org/10.1145/2678025.2701395

one of the support centers that use the CTL platform. Figure 1 illustrates the CTL platform and the network of crisis support centers, each of which are organizations independent of each other and CTL. The crisis support centers are connected to the CTL platform via a web interface; each counselor that is part of one crisis center can see the load and status of every other support center as well as the queue of callers waiting to be served. For example, a counselor whose support center is in Boston can see the status of a counselor who might be a part of a crisis center in Chicago and vice-versa. Incoming callers are accepted regardless of where they are texting from geographically, as this information is not revealed to the counselors.

The counselors staffing a crisis support center are usually a mixture of paid staff and volunteers, who undergo a period of training prior to handling calls as counselors. The number of calls to crisis support centers have historically exceeds demand - callers who may not find an open counselor are made to wait in a queue.

Surveys have unearthed a high attrition rate among crisis counselors, with burnout and low morale as the top two contributing factors given the huge volume of messages they exchange with callers every day [Kinzel and Nanson, 2000]. A stable, supportive work environment, with a concerted effort to reduce counselor cognitive workload and burnout are thought to help in retaining trained counselors [Pratt, 2013]. In this work, we design a suite of tools to minimize counselor cognitive load and maximize counselor-caller interaction time by reducing the time and effort expended on cursory tasks other than communicating with the caller. While maximizing counselor-caller interaction time on its own cannot fully address problems of burn-out and better counselor, it was identified as significant contributor, which we focus on in this work.

CONTEXTUAL INQUIRY

Counselors using the CTL platform use chat-based web application to have counseling conversations with callers. The callers communicate with the counselors via SMS text messages. Each SMS message is limited to 160 characters. There are restrictions on how much information about a caller a counselor is allowed to access. Though callers are uniquely identified by their telephone numbers, these are not available to the counselor. Instead, each caller has a profile, with details and notes on the last three prior conversations; counselors handling new callers create their profiles. At any given time, a counselor handles two or more conversations in parallel, as these are asynchronous calls through text messages and not voice.

We performed a contextual inquiry (as shown in Fig 1) on two counselors for two hours each on three different days. Below, a work flow model involving the counselor, caller and a supervisor who manages each crisis support center. A typical work-flow of a counselor, along with the pain-points is explained below:

1. **Accept caller**: An alert is sounded when a new caller arrives, who is then placed in a waiting list. The counselor then examines the queue and chooses to accept a caller from the waiting list. Choosing a caller from the waiting list is conditioned on serving a mixture of new and repeat callers. Handling new callers involves more work than talking to repeat callers. On some occasions when the counselor was too overloaded, either because of the gravity of a particular conversation or handling more than three conversations in parallel, a conversation is **warm transferred** to another counselor, who has to pick up the conversation from where it was prior to being transferred.
2. **Check caller profile**: Upon accepting a caller, the counselor then glances through the caller profile if available. Details and notes from previous conversations that might have been handled by another counselor are examined. For new callers, the counselor elicits information from them to fill their profiles. Counselors are trained to treat each call, whether it is from a new or repeating caller, as a distinct and important call. Knowledge of prior conversations is not disclosed to the caller, even as it helps create a context in which the counselor begins the counseling session. Reading through prior notes was time consuming; in almost all cases, reading through the entire transcript of previous calls was even more time consuming.
3. **Counseling session**: The counselor then performs a counseling session with the caller, with three key elements for every conversation:
    - ***Risk assessment***: What is the crisis facing the caller? Is she or he a threat either to themselves or to others? For example, a caller with a history of *drug abuse* and *depression* and who is experiencing intense suicidal thoughts might be construed as being a *high risk* caller.
    - ***Issues & emotional state***: What are the main issues facing the caller? For example, a teen with *relationship difficulties* might be *depressed*.
    - ***Action plan***: What concrete steps can the caller to take to tide over the immediate crisis facing them? For example, a caller who has recently *divorced* might be asked to reach out to friends and family and possibly a therapist.
4. **Take notes**: For each conversation, counselors take real-time manual notes to indicate important aspects of the conversation based on the aforementioned key elements. While some counselors take notes at the end of a conversation, others choose to take notes as the conversation is progressing. Conversation notes for a given conversation are sharable across counselors and the time spent in taking manual notes is significant. During contextual inquiry, 38% of counselor's time went in taking notes, even as it delayed talking to other callers currently being served and the ones in the waiting list.
5. **Fill conversation report**: A report is filled at the end of each conversation based on a static template. This report has categories of issues and counselor responses, and is separate from the notes taken by the counselor. Again, filling the conversation report is a time consuming job that

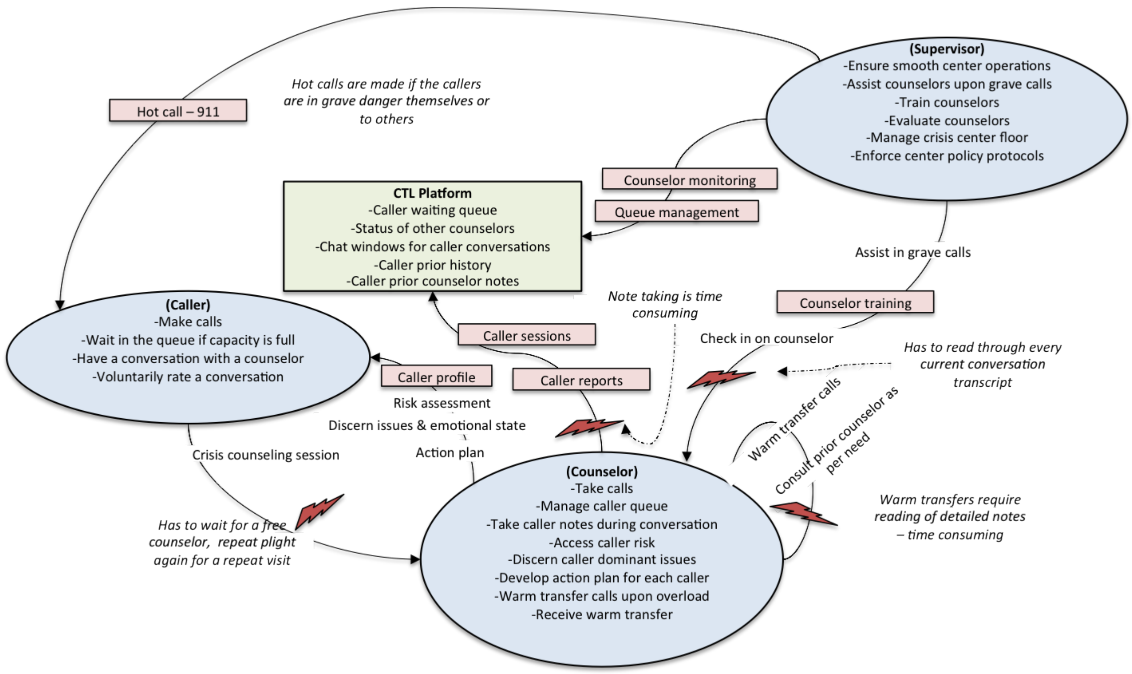

Figure 1. Contextual Inquiry:A work-flow model of a crisis support center using the CTL platform. Four key pain-points, depicted as lightning flashes are as follows: (1) Manual real-time note taking is time consuming; (2) Warm transfer of calls to a new counselor are time consuming because the new counselor has to read prior counselor notes; (3) COunselors notes, when not adequate to describe a conversation, forces repeat users to explain their plight again, and (4) There are no structured ways for a supervisor to examine a set of conversations.

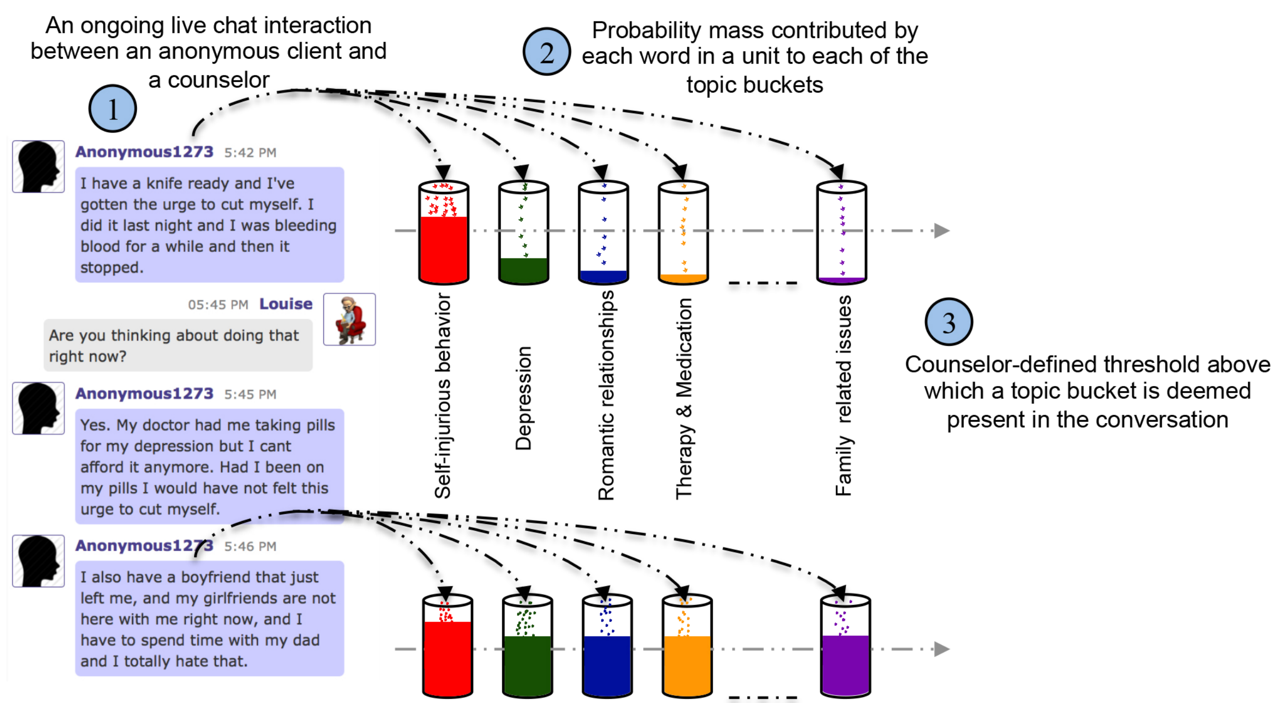

Figure 2. The relationship between the conversation text and the topic model. Words in the conversation each make a contribution to buckets representing topics. The set of topics were determined by an mixed-initiative process, involving annotation of a corpus of conversations by Prevention Science psychologists, in order to be appropriate for the crisis counseling application. A threshold is set for a bucket to be considered significant for that conversation. The conversation shown here is a canned example entered by a volunteer at the time of evaluation. Our IRB prevents us from showing any actual counselor-patient conversations.

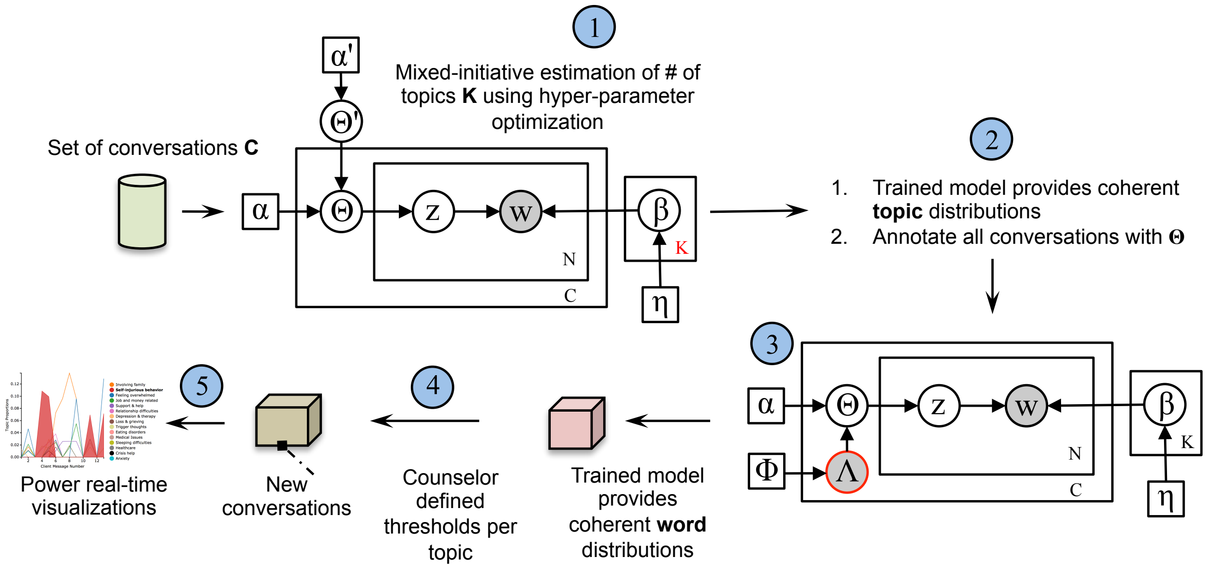

Figure 3. A two-stage mixed-initiative modeling process. (1) Run vanilla LDA with hyper-parameter optimization with prevention science psychologists in the loop using the mixed-initiative method to fetch coherent topics distributions; (2)Use the topic distributions to annotate all the conversations with psychologist-assigned topics; (3) Run a labeled LDA model using the annotated conversations and perform model selection over perplexity curves; (4) Use the word distributions from the model selected and assign user-defined thresholds for each topic; (5) Use the models real-time for each new conversation.

eats into time that could have been served in talking to another current caller or in accepting a new one from the waiting list.

6. **Monitor waiting list**: Even as a counselor accepts calls and handles two or more calls at the same time, they also have to keep an eye on and monitor the waiting list. This is predicated on the belief that any of the callers in the waiting list might in a grave crisis, and minimizing wait time is a priority.

All three counselors that were interviewed during the contextual inquiry had similar pain-points. The task of having to take detailed notes while having conversations in parallel was deemed time consuming. For repeat callers, the manual task of going through prior conversation transcripts and notes was deemed very helpful, but was also time consuming. Filling in a report post-conversation was unanimously considered to be time consuming; the report itself was considered too simple to capture the complexity of issues faced by a caller. None of the three counselors read conversation reports of previous conversations and jumped right in to counselor notes and the conversation transcript.

It was clear from contextual inquiry that there was a need to **(a)** visually summarize the content of a conversation for future use, **(b)** reduce the amount of note-taking by extracting clear themes and issues from a conversation in an automated manner, and finally **(c)** use salient issues extracted from a conversation to populate a report at the end of the counselor-caller interaction, that captures in better detail the complexity of issues faced by the caller.

TOPIC MODELS FOR CONVERSATION SUMMARIES

Traditional supervised learning is an analytic technique. Unsupervised techniques such as LDA are generative, but the problem is that the topics they generate may or may not be interesting and relevant to a particular application or purpose. Our methods provide a way to provide gentle guidance to an unsupervised learning process, while retaining the valuable ability to discover unanticipated salient topics. A visual summary of a conversation is predicated on an thematic understanding of the salient issues faced by a caller, her or his emotional state and reactions and the relationships between them. To train a computational model to summarize a conversation, we received a data set of through a licensing agreement with Crisis Text Line and after an IRB review.

We analyzed a total of 8106 conversations consisting of 469,849 (6,787,627 words) counselor messages and 412,050 caller messages. There were 214 counselors in total. The average length of a counselor message and a caller message were 14.4 and 10.5 words respectively.This data was received de-identified, devoid of any personally identifiable information.

In the machine learning literature, challenges in the computational modeling of interaction analysis are well known: sociolinguistic variation in interaction analysis, the presence of affect, of personality and ascription to community tend towards specificity and uniqueness, whereas much of machine learning and natural language processing tend towards abstraction and generalization [Gianfortoni et al., 2011]. Of the conventional supervised and unsupervised machine learning methods for natural language understanding and text classification, probabilistic graphical topic models have been successfully used in the context of detecting teenage bullying and to help distressed teens find similar stories to theirs [Dinakar et al., 2012].

Mixed-initiative modeling

A key strength of topic models is their ability to generatively model corpora. They have been shown to work comparably with supervised learning models in their ability to model personal teenage stories of distress [Dinakar et al., 2012, Dinakar et al., 2014b] and in document classification [Rubin et al., 2012]. We adopt a two-stage, mixed-initiative modeling process involving prevention science psychologists in the loop as shown in Fig 3.

In the first stage, we are interested in extracting meaningful topics (as validated by the prevention science psychologists), or in other words, a ***coherent topic distributions***. For this we use vanilla LDA with hyper-parameter optimization[Wallach et al., 2009, Hoffman et al., 2010] and use the mixed-initiative method of training LDA models with psychologists in the loop [Dinakar et al., 2012]. This is achieved by having a group of prevention science psychologists construct sentences from every intermediate topic distribution after a training loop (using as few other words as possible) and subsequently achieve agreement with other counselors to see if the particular distribution of words denotes a coherent topic. More details about this approach is fully described in our prior work [Dinakar et al., 2012]. Armed with a good topic distribution, we can then use such the topic distributions to annotate each conversation in the corpus.

Once we have we have an annotated set of conversations, we are then interested in generating a ***coherent word distributions*** for each topic. This is achieved by using a labeled LDA model [Ramage et al., 2009] trained over the set of topics derived from stage 1. A model selected over the best perplexity curves over held-out data then gives us the best coherent word distribution for the corpus. The algorithmic details and its effectiveness as compared to more conventional approaches are detailed in an earlier paper presented to a more computational research venue [Dinakar et al., 2014a] , but it is worth summarizing how our mixed-initiative LDA approach addresses the key requirements described above:

1. **Modeling the thematic distribution of issues and emotional reactions** -Topic models are mixture models [Blei, 2012]. Labeled latent Dirichlet allocation allows each story to be viewed as a mixture of various issues. The set of such issues and reactions across the corpus of stories provides an excellent overview of the corpus of conversations, allowing us to drill down into the constituent set of issues for each conversation and to view how these issues interact with one another at a corpus-wide level.

2. **Modeling the relationships between caller issues and emotional reactions** - By examining the most dominant issue in a conversation, we can examine the relationships between issues and see how they are related to each other. For example, issues related to loss of a job and loss of medical insurance may co-occur more with depression and relationships under duress.

3. **Modeling socio-linguistic variation** - Because topic models define each topic as a distribution over words, it captures relationships between words as pertaining to a topic. For example words denoting a therapist ranging from doc to shrink were all clustered under the same issue of depression and therapy.

The trained model is used as follows to tag every ***unit*** or SMS message sent by a caller. Topics are distributions over corpus-wide vocabulary of words. For each word in a given unit, we see the probability mass contribution of that word to every topic bucket in the model. The topic bucket that cumulatively gives the most mass is the most dominant topic for that unit. The threshold for when a topic bucket is deemed dominant can be set by the counselor. This is depicted visually in Fig 2.

FATHOM INTERFACE

We use the suite of trained topic models to power a series of visualizations to visually summarize a conversation in real time.The same visualizations can also be used to see visual summaries of prior conversations and those that are warm transferred. We create a prototype interface, *Fathom*, that mimics the real CTL interface in its core functionality, with real-time conversations between a counselor using Fathom and a caller using SMS text messaging through a phone (shown in Fig 4). We design five visualizations using the *D3js* framework [1] to summarize a given conversation, with transitions updating real-time, with every unit spoken by the caller as follows:

1. **Quick summary**: In this visualization, topics extracted from the models are presented as a named topic list of colored nodes, with the radius of each node determined by the proportion of the respective topic in the conversation. The most most dominant topic extracted in the conversation appears first, and is the largest node in the list. Each subsequent node representing a topic is less and less dominant in the conversation (see figure 5).

2. **Conversation topic proportions**: A donut chart with the topic proportions determining the slice of the donut for the respective topic. The legend is a miniaturized version of the topic list summary discussed in the previous visualization. We see on hover that the topic 'involving family' constitutes 48% of the issues extracted from this conversation (see figure 6).

3. **Conversation topic evolution (non-cumulative)**: We design a line graph to show a chronological occurrence of topics in the conversation. We plot the unit number received from the caller on the x-axis and plot the topic proportions on the y-axis. We plot the topic proportion of each

[1] www.d3js.org

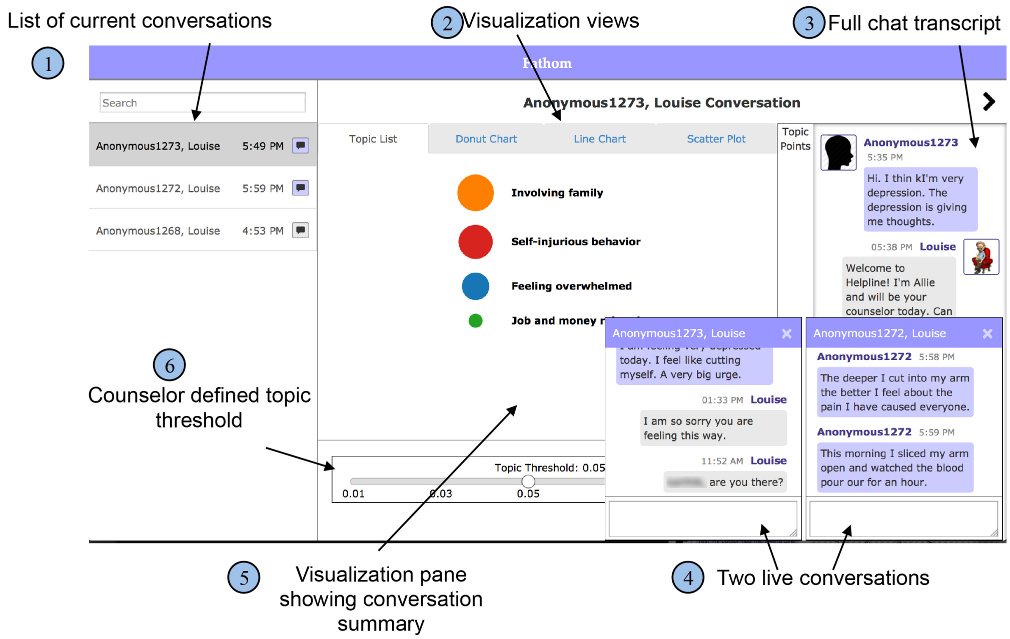

Figure 4. The Fathom interface shows a visualization pane that summarizes a conversation. The control interface omits the visualization engine. Note how a counselor is having more than one conversation in parallel at any given time.The interface here shows two live conversations in progress, with the full transcript of one of the conversations in the background. The visualization pane also has a counselor defined topic threshold setting, where a lower threshold generates more topics, explained in Fig 2. The conversations shown here are canned examples entered by volunteers during our experiment.

topic against the unit number. Selecting a topic in the legend highlights the positions in the graph where the respective topic occurred in the conversation as shown below (see figure 7).

4. **Conversation topic evolution (cumulative)**: We design a line graph to show a chronological occurrence of topics in the conversation. We plot the unit number received from the caller on the x-axis and plot the topic proportions on the y-axis. We plot the topic proportion of each topic against the unit number. Selecting a topic in the legend highlights the positions in the graph where the respective topic occurred in the conversation as shown below (see figure 8).

5. **Visual indexing via scatter plots**: In this visualization, each topic is a row. A topic when dominant (as set by the counselor threshold) is placed as a dot for the unit for which it was detected. The size of the dot is again determined by the proportion of the topic present. The main purpose of designing the scatter plot was to assist in visual indexing, which we explain in the next section.

Visual indexing

The line graphs and scatter plots were designed to allow visual indexing of a conversation. Upon clicking on a point on the line graph or a dot on the scatter plot, the counselor is automatically taken to the part of the conversation transcript where the respective units were tagged with the topic. A visual summary of a conversation, both ongoing and historic, along with affordance for easy visual indexing was designed to give a counselor a quick and easy gist of the dominant topics in a conversation and spare them from having to read

Figure 5. Quick Summary: An ordered list of topics extracted from a conversation. The node radius of a topic, arranged in descending order, is determined by the proportion of the respective topic in the conversation. In this example, the topic 'involving family' is the dominant topic in the conversation.

the entire chat transcript. Our contextual inquiry shows that visual summaries could potentially help a counselor to save time which might be better used in serving existing callers or those in the waiting lines. The figure below illustrates visual indexing via the scatter plot.

EVALUATION

We evaluate the Fathom interface against a control interface (without the visualization engine as described in Figure 4) to discern if the visualization summaries of a conversation might be helpful to get a gist of dominant topics for repeat callers' conversation histories as well as warm-transferred conversations.

Additionally, we also evaluate of the visualization framework could help in each of the key counseling foci of risk assess-

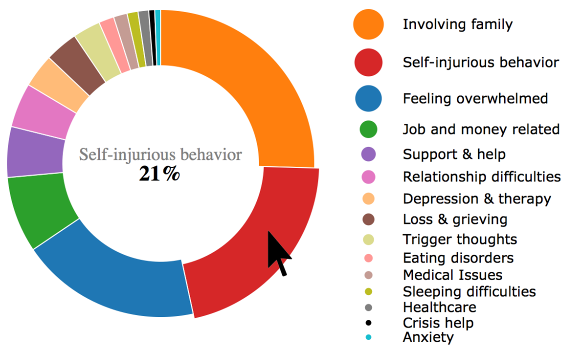

Figure 6. Conversation topic proportions: The slice of the donut is proportional to the dominance of the topic in the conversation. The topic 'involving family' is the most dominant topic in this example, constitutes 48% of the topics extracted from its conversation.

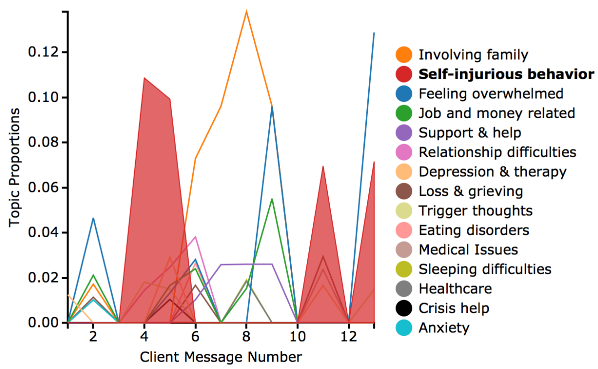

Figure 7. Conversation topic evolution (non-cumulative): The unit or message number, in the order they were received from the caller, is plotted against the topic proportions of topics extracted from those units. In this example, we select 'Self-injurious' behavior from the legend on the right. We see that it spikes at the beginning and at the end of the conversation. s

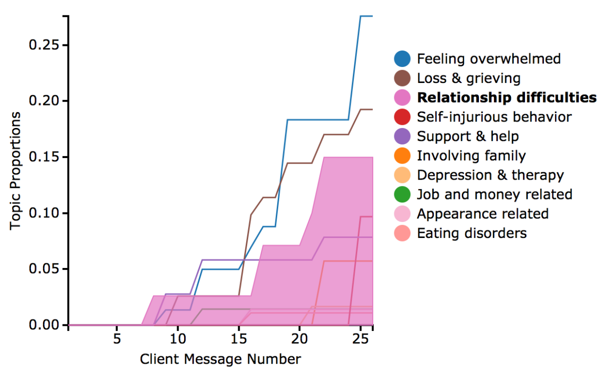

Figure 8. Conversation topic evolution (cumulative): This is a cumulative version of the previous visualization, where one can see if a given topic increases in dominance as the conversation progresses. The unit or message number, in the order they were received from the caller, is plotted against the topic proportions, cumulatively, of the topics extracted from each unit from the caller. In the example below, we select 'Relationship difficulties' in the legend and see how it grows in proportion as the conversation evolves. The topic 'Feeling overwhelmed' is the most dominant issue whose proportion increases as the conversation progresses.

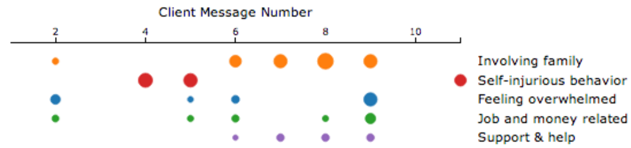

Figure 9. Scatter plot: In this example, we see that unit 9 was tagged for 4 topics, namely 'involving family', 'feeling overwhelmed', 'job and money related' and 'support and help'. The most recent unit was tagged for self-injurious behavior, but a counselor can quickly see at what other points in the conversation a particular topic was present.

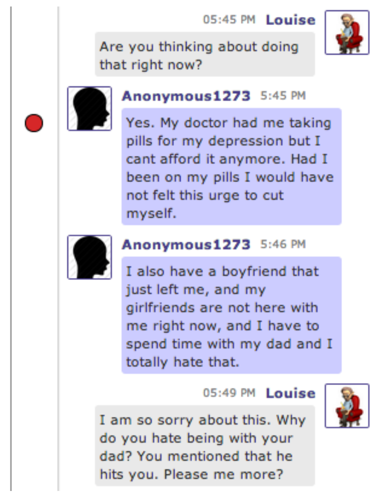

Figure 10. Visual indexing: Upon clicking the second red dot shown in the previous visualization (for self-injurious behavior), the chat transcript automatically scrolls to the corresponding unit in the conversation which was tagged for self-injurious behavior. This visual indexing allows for seamless and quick gist of conversations without having the read the entire chat transcript, saving time that might be better utilized in serving existing callers or the ones still waiting to be accepted by a counselor. The example shown here is a canned one entered by a participant at the time of evaluation.

ment, gist of topics and caller emotional state, and formulating an action plan for the caller.

Participant selection

We selected a total of 7 participants for evaluating Fathom. All 7 participants had undergone training as crisis counselors and three of them worked in a crisis support center currently using CTL as a platform to offer crisis counseling, with the other four actively engaged in the recruitment of counselors and who were well-versed with a plethora of crisis counseling software.

Experiment protocol

Each participant was shown two interfaces, Interface 1, and Interface 2, namely the control interface and Fathom respectively. The user evaluation instructions consisted of two parts. Six pre-selected and existing stories were shown per interface in part 1 to gauge how useful the visualization framework might be with respect to risk assessment, gist of topics and& emotional state of the caller and formulating an action plan for the caller. Participants were asked to imagine these conversations were warm-transferred to them, where they had to pick up these conversations at the point a previous counselor might have handed it to them.

In the second part of the experiment, participants were asked to engage in a live conversation with a member of the experiment team playing the role of a caller, and evaluating each of the visualizations in Fathom that updated dynamically with every unit from the test caller. The questions asked of the participants in both parts are as follows:

Part 1

- **Q1**: Which interface was useful for a risk assessment of a warm-transferred call (Interface 1 or Interface 2), and how useful was this interface? (1 not useful to 5-extremely useful)

- **Q2**: Which interface was useful for eliciting the list of issues and emotional state of a warm-transferred call (Interface 1 or interface 2), and how useful was this interface? (1 not useful to 5-extremely useful)

- **Q3**: Which interface was useful for developing an action plan for a warm-transferred call (Interface 1 or Interface 2), and how useful was this interface? (1 not useful to 5-extremely useful)

- **Q4**: Which interface was useful to get a conversation summary for repeat callers (Interface 1 or Interface 2), and how useful was this interface? (1 not useful to 5-extremely useful)

Part 2

- **Q5**: How useful was it to view the visualizations real-time for conversation note taking? (1 not useful to 5-extremely useful)

- **Q6**: Which visualization was the most useful for quick grasp of a conversation? (Quickly summary, topic proportions, topic evolution (non-cumulative), topic evolution (cumulative), visual indexing via scatter plots)

Participants were also allowed to add open-ended responses for each of the questions if they wished to elaborate on their votes.

Results

The results for each of the questions from Part 1 are tabulated below:

	Control		Fathom	
	Votes	Average Rating	Votes	Average Rating
Q1	71.38%	2.6	28.57%	3.5
Q2	14.28%	3	85.71%	3.9
Q3	57.14%	3	42.85%	3.3
Q4	0%	0	100%	4.42

Table 1. Results from Part 1 show that most of participants overwhelmingly preferred to read specific parts of the transcript at the beginning to discern callee risk. Fathom was rated by 85% and 100% of the participants for the most useful interface for eliciting callee emotional state and their issues, as well as the interface to help get a gist of prior conversations by repeat callees.

For part 2, question **Q5**, Fathom got an average rating of 4, denoting that our visualization framework was indeed useful in real-time conversations in terms of supplementing the manual taking of conversation notes. For **Q6**, 4 participants said that the scatter plot was the best visualization for visual indexing, while 2 preferred the topic list summary and only 1 preferred the non-cumulative line graph.

DISCUSSION

In their open-ended feedback, participants were very positive about the visualization framework of Fathom, but expressed a desire for even more functionality. Visualizing counselor notes along with conversation summaries, was a common feedback. Three participants also wanted certain topics such as suicide ideation and risk to be prominently displayed, no matter how dominant it might be in a conversation. Our evaluation shows that counselors preferred the control interface (where they see read the entire chat transcript) to Fathom for risk assessment (**Q1**) of a warm-transferred call. Verbatim feedback from them mentioned how risk assessment was counselor-specific - that it was a function of more than a set of topics and that the tone and intensity of the caller responses to risk-assessment questions was also a big factor. This shows the need for additional modeling of caller responses to sensitive risk-assessment questions from a counselor, which can happen at any point during a conversation. In terms of developing an action plan (**Q3**) however, opinion was divided. Two counselors thought that dispensing an action plan is closely is conditioned on risk assessment, suggesting that modeling risk-assessment better can also help in formulating action plans for warm-transferred calls.

The verdict for eliciting the emotional state of the caller (along with a list of topics underlying the conversation - **Q2**) and for viewing conversation summaries (**Q4**) for repeat callers was decisively in favor of Fathom. Counselors seemed particularly enthused by the possibility of using Fathom to augment the mandatory caller-report task at the end of conversations.

Overall, our evaluation suggests two takeaways: **(1)** A joint, generative topic modeling of counselor-caller interactions, with a specific focus on risk assessment, action plan formulation, eliciting emotional states and projection of empathy. Embedding counselor expertise in the approximate posterior inference while training is already an ongoing research direction we're pursuing. **(2)** Making risk assessment areas explicit and prominent in the visualizations, as they seem to bear an impact on formulating an action plan as well.

RELATED WORK

The workshop paper [Dinakar et al., 2014a] described an earlier version of Fathom's interface. The present paper reports the Contextual Inquiry, details our mixed-initiative topic modeling, and presents the evaluation with the counselors.

The particular visualizations used by Fathom are described in the visualization toolkit D3JS [Bostock, 2012]. The four visualizations, the quick summary, conversation topic proportions, topic evolution and visual indexing are all standard visualizations in D3JS's library. D3JS's emphasis on data-driven visual transformations makes it most suitable for Fathom's dynamic approach to topic modeling.

The most directly related work to Fathom is in the area of visualizing topic models. Several prior projects have provided different graphical views on topic models, and have also provided means for associating topic models or component words back to the original corpus, either in particular examples or in aggregate [Chaney and Blei, 2012]. The major difference is that Fathom is intended for applying topic modeling to examples that unfold in real time conversations rather than on static documents or static corpora.

The Topic Browser [Gardner et al., 2010] provides a variety of techniques for visualizing topic models, focusing on features that help users understand the context in which the model occurs, such as word clouds and keyword-in-context (KWIC) views. There is also an emphasis on helping the user to understand properties of weighted-word topic models, like document entropy, coherence, and model similarity, that help educate the user as to how the machine learning techniques achieve their summarization of the relationship of documents to corpora. Again, the documents and corpora are not expected to change during the user interaction with the visualization. In Fathom, each unit of a conversation is handled by a back-end suite of topic models to power a front-end visualization.

TextFlow [Cui et al., 2011] is a visually innovative visualization technique that does address the temporal element of evolution of topics in a document. As the authors state, it is less suited to associating topics with particular examples, and they advocate a hybrid approach. Certainly, in future work it would be worth experimenting with unconventional display methods such as this in our context, for highlighting dynamic evolution of topics.

Finally, another relevant area is software for call center management. Visualization is an important component of many call center management tools, of which Dundas [Hedgebeth, 2007] is a typical example. Many provide CRM (Customer Relationship Management) tools, and problem-solution databases to the responders. Some tools provide post-hoc monitoring of the performance of call center personnel and understanding in the aggregate the topics of the incoming calls [Pallotta et al., 2011]. We are not aware of any prior work that uses machine learning techniques such as topic models to provide real-time assistance to crisis counselors.

SUMMARY

Crisis counseling is a job that has a high cognitive load, and managing that load in real-time can make the difference between life and death for some people. Both graphical visualization and machine learning have been proven powerful computational tools for helping people deal with complex situations, and understand them quickly. But these fields have largely developed independently, and it is now time for them to work together to help solve this important problem.

We have presented Fathom, the first system that uses a variety of dynamic visualizations of topic models using the machine learning technique LDA, to help crisis counselors understand caller conversations in real time, and facilitate collaboration between counselors. When a mental health emergency strikes, both people and computers stand ready to lend a helping hand.

ACKNOWLEDGMENTS

This work was funded in part by the Reid Hoffman Fellowship and the Google Faculty grant. We thank the team at Crisis Text Line for their support and help. We thank Emma Kerry, Nate Baum and Ron White of Samaritans for their help and expertise.

REFERENCES

Alison Bryant, J., Sanders-Jackson, A., and Smallwood, A. M. (2006). Iming, text messaging, and adolescent social networks. *Journal of Computer-Mediated Communication*, 11(2):577–592.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Bostock, M. (2012). D3. js. *Data Driven Documents*.

Chaney, A. J.-B. and Blei, D. M. (2012). Visualizing topic models. In *ICWSM*.

Christensen, H., Farrer, L., Batterham, P. J., Mackinnon, A., Griffiths, K. M., and Donker, T. (2013). The effect of a web-based depression intervention on suicide ideation: secondary outcome from a randomised controlled trial in a helpline. *BMJ open*, 3(6).

Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., Qu, H., and Tong, X. (2011). Textflow: Towards better understanding of evolving topics in text. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2412–2421.

Dinakar, K., Chaney, A. J., Lieberman, H., and Blei, D. M. (2014a). Real-time topic models for crisis counseling. *Twentieth ACM Conference on Knowledge Discovery and Data Mining, Data Science for the Social Good Workshop*.

Dinakar, K., Jones, B., Lieberman, H., Picard, R., Rose, C., Thoman, M., and Reichart, R. (2012). You too?! mixed-initiative lda story matching to help teens in distress.

Dinakar, K., Weinstein, E., Lieberman, H., and Selman, R. (2014b). Stacked generalization learning to analyze teenage distress. *International AAAI Conference on Weblogs and Social Media*.

Gardner, M. J., Lutes, J., Lund, J., Hansen, J., Walker, D., Ringger, E., and Seppi, K. (2010). The topic browser: An interactive tool for browsing topic models. In *NIPS Workshop on Challenges of Data Visualization*.

Gianfortoni, P., Adamson, D., and Rosé, C. P. (2011). Modeling of stylistic variation in social media with stretchy patterns. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, DIALECTS '11, pages 49–59, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gould, M. S., Cross, W., Pisani, A. R., Munfakh, J. L., and Kleinman, M. (2013). Impact of applied suicide intervention skills training on the national suicide prevention lifeline. *Suicide and Life-Threatening Behavior*, 43(6):676–691.

Gould, M. S., Munfakh, J. L., Kleinman, M., and Lake, A. M. (2012). National suicide prevention lifeline: enhancing mental health care for suicidal individuals and other people in crisis. *Suicide and Life-Threatening Behavior*, 42(1):22–35.

Hedgebeth, D. (2007). Data-driven decision making for the enterprise: an overview of business intelligence applications. *VINE*, 37(4):414–420.

Hoff, L. A., Hallisey, B. J., and Hoff, M. (2011). *People in crisis: Clinical and diversity perspectives*. Taylor & Francis.

Hoffman, M., Bach, F. R., and Blei, D. M. (2010). Online learning for latent dirichlet allocation. In *advances in neural information processing systems*, pages 856–864.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159–166. ACM.

Kinzel, A. and Nanson, J. (2000). Education and debriefing: Strategies for preventing crises in crisis-line volunteers. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 21(3):126–134.

Pallotta, V., Delmonte, R., Vrieling, L., and Walker, D. (2011). Interaction mining: the new frontier of call center analytics. *Proc. of DART*.

Pratt, M. E. (2013). *The Future of Volunteers in Crisis Hotline Work*. PhD thesis, University of Pittsburgh.

Ramage, D., Hall, D., Nallapati, R., and Manning, C. D. (2009). Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 248–256. Association for Computational Linguistics.

Rubin, T. N., Chambers, A., Smyth, P., and Steyvers, M. (2012). Statistical topic models for multi-label document classification. *Machine Learning*, 88(1-2):157–208.

Schwartz, R. C. and Rogers, J. R. (2004). Suicide assessment and evaluation strategies: A primer for counselling psychologists. *Counselling Psychology Quarterly*, 17(1):89–97.

Smith, A. (2012). 46% of american adults are smartphone owners. *Pew Internet & American Life Project*.

Wallach, H. M., Mimno, D. M., and McCallum, A. (2009). Rethinking lda: Why priors matter. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, *NIPS*, pages 1973–1981. Curran Associates, Inc.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., and Kessler, R. C. (2005). Twelve-month use of mental health services in the united states: results from the national comorbidity survey replication. *Archives of general psychiatry*, 62(6):629–640.