Survey on Seizure Reports from Clandestine Labs

P. Ramina[1], M. Vanitha[2]

[1]Research Scholar, Bharathidasan University, Trichy, Tamilnadu, India

[2]Asst. Professor, Department of Computer Science, J.J. College of Arts & Science, Pudukkotai. Tamilnadu, India

Abstract: *This paper focus on extraction the information and then review the visualization tasks related to the information gathered and then investigate topic modeling approaches. Experimental test bed for event mapping is prepared which uses end to end information retrieval system. Lab reports of seizers patients is collected with respective to time and space.*

Keywords: Information extraction, information visualization, event extraction, data mining, time series, machine learning, spatiotemporal information retrieval

1. Introduction

In this paper, an event is analyzed based on structured data, in the form of time referenced when it took place and the geographical information where it took. And the unstructured events were classified in the free text. This is later detected with the known attributes of the event. Identifying the complicated issues like trends and information about high event activity was lagging and is considered to be the drawback in this research.

Reasons given for the above are given as 1) lack of well defined criteria for activity   2) limitation in existing annotation interfaces.

First the basic approach to event visualization is given. And the general framework makes use of
1) Google Maps
2) The Google tool kit[1].
3) Timeline visualization tools such as MIT SIMILE[2].

It also uses previous work based on gazetteer-based event recognition and syntactic patterns for semantic relationship detection.

Next system is developed for visualization of animal disease for which online news document are used to display reports of methamphetamine lab seizures reports. And then with information retrieved filtering and ranking of the events done.

Topic modeling algorithm is used is applied for computing spatial location, time unit, or combination of both. This allows the data to be more interrogated systematically in order to display geographic regions that are more prone to events.

Choropleth map is used to construct a time composite map within a state or province, or a spatial composite time series by month or year, showing active regions. Event type parameters used to filter events according to the geographic or time line. Finally it has the ability to filter events to display only those that fall within a specified frequency range. For example, the system can be configured to search for seizures of methamphetamine production labs in one country or districts where it is common or rare.

2. Event Visualization Task

2.1 Spatiotemporal Event extraction

It is to identify or extracting based on the specific task, place, relationships or entities. For example a test conducted for contagious animal disease on a farm is an event that may be tied to an epidemic and identified. For example, an event of seizure from methamphetamine production or waste products from dump site.

2.2 Georeferencing and Map visualization

Mapping is done based on the event, took place in certain geographic areas. The result is kept in spatial database management system (SDBMS) using software service like Google maps.For this purpose two alternative access layers with a unified representation and geographic information system (GIS) data model. The first layer is based on Google''s Keyhole Markup Language (KML) and a file-based application programmer interface (API), second layer is based on a PHP interface to a MySQL database implementing the KML schema. Front end as Time-Map used to configure either of the layers.
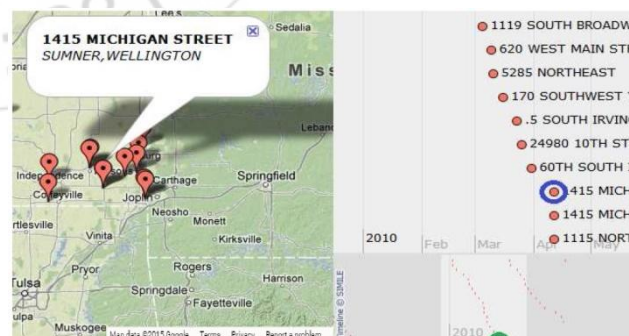

Figure 1: Map and timeline visualization of meth lab seizure events (2004-2011) using Google Maps and MIT SIMILE.

2.3 Timeline Visualization

It is the representation on the map of seizure event took place in year and month and is showed with circled dot and pop-up note. For example the seizure event in April 2010 is represented on the map.

2.4 Thematic Mapping

This mapping is to find the phenomena and trends in each geospatial context. The reason for development of geospatial visualization components is to facilitate information retrieval and multimodal information access using thematic mapping and small multiples. For example Choropleth map is showing the meth lab seizures by state in 2004. [3] This map which is published by the associated Press and featured on the ABC news website in 2005, does not take state population or intrastate population distribution into account.
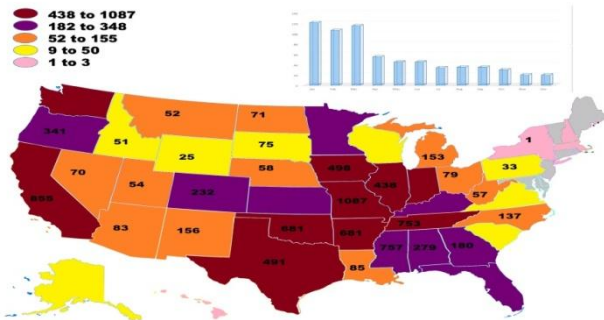

Figure 2: Choropleth map of 2004 meth lab seizures in the USA

2.5 Spatial time series prediction

It is the final time map in this series, which more specific to the domain. Its main objective is epidemiology and criminology. Government agencies meant for agriculture, public health [4] and law uses predictive tools to assist with decision making. Data mining is done with the relevance information of that domain.

3. Topic Modeling

Entity recognition is combined with date and location will help to extract even a micro detail of the event to increase the information transparency. Cluster modeling is done in two ways-: Static Topic model and Dynamic Topic Model.

3.1 Static Topic Models

An Event document is taken from that topic is separated first, and then topic is sampled and generates the word. Markov matrix [8] is used to give word distribution per topic and Dirichlet prior parameter used in generating that matrix. Figure illustrates the kind of generative Bayesian topic model which is used to cluster static collection documents.
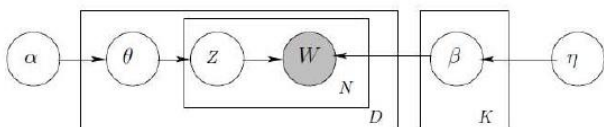

Figure 3: Plate model for Latent Dirichlet Allocation (LDA) in a system with an N-word lexicon, D document, and k topics.

Figure 3 illustrates the kind of generative Bayesian topic model widely used to cluster static collections of documents.

Here Θ is a topic distribution for document Z is the topic sampled from for word W.
β is a Markov matrix,
η is the Dirichlet prior parameter used in generating that matrix.

3.2 Dynamic Topic Models

In the starting of the experiment historical data are used for epizootic disease and math lab seizers. But these data are changing with time as a result arrival of new topics which is tied to observable events.

Blei and Lafferty (2006) proposed a dynamic topic model with fixed topic count K in which each topic‚s word distribution and popularity are linked over time. [5] Older topic modeling algorithms such as latent Semantic Analysis(LSA)[6] that permit K to vary suffer from problems such as proximity of different senses of polysemous word, While variants Probabilistic Latent Semantic Analysis(PLSA) [7] exhibit parameter growth linear in the number of documents D.

3.2.1 Discrete Time, Infinite Topic
They analyzed the birth and evolution of topics in the neural computation community based on the Neural Information Processing Systems. This model evolved topics over discrete time units. This model does not suit less "bursty topic" application such as meth lad seizures or diseases outbreak.

3.2.2 Discrete Time, Infinite Topic
They analyzed the birth and evolution of topics in the neural computation community based on the Neural Information Processing Systems. This model evolved topics over discrete time units. This model does not suit less "bursty topic" application such as meth lad seizures or diseases outbreak.

3.2.3 Continuous time, Finite Topic
This model uses Brownian motion to stimulate the evolution of topics over time and Kalman filtering algorithm for fast inference, the number of topics it samples from is bounded, it severely limits its application in news feed storyline creation and article aggregation. Therefore Limitation of this model is topics covered are in small numbers.

3.2.4 Proposed: Continuous Time, infinite Topic
This model over comes non- bursty news updates and proposed a hybridization of the infinite topic model and continuous time model which combines Dirichlet process(DP) [9] for dynamic topic abstraction and refinement with Brownian motion.
Variational inference algorithm such as variationalkalmanfiltering to factorize the variational distribution [10] over latent variables:
$Q(\beta_{1:T}, z_{1T}, 1:N, \theta_{1:T} | \hat{\beta}, \phi, \gamma =$
$\prod_{k=1}^{K} q(\beta_{1,k}, \ldots, \beta_{T,k} | \hat{\beta}_{1,k}, \ldots, | \hat{\beta}_{T,k}) \times$
$\prod_{t=1}^{T} (q(\theta_t | \gamma_t) \prod_{n=1}^{Nt} q(z_{t,n} | \phi_{t,n}))$
Where β is the word distribution over topics
Is the word distribution over topics for time 1: T,
Word index 1: N, where N is the size of the document.

Paper ID: NOV152762

544

4. Application Test Bed

4.1 Prior Work: Veterinary Epidemiology

Earlier works on Veterinary epidemiology is given which is implemented using Google maps [1]. And more recent version of math Lab application is represented.
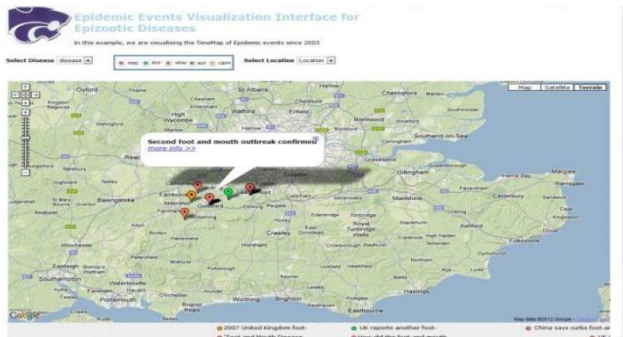

Figure 4: Kansas Information Integration and Analysis Center(KIIAC) for epizootic diseases

4.2 Kansas Meth Lab Seizures

This is the system which forms the test bed for both the visualization techniques and topic modeling techniques [11] intended to raise the precision of the federated system by improving relevance filtering and ranking [12].

5. Experimental Evaluation
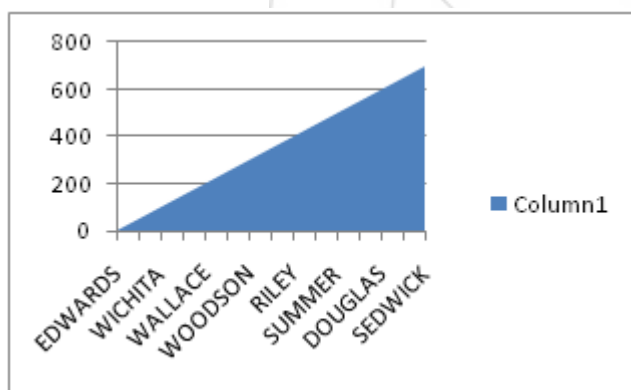
Some statistics for meth lab seizures are showed with column graph of 4942 total meth lab seizures in Kansas countries, for year 2000-2011Abondoned dump site proportion per seizure reports for four Kansas countries is showed in table for over six years.

6. Summary

Survey on visualization of clandestine seizure lab was done, Validated the baseline LDA output using it to filter and rank search results the seizure database represented in the figures and Tables. Feedback from multiple subject experts has been used to evaluate both LDA and DP-based topic models [13].

References

[1] Google (2010) Google maps, Retrieved May 15,2015, from https://maps.google.com.

[2] Massachusetts Institute of Technology. (2008).Retrieved June29,2012, from SIMILE: Semantic Interoperability of Metadata and Information in unlike Environments:

[3] http://simile.mit.edu/wiki/Main_Page

[4] Associated press. (2005, November1). Meth Stats. Retrieved June 29,2012, from ABC news:http://abcn.ws/NYggv5

[5] Brownstein, J., &Feifeld, C. (2007). HealthMap- Global Disease Alert Mapping System. Retrieved January 25,2010, from http://www.healthmap.org

[6] Blei, D.M., & Lafferty, J.D.(2006). Dynamic topic models. Proceedings of ICML 2006, (pp. 113-120). ACM Press.

[7] Landauer, T.K., Foltz, P.W., &Laham, D. (1998). An Introduction to Latent Semantic Analysis. Discourse process, 25,259-284.

[8] Hofmann, T. (2001). Unsupervised Learning by probabilistic Latent semantic Analysis. Machine Learning, 42(1-2), 177-196.

[9] Ahmed, A., & Xing, E.P. (2010). Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. Proceedings of UAI 2010, (pp 20-29). AUAI press.

[10] Wang, C., Blei, D.M., & Heckerman, D. (2008). Continuous time dynamic topic models. Proceedings of UAI 2008, (pp.579-586). AUAI Press.

[11] Blei, D.M (2012, April). Introduction to probabilistic topic models. Communications of the ACM, 55(4), pp. 77-84.

[12] Volkoa, S., &Hsu,W.H(2010). Computational knowledge and information management in veterinary epidemiology.Proceedings of ISI 2010,(pp.120-125).IEEE press.

[13] Volkova, S., Caragea, D., Hsu, W. H., Drouhard, J., &Fowles, L. (2010). Boosting Biomedical Entity Extraction by using Syntactic patterns for Semantic Relation Discovery. Proceeding of WI-IAT 20109pp 272-278) IEEE Press.

[14] Visualization of Clandestine Labs from Seizure Reports: Thematic Mapping and Data Mining Research Directions by William.H and Mohammed Abduljabaar.

Author Profile

Ramina.Pobtained her M.phil degree in Computer Science from Bharathidasan University Thiruchirapalli, in 2007. She is currently a Ph.D.candidate in computer Science Department Bharathidasan Universiy Thiruchirapalli.

Dr. M.Vanitha is the Head, Department of Information Technology, J.J. College of Arts and Science, Pudukkottai. She received her Bachelor's Degree in Mathematics during the year 1995 from SRC, Trichy. She did her Master's degree M.Sc(OR & CA) during the year 1997 at NIT, Trichy, M.Sc in Computer Science at Annamalai University, Chidambaram, M.Phil in Computer Science at Bharathidasan University, Trichy and received Ph.D. in Computer Science, from Mother Teresa University, Kodaikanal.

Paper ID: NOV152762

545