Analyzing Content of Patient-Physician messages with Self-Organizing Maps

Lic (Tech.) Karita H. Ilvonen – Helsinki University of Technology
MD Steven LeVine – Division of Research, Kaiser Permanente
MD, PhD. Joe Terdiman - Division of Research, Kaiser Permanente

Objective

- Part of an extensive research effort, in which our overall objective is to understand the impacts of patient-physician messaging service on provider efficiency and use of resources.
- We evaluate the usefulness of Self Organizing Maps (SOM) technology for analysis of unstructured patient-originated electronic messages.
- A one-by-one hand analysis of a subset of the messages was performed in order to evaluate the accuracy of the topographic method.
- Two specific research questions: 1) What broad categories of requests are submitted by the patients 2) how can the requests be understood in the context of health care information provisioning by primary care physicians to patients with a specific chronic disease.
- Topographic mapping of message types confirmed the patterns seen through manual analysis. Sub-mapping of messages by symptom themes allowed more granular understanding of the clinical decision processes involved.

Methods

- A study cohort of 281 diabetic patients receiving primary care at the Kaiser Permanente Medical Center in Oakland, California was defined.
- The cohort consisted of self-selected patients using the CyberKaiser electronic messaging system during a study period of one year (March 2003-February 2006).
- We combined Electronic Health Record Data, messaging free text data and demographic data for our analysis.
- Diabetics were defined on the basis of ICD-9 codes.

Methods 2

- We used Foundation™ software to analyze the unstructured text of the patient messages and categorized the messages manually into 12 different categories (e.g. appointment request, medical advice, lab result interpretation etc.) based on the initial requests made in them.
- We also recorded the number of request made in individual messages.
- The software used draws Self Organizing Maps (SOMs) from the unstructured data and helps us to circumscribe major categories and to identify the most frequently used terms ('themes') in the messages.

Self Organizing Map

- Self Organizing Map (SOM) was used for the broader context analysis. However, in-depth analysis of the complexity of the messages required categorizing them by hand.
- We examined the initial message of each thread send by the patient within the study period for review.
- Request types were categorized and statistical analysis performed on the results.
- Figure 1 presents and example graphic of a SOM used for the context analysis. This graphic organizes all messages sent by diabetic patients in the cohort to a map outlining five different areas. For example an area where the most common themes are refill, prescriptions and pharmacy includes mostly messages where patients request to refill their prescriptions.

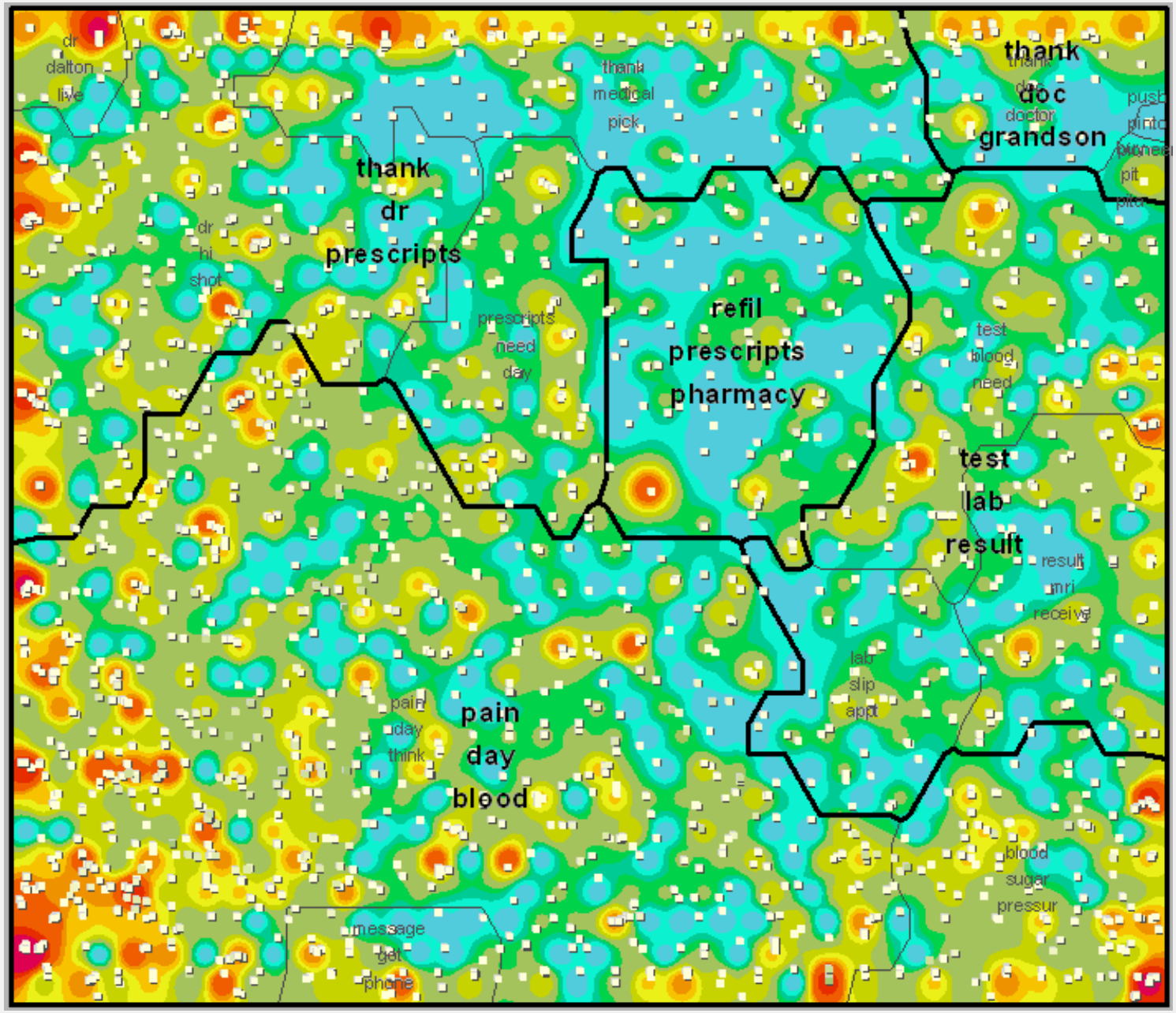

Figure 1: Self Organizing Map (SOM) of all messages sent by diabetic patients in cohort

Using a glossary

- The categorization presented in figure 1 is on a very abstract level because words that are likely to appear in a message with any topic like ‘thank’ and ‘dr’ are still included.
- A diabetes glossary was used to vertically classify the themes identified in unstructured data when using the Foundation Software.
- Vertical classification of unstructured data refers to the process of capturing unstructured data and organizing it so that it can be meaningfully accessed and analyzed according to specific topic(s) as identified in the glossary.
- Glossaries are lists of related words, their meanings and synonyms. We used a glossary of clinical and other diabetes related words in running the following SOM (figure 2).

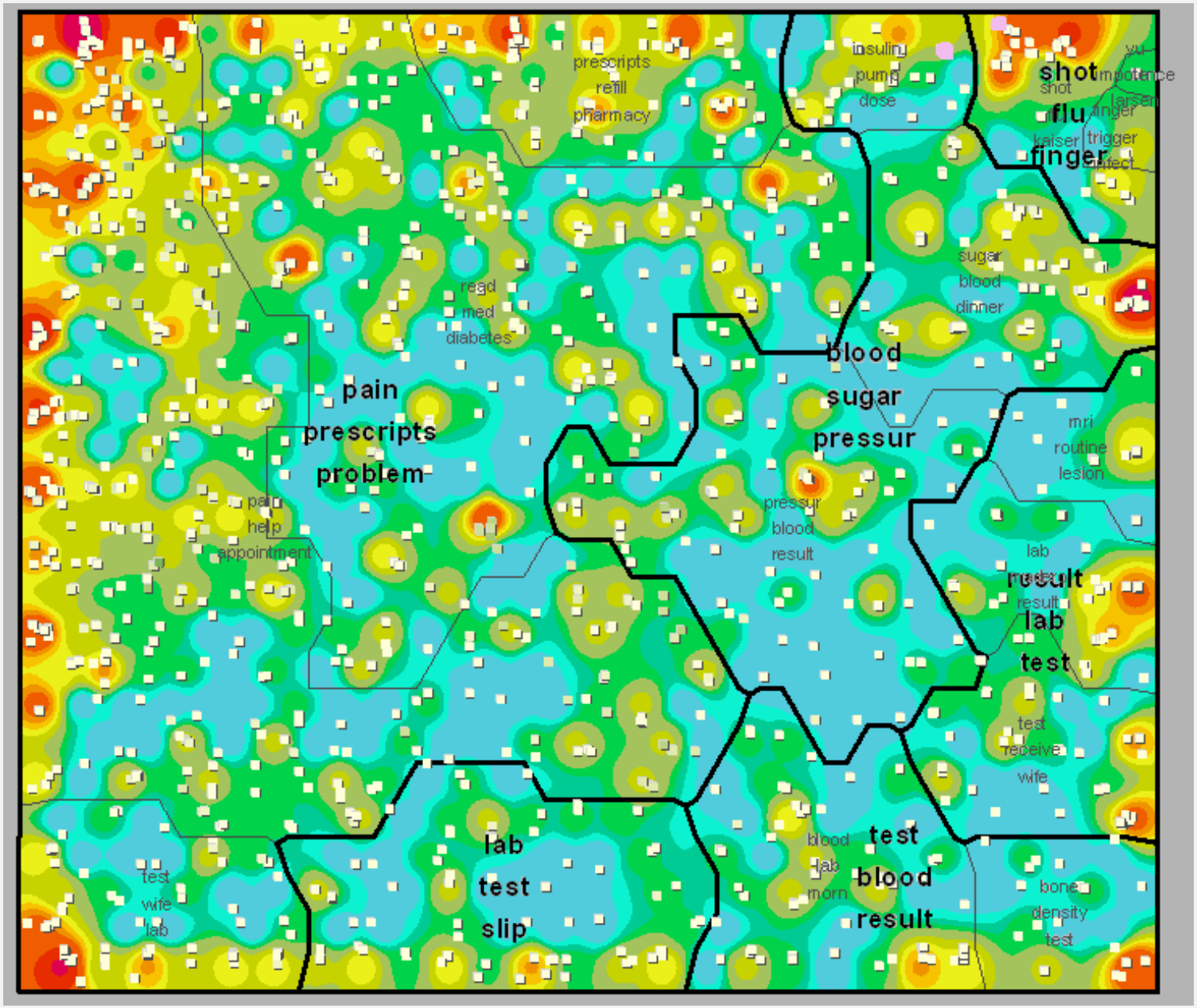

Figure 2: Diabetes messages using a glossary

Results

- The visualization of CyberKaiser messages send by diabetic patients grouped messages by common themes.
- The visualization recognized five large groups of messages talking about appointments, pain and symptoms, prescriptions, 'thank yous' and administrative issues.
- The topographic mapping software was also used for identifying 30 most commonly used words in the messages.
- The one-by-one hand analysis found that most messages have a single request on action on them and provide a lot of background information.

Conclusions

- Self-organizing maps allow parsing of large amounts of text information contained in patient-provider email.
- Context analysis findings and categorization of the requests are expected to advance the future development of these services and answer to some concerns on the messaging as an appropriate medium for healthcare communication..

For more information please contact:

Karita Ilvonen

karita.ilvonen@hut.fi

+1 415 794 0767