Exploring Interval Graphs of Rare Diseases in Retrospective Analysis of Outpatient Records

Svetla Boytcheva

Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, Bulgaria

svetla.boytcheva@gmail.com

Abstract. This paper deals with investigation of complex temporal relations between some rare disorders. It proposes an interval graphs approach combined with data mining for patient history pattern mining. The processed data are enriched with context information. Some text mining tools extract entities from free text and deliver additional attributes beyond the structured information about the patients. The test corpora contain pseudonymised reimbursement requests submitted to the Bulgarian National Health Insurance Fund in 2010-2015 for more than 5 million citizens yearly. Experiments were run on 2 data collections. Findings in these two collections are discussed on the basis of comparison between patients with and without rare disorders. Exploration of complex relations in rare-disease data can support analyzes of small size patient pools and assist clinical decision making.

Keywords: Big Data Analytics, Knowledge Discovery, Interval Graphs, Data Mining, Clinical Text processing.

1 Motivation

The importance of rare disorders research is underestimated. Usually they are considered as low prevalence disorders and small size patient pools that causes less attention. According to the recent research when rare diseases are considered together (approx. 6,000-8,000 depending on the granularity of disease definition) about 5% of the world population suffers from some of them [1]. For some of the rare diseases there are not available standardize codes [2] according the ICD-10[1]. Currently one ICD-10 code can refer to a group of different rare disorders. Recently there is no systematic review of the rare diseases, except for genetic related research. Another obstacle in rare disorders study is their phenotypic heterogeneity. The majority of deceases in this category are genetic or congenital malfor-

[1] International Classification of Diseases and Related Health Problems 10th Revision. http://apps.who.int/classifications/icd10/browse/2015/en

mations, but there are also some rare cancers, auto-immune, toxic and infectious diseases. The first two categories are usually diagnosed in childhood. About 50% of the rare disorders are diagnosed in the adulthood[2] and we will focus our study on them.

For investigation of rare diseases complex temporal relations we can use retrospective analysis of population data. We need to filtering events with common properties and similar significance.

Two of the widely used data mining (DM) approaches for patterns search are [3]: (i) frequent pattern mining (FPM) viewing the events (objects) as unordered sets and (ii) frequent sequence mining (FSM). The difference between them is that in the first case the event order does not matter. Most FPM and FSM methods do not consider contextual information about extracted patterns. Further development of such DM methods is needed [4].

Another hot issue in medical DM is harnessing unstructured data from clinical narratives. They are an underused data source that has much greater research potential than is currently realized [5,6]. Advanced research projects apply NLP as a first step in mining entities from free text and use the latter as input to subsequent DM tasks.

Application of data mining techniques is not appropriate in the task for rare disorders analysis due to the small size of the patient pool. Such techniques like discovering frequent patterns of diseases can be applied to relatively big repository of data.

We need to consider not only the partial order between different episodes, but also their overlapping and the duration of different groups of disorders coexistence. For modeling events and their relations we propose approach based on interval graphs [7]. For each patient history is generated individual interval graph for episodes in different windows. The generated interval graphs are mapped onto interval graphs of patients without rare disorders. The experiments explicate some population specific relations. We also discuss the effects of context information - age, gender and demographics on temporal relations.

The paper is structured as follows. Section 2 presents related work especially in the areas of pattern search and interval graphs. Section 3 presents the data corpora we use, Section 4 introduces the methods. Section 5 discusses current experiments and results. Section 6 sketches further work and the conclusion.

2 Related Work

One of the first attempts for mining patients' history in big data scope – processing over 1.6 million of patient histories was reported by Patnaik et al. [8,9]. Three tasks are addressed in their research: mining parallel episodes, tracking serial extensions, and learning partial orders. Patnaik et al. [9] present streaming algorithm for mining frequent episodes over a window of recent events in the stream. The system EMRView for mining the precedence relationships to identify and visualize partially ordered information was demonstrated.

Many research efforts for visualization and analysis of periodical data for single patient or searching patterns for a cohort of patients we reported in [10,11,12]. Monroe et al. [13] presents a system with visual tools with application in Electronic Health Records (EHRs) that allows the user to narrow iteratively the process for mining patterns to the desired target.

Other research achievements are in the area of temporal events presentation and mining. Lee et al [14] proposes a method for temporal event matrix representation and a learning framework that discovers complex latent event patterns or Diabetes Mellitus complications. Yang et al. [15] present

[2] The International Rare Diseases Research Consortium (IRDiRC)
http://www.irdirc.org/rare-diseases-research/

application of temporal event sequence mining for mining patient histories. Gyet and Quiniou [16] propose recursive depth-first algorithm QTIPrefixSpan that explores the extensions of temporal patterns. Further they extract temporal sequences with quantitative temporal intervals with different models using a hyper-cube representation and develop a version of EM algorithm for candidates' generation [17].

Interval graphs and interval trees have many real applications in modeling time dependences. Such applications are in protein sequencing, job scheduling, file organization and so on [23].

The classic FPM algorithms generate all possible frequent patterns (FPs). Summarized information for data relations can be extracted as maximal frequent itemsets (MFI). One of the first such algorithms MFCS is proposed by Dao-I Lin, Zvi M. Kedem who combine two way searches – top-down and bottom-up [18]. Other efficient algorithms for MFI are GenMax [19], FPMax [20], MAFIA [21] and NSFI [22].

3 Materials

We deal with a repository of pseudoanonymous Outpatient Records (OR) in Bulgarian language provided by the Bulgarian National Health Insurance Fund (NHIF) in XML format. The data repository currently contains more than 224 million ORs submitted in 2010-2015 for more than 5 million citizens yearly. In Bulgaria ORs are produced by the General Practitioners and the Specialists from Ambulatory Care for every contact with the patient. Only the administrative information is presented in structured XML format. It contains data for date and time of the visit; pseudonymised personal data, age, gender; visit-related information; diagnoses in ICD-10[3]; NHIF drug codes for medications that are reimbursed; a code if the patient needs special monitoring, etc. The free text OR fields processed by NLP tools include information for anamnesis (case history, previous treatments. family history, risk factors), status- (current patient state, height, weight, BMI, blood pressure etc.), clinical tests, and prescribed treatment.

Our experiments for rare disorders relations exploration consider two separate collections of ORs. They contain data about patients suffering from Schizophrenia (ICD-10 code F20) and Hyperprolactinaemia (ICD-10 code E22.1). Hyperprolactinaemia is classified as rare disorder. The collections are extracted by using a Business Intelligence tool (BITool) [4] from the repository of ORs for approx. 5 million patients for a 3-years period.

4 Methods

System architecture is shown on Fig. 1. For ORs preprocessing we use some text mining modules to convert the raw text descriptions to structured event data. We developed a drug extractor using regular expressions to describe linguistic patterns [24] and numerical value extractor [25] for clinical lab test and status values extraction like body mass index (BMI), weight, heigh, blood pressure (Riva Roci – RR), etc.

[3] International Classification of Diseases and Related Health Problems 10th Revision. http://apps.who.int/classifications/icd10/browse/2015/en

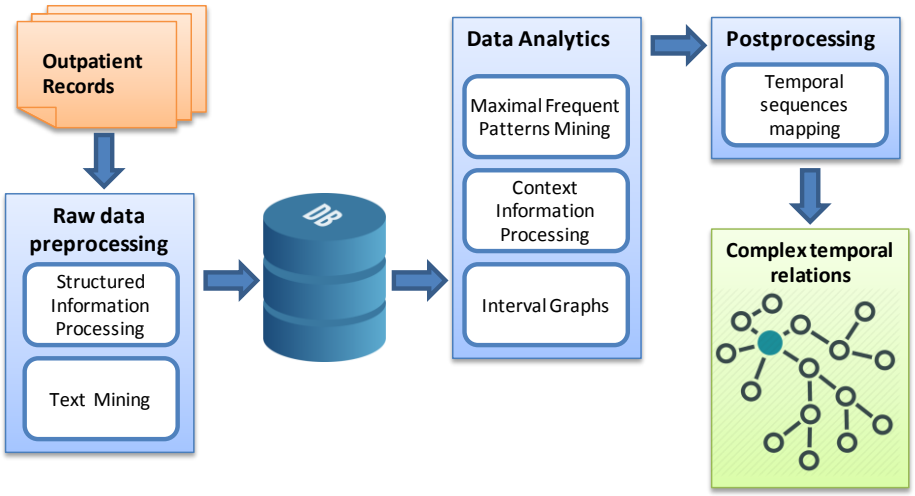

Fig. 1. System Architecture

In preprocessing step we accumulate for each patient its history. For the collection S of ORs we extract the set of all different patient identifiers $P = \{p_1, p_2, \dots, p_N\}$. This set corresponds to transaction identifiers (*tids*) and we call them *pids* (patient identifiers). We consider each patient visit to a doctor (i.e. each OR) as a single event. For each patient $p_i \in P$ an event sequence of tuples $\langle event, timeInterval \rangle$ is generated: $E(p_i) = \left(\langle e_1, t_1 \rangle, \langle e_2, t_2 \rangle, \dots, \langle e_{k_i}, t_{k_i} \rangle\right)$, $i = \overline{1, N}$. The time interval for a event is generated with starting time the date visit of ORs. Note that ORs always contain dates. The end time for chronic diseases[4] is considered the last date in the period of the ORs collection. For non-chronic disorders the end time is the time for next visit to the phisitian (specialist or general practitioner) who is qualified to assign this diagnosis. In case there is no OR for such visit the end time is associated like for chronic diseases. All dates in time intervals are replaced with relative time – to the first event in the sequence we assign time 0, and for all other events the timestamp is converted to the number of days distance from the first event. Let $\mathcal{E}$ be the set of all possible events and $\mathcal{T}$ be the set of all possible times intervals.

The next step includes a pipeline of several data analytics – aplying algorimt for MFI mining for comorbidities in the collections and context information processing. For patients with rare diseases are generated interval graphs.

4.1 Mining Maximal Frequent Itemsets and Context information Processing

In this step we are searching for as many as possible associations between chronic diseases – so called comorbidities. Let $\mathcal{C} = \{c_1, c_2, \dots, c_p\}$ be the set of all chronic diseases, which we call *items*. Each subset of $X \subseteq \mathcal{C}$ is called an *itemset*. We define a projection function $\pi: (\mathcal{E} \times \mathcal{T})^N \to \mathcal{C}^N: \pi\big(E(p_i)\big) = C(p_i) = \left(c_{1i}, c_{2i}, \dots, c_{m_i}\right)$, such that for each patient $p_i \in P$ the projected time sequence contains only the first occurrence (onset) of each chronic disorder recorded in $E(p_i)$. Let $D \subseteq P \times \mathcal{C}$ be the set of all itemsets in our collection after projection π in the format $\langle pid, itemset \rangle$. We will call D *database*. We are looking for itemsets $X \subseteq \mathcal{C}$ with frequency ($\sup(X)$) above given $minsup$. Let $\mathcal{F}$ denote the set of all frequent itemsets, i.e. $\mathcal{F} = \{X | \, X \subseteq \mathcal{C} \; and \; \sup(X) \geq minsup\}$. A frequent item-

[4] Chronic diseases, WHO, http://www.who.int/topics/chronic_diseases/en/

set $X \in \mathcal{F}$ is called *maximal* if it has no frequent supersets. Let $\mathcal{M}$ denote the set of all maximal frequent itemsets, i.e. $\mathcal{M} = \{X | \, X \in \mathcal{F} \; and \nexists \, Y \in \mathcal{F}, such \; that \; X \subset Y\}$. Then each subset of $X \in \mathcal{F}$ is also frequent itemset.

We apply a tabular method MixCO [4] for MFI mining using a vertical database, depth-first traversal as well as set intersection and diffsets. The context information is represented as attribute-value tuples for each patient. For the set of MFI $\mathcal{M}$ with cardinality $|\mathcal{M}| =$ K we have K classes of comorbidities. We apply classification of multiple classes in order to generate rules for each comorbidity class. We use large scale multi class classification because we deal with a big database and a large group of comorbidity classes. We use SVM and optimization based on block minimization method described by Yu et al. [26, 4].

4.2 Interval graphs

Definition: *A graph G(V,E) is an* ***interval graph*** *if there exists a collection of intervals V such that for any* I_i , $I_j \in V$, $I_i \neq I_j$ *there is an edge* $a_{ij} \in E$ *incidence with intervals* I_i *and* I_j *if and only if* $I_i \cap I_j \neq \emptyset$ *, i.e. there is an intersection (overlapping) between intervals* I_i *and* I_j .

Let $\mathcal{B} = \{b_1, b_2, \dots, b_p\}$ be the set of all diseases. We define a projection function $\chi: (\mathcal{E} \times \mathcal{T})^N \rightarrow (\mathcal{B} \times \mathcal{T})^N$: $\chi\big(E(p_i)\big) = B(p_i) = (\langle b_{i1}, t_{i1}\rangle, \langle b_{i2}, t_{i2}\rangle, \dots, \langle b_{im}, t_{im}\rangle)$, such that for each patient $p_i \in P$ the projected time sequence contains only the first occurrence (onset) of each chronic disorder recorded in $E(p_i)$.

We define individual interval graph $G_i(V_i, E_i)$ for each patient $p_i \in P$, $i = \overline{1,N}$, where time intervals of events are normalized to relative time and $V_i = B(p_i)$. There is an edge between events $\langle b_k, t_k\rangle$, and $\langle b_m, t_m\rangle$, $m \neq k$ if and only if relative time intervals overlap $t_k \cap t_m \neq \emptyset$.

The interval graph is generated using naïve approach with logarithmic time complexity. All events and time stamps are sorted in the database by start date.

Example: In this section we introduce a simple synthetic example (first two columns of Table 1) to illustrate the proposed method. Let assume that the period is one year with the last date Dec 31. The events and time intervals after normalization are presented in the last two columns of Table 1 and on Fig. 2. The generated interval graph is presented on Fig. 3, where E22.1 and M05.0 are ICD-10 codes of rare diseases.

Table 1. Example database for diseases with time intervals

Date	ICD-10 codes in OR
Jan 14	N95.1,E22.1,N60.9
Jan 20	E22.1
Jan 21	H53.8
Jan 22	R03.0
May14	E22.1, J11.8
Jul 03	E22.1,E11.9,M06.4
Aug 20	M05.0
Sep 25	E22.1,E11.9,M06.4
Oct 23	E22.1,E11.9,M06.4

ICD-10	Time interval
N95.1	[14,20)
N60.9	[14, 20)
E22.1	[14,365)
H53.8	[21,365)
R03.0	[22,365)
J11.8	[134,184)
E11.9	[184,365)
M06.4	[184,365)
M05.0	[232,365)

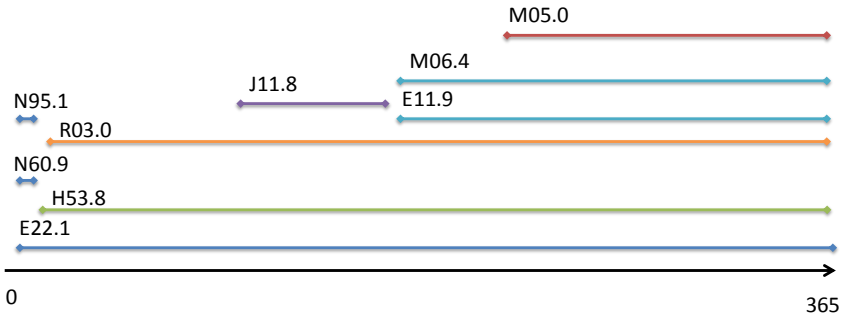

Fig. 2. Example for events with time intervals

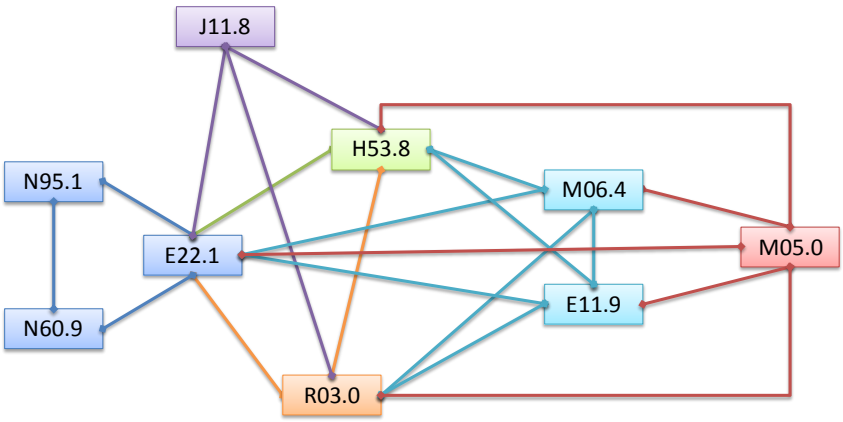

Fig. 3. Example for the generated interval graph

4.3 Temporal sequences mapping

In this task we use the results generated from both MFI and interval graphs. MFI cluster the collection on subsets depending on their support (Fig. 4). We map consecutively the interval graphs onto each of these clusters. For each interval graph we solve the task for induced subgraph isomorphism. Here we use the polynomial time solution proposed by Heggernes in [27].

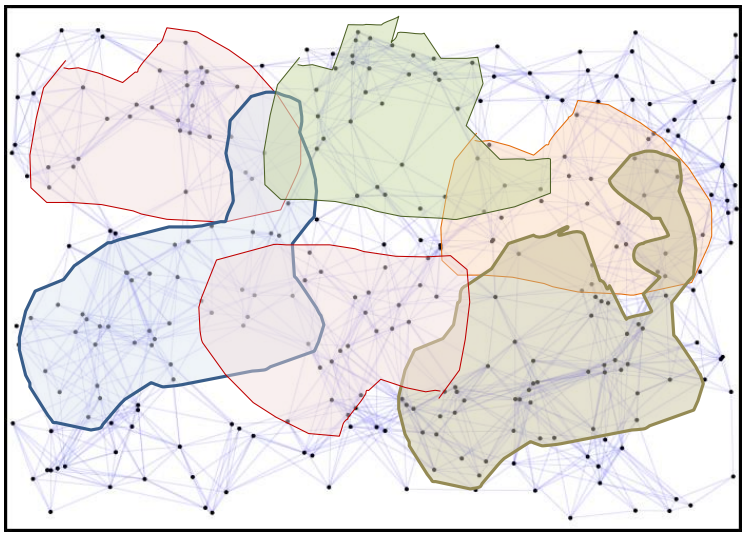

Fig. 4. MFI support in the collection

Algorithm for sequences mapping

//Initial Call: *Collection* S, *The set DV of MFI in the collection* S, $B(\mathrm{p})$ – event sequence for patient p

```
1   D ← {b_k|⟨b_k, t_k⟩ ∈ B(p)} //all deceases in the events in B(p)
2   Generate interval graph G for B(p)
3   foreach M ∈ DV
4     if D ∩ M ≠ ∅ then
5     K ← sup( M )
6       foreach k ∈ K
7           generate interval graph G_k for B(k)
8           Find induced subrgaph H of G isomorhic to G_k
9           R ← R ∪ H
10  Return R
```

5 Experiments and Findings

Our experiments for pattern search are made on two ORs. They contain data about patients suffering from Schizophrenia (ICD-10 code F20) and Hyperprolactinaemia (ICD-10 code E22.1). These collections are of primary interest for our project because Hyperprolactinaemia is one of the rare diseases with high prevalence. Schizophrenia is characterized as one of the chronic diseases with high heterogeneous variety of comorbidities. The results of MIxCO for generated MFI with $minsup = 80$ for S1 and $minsup = 45$ for S2 are presented in Table 4.

Table 2. Generated MFI from data collections S1 and S2

Collection	S1	S2
Main diagnosis	Schizophrenia	Hyperprolactinaemia
ICD10 code	F20	E22.1
Patients number	45,945	9,777
ORs	1,682,429	288,977
Period	3 years	3 years
Total MFI	204	316
Longest MFI	6	5
ICD-10 codes	5,790	4,697
Chronic diseases	227	228

The list of rare diseases in Bulgaria[5] for which are assigned ICD-10 codes in ORs includes only 112 rare disorders in total. In Table 5 are presented the results for patients with rare diseases in both collections. For each of them is generated interval graph. For collection S2 are counted only patients with rare diseases other than Hyperprolactinaemia.

Table 3 Rare diseases in data collections S1 and S2

Collection	S1	S2
Main diagnosis	Schizophrenia	Hyperprolactinaemia
Patients number	45,945	9,777
Patients with rare diseases	480	1,543
Distinct Rare diseases	23	25

There are 8 rare diseases in S1 with single patient only and in S2 there are 6 such diseases. Among those unique cases in S1 and S2 for 11 rare diseases we have single patient only in both collections.

The rare disease with the highest prevalence in S1 is Hyperprolactinaemia (ICD-10 E22.1). In total 316 patients have both Hyperprolactinaemia and Schizophrenia. The rare disease with the highest prevalence in S2 is “Other hyperfunction of pituitary gland” (ICD-10 E22.8). In total 1,164 patients have both Hyperprolactinaemia and E22.8. The distribution of rare diseases by groups in ICD-10 classification for S1 and S2 is shown in Table 6.

[5] List of Rare Diseases in Bulgaria
http://ncphp.government.bg/files/komisia_rare_diseases/Zapoved_SpisakRB.pdf

Table 4 Classes according to ICD-10 of rare diseases in collections S1 and S2

ICD-10	S1	S2
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	14	2
Disorders of other endocrine glands	444	1,491
Metabolic disorders	5	3
Diseases of myoneural junction and muscle	7	0
Pulmonary heart disease and diseases of pulmonary circulation	1	0
Noninfective enteritis and colitis	4	10
Diseases of the musculoskeletal system and connective tissue	2	30
Congenital malformations, deformations and chromosomal abnormalities	3	7

Another interesting finding is that there are many patients that suffer from more than two rare diseases. On the other hand there are rare diseases with high prevalence associated only with patients in one of the collections but not presented in the other one. Such rare diseases are "Hereditary factor VIII deficiency" (ICD-10 D66) only present in S1. "Systemic lupus erythematosus with organ or system involvement" (ICD-10 M32.1) is only present in S2 and only for women in age 26-47. Almost all of them have also diagnosis "Other nontoxic goiter" (ICD-10 E04).

The context information shows that the majority of patients with rare diseases (over 70%) are women in age 15-44 years old.

6 Conclusion and Further Work

This paper presents work in progress within a research project that aims to elaborate the existing DM methods in order to investigate in depth complex relations in medical events. Although we have big national wide clinical data we cannot use for this task standard techniques for processing, because the patient pool for rare diseases is too small.

Usage of some of the other patient data as a complement to the data for patients with rare disorders shows the capacity of the big data. The proposed method enables to build hypotheses concerning the causality relationships among the factors that trigger the formation of rare diseases.

The future work concerns in-depth analyzes and experiments with various subsets of Outpatient Records.

7 Acknowledgements

The research work presented in this paper is partially supported by the grant *SpecialIZed Data MIning MethoDs Based on Semantic Attributes* (IZIDA), funded by the Bulgarian National Science Fund in 2017–2019. The team acknowledges also the support of Medical University – Sofia, the Bulgarian Ministry of Health and the Bulgarian National Health Insurance Fund.

8 References

1. Mertz, L. Turning the Unknown into Known: Data Mining Is Increasingly Used to Prospect for Rare-Disease Biology and Treatments. IEEE pulse, 8(1), 28-32 (2017).
2. Aymé, S., Bellet, B., & Rath, A. Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. *Orphanet Journal of Rare Diseases*, *10*, 35 (2015).. http://doi.org/10.1186/s13023-015-0251-8
3. Zaki, M. J., and Meira Wagner Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms.* Cambridge University Press, (2014).
4. Boytcheva, S., Angelova, G., Angelov, Z., & Tcharaktchiev, D. Mining comorbidity patterns using retrospective analysis of big collection of outpatient records. Health Information Science and Systems, 5(1), 3 (2017).
5. Jensen, P., L. Jensen and S. Brunak. *Mining electronic health records: towards better research applications and clinical care*. Nature Reviews Vol. 13, 395-405 (June 2012).
6. Suominen, H. *Text mining and information analysis of health documents.* Artificial Intelligence in Medicine 61, 127–130 (2014).
7. Fishburn, P. C. Interval graphs and interval orders. Discrete Mathematics, 55(2), 135-149 (July 1985)
8. Patnaik, D., L. Parida, P. Butler, B. J. Keller, N. Ramakrishnan, and D. A. Hanauer. Experiences with Mining Temporal Event Sequences from Electronic Medical Records: Initial Successes and Some Challenges. In *Proc. 17th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'11)*, San Diego, pp. 360-368 (August 2011).
9. Patnaik, D., S. Laxman, B. Chandramouli, and N. Ramakrishnan. Efficient Episode Mining of Dynamic Event Streams, in *Proc. of the IEEE Int. Conf. on Data Mining (ICDM'12)*, Brussels, Belgium, pp. 605-614 (December 2012)
10. Wang, T., C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), ACM, New York, NY, USA, pp. 457-466 (2008)
11. Gotz, D., Fei Wang, and A. Perer. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *Journal of Biomedical Informatics*, Vol. 48, pp. 148-159 (April 2014).
12. Rind, A., Taowei David Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman. Interactive Information Visualization to Explore and Query Electronic Health Records, *Journal of Foundations and Trends® in Human–Computer Interaction* 5(3), pp. 207-298. (2013).
13. Monroe, M., Rongjian Lan, Hanseung Lee, C. Plaisant, and B. Shneiderman. Temporal Event Sequence Simplification, in *IEEE Transactions on Visualisation and Computer Graphics*, 19(12), pp. 2227-2236. (December 2013)
14. Lee, N., A.F. Laine, Jianying Hu, Fei Wang, Jimeng Sun, and S. Ebadollahi. Mining electronic medical records to explore the linkage between healthcare resource utilization and disease severity in diabetic patients. *Proc. First IEEE Int. Conf. on Healthcare Informatics, Imaging and Systems Biology (HISB)*, pp. 250 – 257. (2011)
15. Yang, J., J. McAuley, J. Leskovec, P. LePendu, and N. Shah. Finding progression stages in time-evolving event sequences. In *Proc. of the 23rd international conference on World wide web* (WWW '14). ACM, New York, NY, USA, pp. 783-794. (2014)
16. Guyet, T., and R. Quiniou. Mining temporal patterns with quantitative intervals. In Zighed D., Z. Ras, and S. Tsumoto (Editors): *Proc. of the 4th Int. Workshop on Mining Complex Data, IEEE ICDM Workshop*, pp. 218-227 (2008).
17. Guyet, T., and R. Quiniou. Extracting temporal patterns from interval-based sequences. In *Proc. 22nd Int. Joint Conference on Artificial Intelligence*, pp. 1306-1311 (2011).
18. Dao-I Lin, Zvi M. Kedem. Pincer Search: A New Algorithm for Discovering the Maximum Frequent Set, in Advances in Database Technology - EDBT'98, 6th International Conference on Extending Database Technology - 1998, pp. 105-119,(1998).

19. Gouda K., M. J. Zaki, GenMax: An Efficient Algorithm for Mining Maximal Frequent Itemsets. *Data Mining and Knowledge Discovery*. 11(3), 223–242 (2005).
20. Grahne, G., & Zhu, J. Efficiently using prefix-trees in mining frequent itemsets. In *FIMI* (Vol. 90). (2003, November).
21. Burdick, D et al MAFIA: A maximal frequent itemset algorithm. *IEEE transactions on knowledge and data engineering*, *17*(11), 1490-1504. (2005).
22. Vo, B., Le, T., Coenen, F., & Hong, T. P. Mining frequent itemsets using the N-list and subsume concepts. *Int. Journal of Machine Learning and Cybernetics*, *7*(2), 253-265 (2016).
23. Pal, A., & Pal, M. Interval tree and its applications. *Advanced Modeling and Optimization*, *11*(3), 211-224 (2009).
24. Boytcheva, S. Shallow Medication Extraction from Hospital Patient Records. In: Koutkias, V., J. Niès, S. Jensen, N. Maglaveras, and R. Beuscart (Eds.), *Studies in Health Technology and Informatics series*, Vol. 166, IOS Press, 119-128, (2011).
25. Boytcheva, S., G. Angelova, Z. Angelov, and D. Tcharaktchiev. Text Mining and Big Data Analytics for Retrospective Analysis of Clinical Texts from Outpatient Care. *Cybernetics and Information Technologies*. Volume 15, Issue 4, 58–77 (November 2015).
26. Yu, H. F., Hsieh, C. J., Chang, K. W., & Lin, C. J. Large linear classification when data cannot fit in memory. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *5*(4), 23 (2012).
27. Heggernes, P., Meister, D., & Villanger, Y. Induced subgraph isomorphism on interval and proper interval graphs. *Algorithms and Computation*, 399-409 (2010).