Towards A Task Taxonomy for Analyzing Electronic Health Record Cohorts from Mild Traumatic Brain Injuries

Xiaohui Bian, Lu Liu, Jesus J. Caban, *Member, IEEE*, Gaoqi He, and Jian Chen, *Member, IEEE*

Abstract— We present an initial study result towards a task taxonomy for visualization design for electronic health records (EHR) data analysis through a combination of the data-driven and user-driven approaches. We discuss the process involved in the generation and the initial results which contain more than 100 tasks. The taxonomy will be of value in the design and evaluation of visualization design and tool development of temporal EHR environment.

Index Terms—Task taxonomy, temporal heterogenous data, electronic health records, data- and user-driven methods

1 Introduction

An electronic Health Record (EHR) is the digital version of a patient's medical history, which is often maintained by the providers over time and may include all of the key administrative clinical data relevant to that person's care from a particular provider, including demographics, progress notes, problems, medications, vital signs, medical history, lab testing results, and so on. As a result, analyzing EHR produces great opportunity for patients and healthcare providers to understand patterns of illness and patterns of cares, and subsequently improve patient care by (1) reducing the incidence of medical error by improving the accuracy and clarify of medical records, (2) making the health information available, and (3) reducing medical errors. Due to the EHR collections in the past decades, it is for the first time that we can analyze patients not only by doctor's personal or the patient's own experience, but also by the sound data-driven pattern findings.

It is arguably that to support any design, analysis, and evaluation of EHR data, one must know the questions users want to answer. And such an answer list must address the *data* attributes: the EHR data is different from other temporal data in that the data are not only vary with time and space, but also heterogeneous in their temporal dimension. For example, two different person with the same disease may show a dramatically different hospital visit trajectory and physiology, but may show the same diagnosis, drug orders, and clinical outcomes. This example shows that pattern finding must determine what data attributes to include or ignore and this decision is often made *during* rather than *prior to* the data exploration. As a result, addressing and considering data collection imposes on the type of questions on can answer with a visualization tool.

Many temporal event taxonomy focuses on the visualization methods. For example, Bach et al. reviews the visualization method on space-time cube operations including many temporal visualization techniques and interactive methods (e.g., drill-up and -down abstraction, filtering) [2]. Several specific graph visualization methods are Shneiderman and Yi's and Bach et al.'s for graph visualizations. There are also generic task taxonomies, e.g., the Schneiderman's tasks by data types [13].

Inspired by these trends on task taxonomy, we have begun to study the task taxonomy for EHR cohort. Because data drives analysis and time and associated data from a part of *what* is being visualized, user tasks are related to the question *why* something is visualized and visual and interaction design address *how* data and tasks relate to each other when something is shown on the screen. We have constructed an initial prototypical task space using the data-driven Andreinco and Andrienco (AA) model [1]. This Andreinco and Andrienco model has led to successful extensions in temporal graph visualization [5] and interaction taxonomy [11].

Our proposed methodology for unearthing the task space for EHR data includes three broad stages: (1) the first data-driven stage is to perform the task analysis using the Andrienco and Andrienco method; (2) the second user-driven stage is to ask the stakeholders (medical professionals and patients) to rank and explain the usefulness of these tasks to aid their data analysis; and (3) the third stage to rank, merge and release the task space to the community. The major contribution in this paper is to report the task space we have derived from the first data-driven stage. We will make the task list available to the workshop participants to collect their suggestions on the methodology and usefulness of our results. We will also collect some preliminary data at the workshop.

2 Related Work

Temporal EHR data analysis. Many temporal pattern analysis tools have addressed important pattern search from EHR data. Fails et al. define the temporal patterns as sequences of events with interevent time spans, define attributes of the events [4]. Their tool takes a multivariate and categorical nature of events and defines a temporal pattern as a sequence of point events separated by time span. That tool supports within and across data entries in patient histories. For EHR data and other temporal event analysis, Wongsuphassawat and Gotz designed [14] and studied [15] Outflow and suggest that temporal trajectories should be aggregated otherwise would generate many branches. Their tasks evaluated include travels graph using labels, interpret proportion, outcome, and graph traversal tasks. Perer, Wang, and Hu studied correlation by integrating and combining mining and visual analysis approaches [9]. LifeLines provides a personal record browsing mechanism in a hierarchical timeline visualization to show a person's medical histories organized by facets, such as hospital visits, diagnosis, medications, and lab testing outcomes [10]. The user can adjust the time scale, filter records, zoom and accessing details on demand and encoding methods such as line thickness and color show event attributes (e.g., significance and relationships to other events).

Task spaces. There exist many task space in general and for temporal event visualization techniques. Schulz et al. propose a general design space of visualization tasks in information visualization domain with the aim to establish recurring tasks and to improve the design and evaluation of visualization techniques [12]. That task list is a meta-list which states task's goal (exploratory or confirmatory analysis or presentation), means (how is a talk being carried out? e.g, navigation, re-organization or relation), characteristics (what does a task seek?), target and cardinality of data entries (where in the data does a task op-

- *Xiaohui Bian, Lu Liu, and Jian Chen are with University of Maryland, Baltimore County, E-mail: liulumy, jichen@umbc.edu.*
- *Jesus J. Caban is with Walter Reed National Military Medical Center. E-mail: jesus.j.caban.civ@mail.mil.*
- *Xiaohui Bian and Gaoqi He are with East China University of Science and Technology. E-mail: gaoqihe@gmail.com.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

erate?), order of the tasks (when is a task performed?) and the user types (who is executing a task?) While this list serves as a high-level description with an attempt to put all tasks into a common hood, our task analysis is intended to link task to data attributes in a specific domain in order to prompt domain-specific design to address critical issues in EHR analysis.

Miksch and Aigner connects data, user, and tasks in a triangle to analyze temporal events to show how defining this data-user-task triangle has benefits the visualization process to address a given task [8]. Our task is similar to theirs in terms of linking these three component and different from their methodology in that we study the task space formation instead of the design aspects. Similar to ours, Lammarsch et al. also make use of AA framework for task analysis along the structure of time [6], categorized according to scale, scope, arrangement, viewpoints, granularities, time primitives, determinacy,

Several domain-specific tasks are useful. For example, Macachren presents a list of tasks in maps: existence of an entity, temporal location, time interval, temporal texture, rate of change, sequence, and synchronization [7]. Andrienko and Andrienko also provides several geospatial data lists using their own AA framework [1].

For the design of any interactive system, it is also crucial to consider the *task granularity*, which defines whether or not a visualization system design execute a goal concurrently as a separate task. In general, tasks that can be done in parallel do not mean that they should be done in parallel. This is because finer level of granularity incurs various overheads, e.g., overheads associated with fluid connection between tasks, and the associated communication overheads between tasks. One should be noted that this task granularity is different from the data granularity which can mean temporal scales in temporal data analysis (e.g., tasks by hour, or day or month). Roth describes five objective primitives: identify, compare, rank, associate, and delineate [11]. Theoretically, these verbs can be integrated and combined with the AA framework to produce more tasks. Our method also follows Roth's to actually validate our task space with real users. Ours also considers domain-specificity to produce task useful to the EHR domain related to preparing, cleaning-up, saving and editing and annotating as instances of enabling tasks, similar to those presented in the Brehmer and Munzner paper [3].

3 Task Taxonomy

This section describes our prototypical task list formation.

3.1 Initial Exploration

A human-curated dataset of milld traumatic brain injury (mTBI) with 5,673 clinical encournters about 45 patient history stamped ICD code9 and also other categories of data such as demographics information (patient id, gender, age, age_group), injury information (age when the first TBI occurred, days from first TBI, data of injury, type of injury code), encounter information (encounter id and data and provider's speciality, type and line), other flags (TBI $encounter_flag$, $warRelated_flog$, how many days before and after the injury), diagnosis (stress, $post-traumatic$ stress disorder or PTSD, speech, anxiety, depression, headache, sleep, audiology, vision, neurologic, Alzheimer, cognitive, PCS, endocrine, skull injury and non skull injury). This dataset is curated for two main concerns with a military hospital; These patients represent those who stayed at the military hospital for a three-months treatment of PTSD. Table 1 shows an example patient entry. This dataset is securely transferred to the University of Maryland, Baltimore County for subsequent analysis.

3.2 The Data Model

The AA framework starts at the most general and abstract level to first identify data facets and derives *referential* and *characteristic* components of the data, and then derive a rather complete set of tasks by a formal set of rules. For example, a *look-up* task will be phrased like "look for the characteristics at a given reference". The referential components (aka referrers) are how the data is organized which is similar to the header row in the excel spread sheet. The characteristics are components in data.

Table 1. An Example Entry in the EHR mTBI Database.

Property	Value
patient ID	1234
Gender	Female
Age	29
Age Group	25 to 34
Age_{TBI}	35
Days from the 1^{st} TBI	-2302
Date of Injury	20150120
Type of injury code	NSFINJ
Encounter id	533116955
Encounter date	20081001
Encounter source	$DC-OUTPT$
Provider speciality	Nurse practitioner (obstetrics and Gynecology)
$Pre_{maxdays}$	-3007
$Post_{maxdays}$	223
Diagnosis	N/A

3.3 The Task Model

The AA framework simplifies the Bertin's three-level reading to two-levels to define elementary and synoptic tasks. Here, elementary tasks are those targeting single objects whilst synoptic ones concern a collection of objects. We have derived a set of tasks. Note that often the diagnoses below can be replaced. As a result, the tasks below are only a subset of a complete taxonomy.

3.3.1 Lookup tasks

Direct lookup

What are symptoms of patient A when the daysFrom1stTBI is N?

Inverse lookup

Which patients have the symptoms X, and when?

Who have the symptom X when the daysFrom1stTBI is N?

When does patient A have the symptom X?

3.3.2 Comparison tasks

Direct comparison

Compare the symptoms that patient A has when the daysFrom1stTBI is N with the symptoms X.

Compare the symptoms that patient A has when the days−from−1st−TBI is N with the symptoms that patient B has when the daysFrom1stTBI is M.

Inverse comparison

Compare the patients who have the symptoms X with the patients who have the symptoms Y.

Compare the patients who have the symptoms X when the daysFrom1stTBI is N with the patients who have the symptoms Y when the daysFrom1stTBI is M.

Compare the daysFrom1stTBI when patient A has symptoms X and symptoms Y.

3.3.3 Relation seeking tasks

Which patients have the same symptoms with other patients from the same age group when their daysFrom1stTBI are the same?

Which patients have the same symptoms with the symptoms that patient A has when the daysFrom1stTBI is N?

Find the patients whose symptoms at the daysFrom1stTBI N have no overlap with their symptoms at the daysFrom1stTBI M?

3.4 Synoptic tasks

3.4.1 Pattern identification task

Pattern definition

As for patient A, which symptom occurs most often in the specific period?

What is the symptoms transfer trajectory of patient A (all patients) in the specific period?

What is the frequency distribution of symptoms of patient A in the specific period?

As for all the patients, which symptom occurs most often when the daysFrom1stTBI is N?

As for the female patients, which symptom occurs most often when the daysFrom1stTBI is N?

As for the patients of the specific age group, which symptom occurs most often when the daysFrom1stTBI is N (in the specific period)?

What is the frequency distribution of symptoms of all the (female) patients when the daysFrom1stTBI is N?

What is the frequency distribution of symptoms of the patients of the specific age group when the daysFrom1stTBI is N?

As for all the (female) patients, which symptom occurs most often in the specific period?

What is the symptoms transfer trajectory of the female patients in the specific period?

What is the frequency distribution of symptoms of the female patients (of a specific age group) in the specific period?

Pattern search

As for patient A, in which period does the specific symptom occur most often?

In which period does the frequency distribution of symptoms of patient A conform to the specific distribution?

In which period does the symptoms transfer trajectory of patient A conform to the specific trajectory?

As for patient A, on which day does the specific symptom occur most often?

To the patients of which age group, the specific symptom occurs most often when the daysFrom1stTBI is N?

To the patients of which gender, the specific symptom occurs most often when the daysFrom1stTBI is N?

To the patients of which age group, the frequency distribution of symptoms conforms to the specific distribution when the daysFrom1stTBI is N?

To the patients of which gender, the frequency distribution of symptoms conforms to the specific distribution when the daysFrom1stTBI is N?

To the patients of which age group, the amount of symptoms is the most when the daysFrom1stTBI is N?

To the patients of which gender, the amount of symptoms is the most when the daysFrom1stTBI is N?

To the patients of which age group, the specific symptom occurs most often in the specific period?

To the patients of which gender, the specific symptom occurs most often in the specific period?

To the patients of which age group, the frequency distribution of symptoms conforms to the specific distribution in the specific period?

To the patients of which gender, the frequency distribution of symptoms conforms to the specific distribution in the specific period?

To the patients of which age group, the symptoms transfer trajectory conforms to the specific trajectory in the specific period?

To the patients of which gender, the symptoms transfer trajectory conforms to the specific trajectory in the specific period?

To the patients of which age group, the amount of symptoms is the most in the specific period?

To the patients of which gender, the amount of symptoms is the most in the specific period?

3.4.2 Behavior comparison

Direct comparison We have omitted the direct comparison tasks which can be founded online at https://docs.google.com/document/d/1AmPEIuo42NTL-vU7LTj9LyOTuCrzeUMhofKmtzmnlSQ/edit?usp=sharing. ‘

Inverse comparison

What is the relationship between the period when the specific symptom occurs most often of patient A and the specific period?

What is the relationship between the period when the frequency distribution of symptoms of patient A conform to the specific distribution and the specific period?

What is the relationship between the period when the symptoms transfer trajectory of patient A conform to the specific trajectory and the specific period?

What is the relationship between the time point when patient A has the most symptoms and the specific time point?

What is the relationship between the age group that to the patients of which, the specific symptom occurs most often and the specific age group when the daysFrom1stTBI is N?

As for the female patients, dose the specific symptom occur most often when the daysFrom1stTBI is N?

What is the relationship between the age group that to the patient of which, the frequency distribution of symptoms conforms the specific distribution when the daysFrom1stTBI is N and the specific age group?

Dose the female patients? frequency distribution of symptoms conform the specific distribution when the daysFrom1stTBI is N?

What is the relationship between the age group that the patients of which have the most symptoms and the specific age group when the daysFrom1stTBI is N?

Do the female patients have the most symptoms when the daysFrom1stTBI is N?

What is the relationship between the age group that to the patients of which, the specific symptom occurs most often and the specific age group in the specific period?

As for the female patients, dose the specific symptom occur most often in the specific period?

What is the relationship between the age group that to the patients of which, the frequency distribution of symptoms conforms the specific distribution in the specific period and the specific age group?

Dose the female patients? frequency distribution of symptoms conform the specific distribution in the specific period?

What is the relationship between the age group that to the patients of which, the symptoms transfer trajectory conforms the specific trajectory in the specific period and the specific age group?

Dose the female patients? symptoms transfer trajectory conform the specific trajectory in the specific period?

What is the relationship between the age group that the patients of which have the most symptoms and the specific age group in the specific period?

Do the female patients have the most symptoms in the specific period?

What is the relationship between the period when the specific symptom occurs most often of patient A and the period when the specific symptom occurs most often of patient B?

What is the relationship between the period when the frequency distribution of symptoms of patient A conform to the specific distribution and the period when the frequency distribution of symptoms of patient B conform to the specific distribution?

What is the relationship between the period when the symptoms transfer trajectory of patient A conform to the specific trajectory and the period when the symptoms transfer trajectory of patient B conform to the specific trajectory?

What is the relationship between the time point when patient A has the most symptoms and the time point when patient B has the most symptoms?

What is the relationship between the age groups that to the patients of which, the specific symptom occurs most often when the daysFrom1stTBI is N and M?

Are the genders that to the patients of which the specific symptom occurs most often the same when the daysFrom1stTBI is N and M?

What is the relationship between the age groups that to the patients of which, the frequency distribution of symptoms conforms to the specific distribution when the daysFrom1stTBI is N and M?

Are the genders that to the patients of which the frequency distribution of symptoms conforms to the specific distribution the same when the daysFrom1stTBI is N and M?

What is the relationship between the age groups that the patients of which have the most symptoms when the daysFrom1stTBI is N and M?

Are the genders that the patients of which have the most symptoms when the daysFrom1stTBI is N and M the same?

What is the relationship between the age groups that to the patients of which, the specific symptom occurs most often in the period a and b?

Are the genders that to the patients of which the specific symptom occurs most often the same in the period a and b?

What is the relationship between the age groups that to the patients of which, the frequency distribution of symptoms conforms to the specific distribution in the period a and b?

Are the genders that to the patients of which the frequency distribution of symptoms conforms to the specific distribution the same in the period a and b?

What is the relationship between the age groups that to the patients of which, the symptoms transfer trajectory conform the specific trajectory in the period a and b?

Are the genders that to the patients of which the symptoms transfer trajectory conform the specific trajectory the same in the period a and b?

What is the relationship between the age groups that the patients of which have the most symptoms in the period a and b?

Are the genders that the patients of which have the most symptoms in the period a and b?

What is the relationship between the period when the specific symptom occurs most often of patient A and the specific time point?

What is the relationship between the period when the frequency distribution of symptoms of patient A conform to the specific distribution and the specific time point?

What is the relationship between the period when the symptoms transfer trajectory of patient A conform to the specific trajectory and the specific time point?

What is the relationship between the time point when patient A has the most symptoms and the specific period?

What is the relationship between the age group that to the patients of which, the specific symptom occurs most often and the specific age when the daysFrom1stTBI is N?

What is the relationship between the age group that to the patient of which, the frequency distribution of symptoms conforms the specific distribution when the daysFrom1stTBI is N and the specific age?

What is the relationship between the age group that the patients of which have the most symptoms and the specific age when the daysFrom1stTBI is N?

What is the relationship between the age group that to the patients of which, the specific symptom occurs most often and the specific age in the specific period?

What is the relationship between the age group that to the patients of which, the frequency distribution of symptoms conforms the specific distribution in the specific period and the specific age?

What is the relationship between the age group that to the patients of which, the symptoms transfer trajectory conforms the specific trajectory in the specific period and the specific age?

What is the relationship between the period when the specific symptom occurs most often of patient cohort A and the time point when patient A has the specific symptom?

What is the relationship between the period when the frequency distribution of symptoms of patient A conform to the specific distribution and the time point when patient A has the specific symptom?

What is the relationship between the period when the symptoms transfer trajectory of patient A conform to the specific trajectory and the time point when patient A has the specific symptom?

What is the relationship between the period when the specific symptom occurs most often of patient A and the time point when patient A has the most symptoms?

What is the relationship between the period when the frequency distribution of symptoms of patient A conform to the specific distribution and the time point when patient A has the most symptoms?

What is the relationship between the period when the symptoms transfer trajectory of patient A conform to the specific trajectory and the time point when patient A has the most symptoms?

Relation seeking task

As for patient A, in which period is the symptoms transfer trajectory similar to the previous period?

When the daysFrom1stTBI is N, to the patients of which age group is the symptom that occurs most often the same with the patients who are three years younger than them?

When the daysFrom1stTBI is N, to the patients of which age group is the frequency distribution of symptoms similar with the patients who are three years younger than them?

In the specific period, to the patients of which age group is the symptom that occurs most often the same with the patients who are three years younger than them?

In a specific period, to the patients of which age group is the frequency distribution of symptoms similar to the patients who are in different age groups.

In a specific period, to the patients of which age group is the symptoms transfer trajectory similar with the patients who are three years younger than them? ' As for patient A, in which period is the symptoms transfer trajectory similar to the specific period?

4 Conclusion

We have reported an approach to task analysis of temporal EHR data. Our next step is to continue our collaboration with end users and revise and evaluate the task list.

Acknowledgments

This work was supported in part by a grant from the DoD: USAMRAA-13318046.

References

[1] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.

[2] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A review of temporal data visualizations based on space-time cube operations. In *Eurographics conference on visualization*, 2014.

[3] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.

[4] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pp. 167–174. IEEE, 2006.

[5] N. Kerracher, J. Kennedy, and K. Chalmers. A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 21(10):1160–1172, 2015. doi: 10.1109/TVCG.2015.2424889

[6] T. Lammarsch, A. Rind, W. Aigner, and S. Miksch. Developing an extended task framework for exploratory data analysis along the structure of time. In *Proc. Eurographics International Workshop on Visual Analytics (EuroVA)*. Citeseer, 2012.

[7] A. M. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 1995.

[8] S. Miksch and W. Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014.

[9] A. Perer, F. Wang, and J. Hu. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of biomedical informatics*, 56:369–378, 2015.

[10] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman. Lifelines: using visualization to enhance navigation and analysis of patient records. In *Proceedings of the AMIA Symposium*, p. 76. American Medical Informatics Association, 1998.

[11] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2356–2365, 2013.

[12] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[13] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pp. 336–343. IEEE, 1996.

[14] K. Wongsuphasawat and D. Gotz. Outflow: Visualizing patient flow by symptoms and outcome. In *IEEE VisWeek Workshop on Visual Analytics in Healthcare*, pp. 25–28. American Medical Informatics Association, 2011.

[15] K. Wongsuphasawat and D. Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2659–2668, 2012.