

ENERGETIC APPROACH OF LOW LATENT TEMPORAL SUMMARIZATION WITH SEQUENTIAL PROCESS MINING

Kishore Kumar M

Research Scholar,

Department of CSE ,

JNTUH College of Engineering, Hyderabad, Telangana.

kishore.mamidala@gmail.com

Dr. Sanampudi Suresh Kumar

Assistant Professor & Head,

Department of Information Technology ,

JNTUH College of Engineering Jagtial, Nachupally (Kondagattu),

Kodimyal mandal, Karimnagar Dist. Telangana.

sureshsanampudi@gmail.com

April 27, 2018

Abstract

In today's internet world, Search Engines are the major sources which are helpful to provide the recent information to users all over the globe without any delay. However the process gets strucked with some challenges such as dynamic data overloading in the web platform. At this time search engines feels struggle and confuse to retrieve which content at which time. So, a specialized data summarization process is required to summarize the temporal information in an efficient way. Temporal Summarization is a methodology, which summarize the data instantly in a correct manner

with proper logics and as per time perspective of whenever the event occurs. Lots of reviews are generated in and increased manner with respect to time and as per the arrival of new events. All the past researches are available only for processing some small amount of data or some period of time, but users requires viewing the instant data with time semantic manner, events should be noticed immediately as per time perspective without the matter of size constraint. So a new methodology is required to implement to summarize the temporal information more efficiently, which is called Low Latent Temporal Summarization along with Sequential Process Mining principles. This sequential process mining is the classic technique, which arrange all the data into the database in a sequential manner and provide to the processing medium in the order of indexing based structure and the processing medium process that data sequentially one-by-one. So the indexed data is easy to proceed for next level of processing, in this stage a Low Latent Temporal Summarization approach is applied to the indexed data and arrange all the important and time-dependent data to the visibility of users in proper and efficient manner.

Key Words: Temporal Summarization, Sequential Process Mining, Low Latency.

1 INTRODUCTION

Information Maintenance and Retrieval is the major task in the mining industry, which is handled by lots of advanced algorithms and techniques. All are expecting the data or information to be retrieved properly based on timely manner without any delay and expect to show the information based on priority wise such as giving importance to information which are most important and which are in normal constraint like that. But the information summarization is quite complex at the time of unexpected new events occurs such as earthquakes, volcano exploration, hurricanes and so on. As the data gave are inadequate, excess and loud. Data get to is critical amid unforeseen news occasions.

It must be given on critical premise particularly to the individuals who are close and worry to the occasion. Temporal/Transient Summarization is an approach that consequently condenses such

information in a sensible way as for the time being of occasion happenings. Audits are shaped incrementally with precise time, when recent information are open. Loads of change and course of action frameworks, the errand for human fashioners is less to find a result to an issue, yet rather to discover an extent of decisions and see the progressions fundamental to them. Certified interest over complex course of action spaces needs development that can enable a human assessment producer to comprehend the key parts of hopeful plans. This framework offers an approach to manage shortening brief methodology that effort on seeing stunning common edges. These systems look for surfaces or surprising transient segments; draw on a saved space theory to goad the request technique. An assessment of the technique on a course of action of mission outlines shows the capacity of the approach.

2 LITERATURE SURVEY

In the year of 2010 the author Andre Bittar, proposed a paper titled "Building a Time Bank for French: A Reference Corpus Annotated According to the ISO- TimeML Standard" in that the author summarized some important details such as the main points in the creation of the French TimeBank (Bittar, 2010), a reference corpus annotated according to the ISO-TimeML standard for temporal annotation. A number of improvements were made to the markup language to deal with linguistic phenomena not yet covered by ISO-TimeML, including cross-language modifications and others specific to French. An automatic preannotation system was used to speed up the annotation process. A preliminary evaluation of the methodology adopted for this project yields positive results in terms of data quality and annotation time. In the year of 2006 the author Roser Sauri et al, proposed a paper titled "TimeML Annotation Guidelines version 1.2.1" in that the author described such as the document is organized as follows. The next section explains what the TimeML tags (XML elements) are and how to annotate them. It also specifies for each tag what its attributes are and provides a BNF definition for the tag and its attributes. While this exposition contains many examples illustrating what and how to tag, the examples focus, for clarity's sake, on the tag under dis-

cussion at any given point. The appendix found the end of this chapter provides a set of fully annotated examples, illustrating all of the interactions between the various entity and relational tags. For the sake of convenience, I&P(02) will be used to refer to Ing02 and TIDES(02), to Fer02 throughout the whole document.

3 TEMPORAL SUMMARIZATION

Temporal Summarization incorporates effective checking and maintenance of data related with an occasion on time. When learning about Temporal Summarization with respect to transience, it is preceding characterize the fundamental ideas identified with time which empowers us to increase valuable bits of knowledge into how Temporal data is passed on in regular composed dialect. In textual representations there are two main constraints called Events and States. Events indicate the occasions and states indicate the position and timeline of the event. Temporal summarization process is handled based on these two constraints without any interruptions.

Temporal Expression: It can be characterized as any component in dialect that lexicalizes the idea of time regarding perceived and by and large quantifiable Temporal units [1]. Concurring to Roser Saur et al., there are four classifications of transient articulations as per the sorts of moments or interims they depict: dates, times, terms and sets [2]. In regular dialect, Temporal Relations are the associations between temporal substances. The connection might be between two occasions, between two Temporal articulations or between an occasion and a worldly articulation. Robert E. Longacre characterizes a Temporal connection as 'a between propositional connection that imparts the synchronization or requesting in time of occasions or states' [3]. Temporal data is passed on in regular dialect through two linguistic components to speak to time and worldly relations.

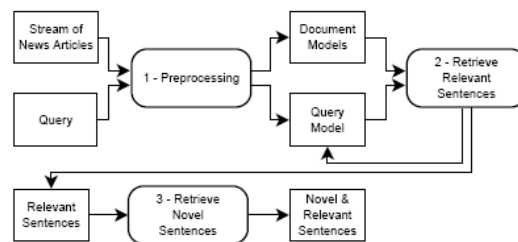


Fig.1. Temporal Summarization Outline

(a) Tense: Tense is a Temporal semantic system that communicates the time at which or amid which an occasion takes place. Hans Reichenbach built up a hypothesis 'The tenses of Verbs', in which tense gives data about the accompanying times: Speech Time (S), Event Time (E) [4].

(b) Aspect/Viewpoint: Aspect/Viewpoint is the second gadget communicating time in common language. Two sorts of angles are communicated in dialect, specifically linguistic perspective and lexical angle. In the writing of Aspect, Zeno Vandler in his article 'Verbs and Times' proposed an underlying refinement amongst occasions and states and after that orders the occasion articulations into three aspectual classes or Aktionsarten: exercises, achievements, and accomplishments [5].

Temporal Summarization contains three important steps such as:

- (i) Data Preprocessing
- (ii) Temporal Data/Information Identification
- (iii) Checking Data/Information Novelty

(i) Data Preprocessing

Data Preprocessing is the important stage of temporal summarization, which orders the data and process the data for further order of processing. In this proposed approach we are using an advanced Sequential Process Mining technique, which indexes the data into procedural manner, which arrange all the data into the database in a sequential manner and provide to the processing

medium in the order of indexing based structure and the processing medium process that data sequentially one-by-one.

Procedure: Sequential Process Mining

Input: Dataset or Database

Output: Indexed Data based on Time Constraint

- (a) Read the whole data or information from the input dataset/database.
- (b) Parse the data/information into group of sentences.
- (c) Identify the time constraints specified into the sentences.
- (d) Construct the indexing tree based data accumulation structure.
- (e) Extracting the important information from the sentences.
- (f) Order the data sentences based on the indexing rule with sequential manner.
- (g) Resulting the sequential ordered indexed data with time oriented strategies.

(ii) Identification of Temporal Data/Information

Temporal Information Processing incorporates two noteworthy tasks. They are:

- (a) One is to naturally perceive and extricate worldly elements in story [to be specific fleeting articulations and occasions], and
- (b) To find temporal relations between these elements and constructing the kind of each remembered one.

The first is named as Temporal Information Identification, which comprises of recognizing and separating temporal substances from common dialect writings. For occasions, data recognizable proof is discovering which literary element constitutes an occasion.

For instance, in the sentence 'He battled with the shrewdness', Filatova and Hovy [6] consider that the whole condition speaks to an occasion. While for Pustejovsky et al. [7], the proper traverse is quite recently the verb gathering or simply the groups head word ['battled' in the case']. Besides, occasion acknowledgment includes the recognizable proof of a few ascribes identified with every occasion which rely upon the theme being considered. Pustejovsky et al., characterized five characteristics in their Time Bank news corpus [8] [tense, aspect, methodology, extremity and class].

A few methodologies were proposed to distinguish occasions in regular dialect substance. R.Faiz proposed a way to deal with recognize significant sentences in news articles for extraction of occasion data [9]. Afterward, ElKhlifi and R.Faiz proposed another auto-

matic approach which is to comment on occasions in News writings [10]. For transient articulations, Temporal Information Identification or substance acknowledgment adds up to recognize its diverse sorts, for example, times, dates, spans, frequencies and to locate its comparing esteems. It should be possible effectively by a commented on corpus including a decent explained construction.

(iii) Checking Novelty of Data/Information

Checking the novelty involves the representations of temporal events and temporal specifications within the process of natural language texts. This portion has highly importance because of its potential appliances like Summarizing, Questioning, answering and so on. Low Latent Temporal Summarization is applied over here in the indexed data/information to summarize the data according to the time and the event when it occurs. The quantitative approach includes the time stamps separated from the web sources, for instance, Publication Time of a posted article.

The subjective approach includes the circumstance where coordinate time stamps can't be separated, however can be derived by examining the substance. For this, James Allen built up a hypothesis, later named as 'Allens Interval Algebra' [12], which gives a reasonable model of time that catches the distinctive pathways in which outcomes might be identified with each other. Allen considers that each worldly articulation or occasion can be introduced as a temporal interim having a begin point and an end point on a timeline.

The Content unit selection and process is the important task in the temporal novelty checking process that is summarized by the following equation:

$$t_i = \{S_j\}; S_j = \{T_p, T_r, T_{trigger}, Scope\}, i, j = 1 \dots n \quad (1)$$

Where i is the point of time, j is the indexed Sentence, s is the scope of the sentence, Tp is the time of publication, Tr is the real-time process scenario as well as the scope is summarization which containing events.

Procedure: Low Latent Temporal Summarization

Input: Sequentially Indexed Data

Output: Detection of Important Temporal Summary (ex. important news based on time constraint)

- (a) Read the indexed data or information from the sequential results.
- (b) Check for novelty based on peaks and climbs with respect to start time of event and end time of event.
- (c) Eliminate the unwanted climbs and steadily changing the time constraints and check for relative importance between key points.
- (d) Compare the present sentence with the intersection scenarios provided by the previous step.
- (e) Resulting the important temporal data with proper low latent classification process.

4 CONCLUSION

The proposed approach clearly describes the process of Temporal Summarization in several aspects. Lots of researchers and analysts showings more interests on Temporal Summarization and its significance is improved day-by-day because of its intersection of linguistics, philosophy and symbolic Artificial Intelligence. In this proposed Low Latent Temporal Summarization all the summarizing concepts are covered, beginning with numerous time constraints and temporal information accessing. Through this proposed approach several Temporal Data/Information concepts are visualized, which contains the process of examining and extraction of various key points such as data preprocessing, significant identification of temporal data summarization and checking the novelty of data which is available into the dataset or database. These natures are clearly analyzed with the help of sequential process mining as well. For all the entire approach is helpful for analyzing and propose the Temporal Information Summarization process more efficiently and its significance is better.

References

- [1] Andr Bittar, (2010) Building a TimeBank for French: A Reference Corpus Annotated According to the ISO- TimeML Standard. PhD thesis, Universit Paris Diderot EFR de Linguis-

tique, cole doctorale de sciences du langage, laboratoire ALPAGE.

- [2] Roser Saur et al., (2006) TimeML Annotation Guidelines version 1.2.1. [3] Robert E. Longacre, (1983) The Grammar of Discourse. New York. Plenum Press
- [3] Hans Reichenbach, (1947) The tenses of verbs. Elements of Symbolic Logic. New York. Macmillan & Co.
- [4] Zeno Vendler, (1967) Verbs and Times, chapter 4, pp 97121. Cornell University Press, Ithaca, New York.
- [5] E. Filatova & E. Hovy, (2001) Assigning time-stamps to event-clauses. In Katz, G., Pustejovsky, J., and Schilder, F., editors, the workshop on Temporal and spatial information processing, Dagstuhl Seminar Proceedings, Morristown, NJ, USA, Association for Computational Linguistics.
- [6] J. Pustejovsky, J. Castano, R. Ingria, R. Sauri, R. Gaizauskas, A. Setzer & G. Katz, (2003a) TimeML: Robust specification of event and temporal expressions in text, In Fifth International Workshop on Computational Semantics. IWCS-5.
- [7] J. Pustejovsky, R. Hanks, A. Sauri, R. See, A. Gaizauskas, A. Setzer, B. Radev, D. Sundheim, L. Day, L. Ferro & M. Lazo, (2003b) The timebank corpus. In Corpus Linguistics. Lancaster.
- [8] R. Faiz, (2006) Identifying relevant sentences in news articles for event information extraction. International Journal of Computer Processing of Oriental Languages (IJCPOL), World Scientific, Vol. 19, No. 1, pp. 1-19.
- [9] A. Elkhilfi & R. Faiz, (2009) Automatic Annotation Approach of Events in News Article, International Journal of Computing and Information Sciences (IJCIS), Vol.7, N1, pp. 19-28.
- [10] Suresh Kumar Sanampudi & G. Vijaya Kumari, (2010) Temporal reasoning in natural language processing: A survey, International Journal of Computer Applications, 1(4), pp5357.

- [11] J. Allen, (1983), Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11), pp, 832843.
- [12] A. Setzer & R. Gaizauskas, (2002.) On the importance of annotating temporal event-event relations in text. In the Third International Conference on Language Resources and Evaluation. LREC2002.
- [13] Ruifang He, Bing Qin, Ting Liu, Sheng Li Information Retrieval Lab, Harbin Institute of Technology , (2010) Cascaded Regression Analysis Based Temporal Multi-document Summarization, *Informatica* 34 (2010) 119124.
- [14] Xiaojun Wan, Institute of Computer Science and Technology, Peking University, Beijing 100871, China, TimedTextRank: Adding the Temporal Dimension to Multi-Document Summarization.