



University
of Glasgow

Fong, A.C.M. (2015) Conceptual Analysis for Timely Social Media-Informed Personalized Recommendations. In: 2015 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, USA, 09-12 Jan 2015, pp. 150-151. ISBN 9781479975426.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/123182/>

Deposited on: 24 August 2016

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Conceptual Analysis for Timely Social Media-Informed Personalized Recommendations

A. C. M. Fong, *Senior Member*, IEEE
University of Glasgow
acmfong@gmail.com

Abstract—Integrating sensor networks and human social networks can provide rich data for many consumer applications. Conceptual analysis offers a way to reason about real-world concepts, which can assist in discovering hidden knowledge from the fused data. Knowledge discovered from such data can be used to provide mobile users with location-based, personalized and timely recommendations. Taking a multi-tier approach that separates concerns of data gathering, representation, aggregation and analysis, this paper presents a conceptual analysis framework that takes unified aggregated data as an input and generates semantically meaningful knowledge as an output. Preliminary experiments suggest that a fusion of sensor network and social media data improves the overall results compared to using either source of data alone.

I. INTRODUCTION

In a world that is increasingly becoming wired with sensors, individual smartphone users also contribute to a joint sensor-social network [1] with nodes made up of human and machine sensors. Fig. 1 captures the essence of this type of socio-sensor network that comprises heterogeneous nodes comprising humans and machines. A characteristic of such a network is that both the volumes and usefulness of data coming in from the various nodes vary as a function of time and, in the case of mobile nodes, location.

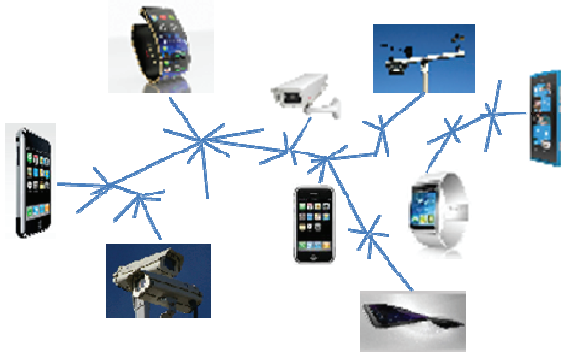


Fig. 1. Ad-hoc socio-sensor network made up of heterogeneous nodes. Volumes and usefulness of data generated by these nodes vary with time and location.

Not only is the amount of data generated by such *ad-hoc* human-machine networks unprecedented, the potential use of such vast socio-sensor data is significant. One possible use of such vast amounts of data is to discover hidden knowledge for consumer applications. One specific example is to extract useful knowledge from the socio-sensor data to provide spatiotemporal information such as trending consumer

products and services at a macro level. This macro-level knowledge complements personal information such as a user's personal attributes (likes and dislikes), physical location and emotional state, etc. to generate location-based, useful and timely personalized recommendations.

This paper builds on the authors' prior work in consumer recommendations [2], and proposes the use of conceptual analysis to discover useful knowledge from socio-sensor data. The conceptual analysis approach is potentially complementary to the kind of ontology-powered knowledge base approach as presented in [3], and this fusion is a subject of further research. It should be emphasized that the proposed approach is adaptable for other purposes, e.g. provide directions based on instantaneous traffic conditions and personal preferences in terms of mode of transport (bus, train, cycling) and choice of routes (highways or not).

The rest of this paper is organized as follows. Section II presents an overview of conceptual analysis within a multi-tier framework inspired by EventShop [4]. Section III demonstrates how the conceptual analysis framework can be applied. It also includes preliminary experimental results that validate the usefulness of aspects of the proposed framework. Finally, Section IV concludes the paper.

II. OVERVIEW

The fundamental premise of the conceptual analysis framework is to blend together heterogeneous data gathered from different sources that may be scattered around different physical locations. The essence of this framework is captured in Fig. 2, which shows the blending and pre-processing of multi-source, heterogeneous data, followed by analysis.

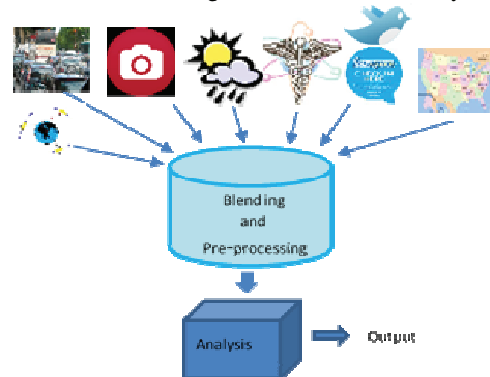


Fig. 2. Blending and preprocessing of multiple data sources, followed by analysis.

III. CONCEPTUAL ANALYSIS FRAMEWORK

A. Formal Concepts

There are two flavors of formal concepts. The first is based on the traditional formal concept analysis FCA [5]. The second is a fuzzified version of FCA that attempts to better model real world phenomena, where property–value associations are not always clear cut. By way of example, a (property, value) pair can be something like (color, blue) or (texture, smooth). Traditionally, uncertainty and vagueness information that describe real-world concepts has been handled using probabilistic models, certainty factors, fuzzy sets, belief functions, etc. [6], all have their respective strengths and weaknesses.

Formal concepts stretch beyond formal context by examining the membership values, and making sure that those membership values associated with the object – attribute pairs for each formal context meet a certain predefined threshold to be included. A fuzzy formal concept, then, is a pair that contains both an extent and an intent.

B. Data Structure and Format

Although a hierarchical structure is often favored by computer scientists, it is not the most appropriate for conceptual analysis. It imposes unnecessary order among the concepts that can be counter-productive in some cases to model inter-relationships between real world concepts. A hierarchical structure would be useful for deriving taxonomy information necessary for ontology generation [7], but in the present application a lattice structure would be more suitable.

In terms of data format, the key idea is to find a balance between fidelity and efficiency not unlike a rate-distortion tradeoff. In particular, the amount of information content generally increases with an increasing payload (i.e. using more bits). However, the downside is more latency, more computing power and storage requirements. Following on from [4], a reasonable tradeoff is to attach these attributes to each data item: time, space, and semantic meaning. These attributes correspond to timestamp, geographical location (coordinates) and a descriptive tag drawn from a finite alphabet. Data coming in from various heterogeneous sources are brought together using this unified format.

C. Application Example

In this illustrative example, mobile subscribers receive timely and personalized messages about an impending situation so that they can take appropriate actions to stay out of harm's way.

Scenario: law enforcement agencies are taking concerted effort in crowd control at an unexpected event. There is a sudden and very significant increase in social media traffic (tweets, etc.) mentioning civil unrest in the Downtown area. Security camera footages at the scene seem to corroborate.

Data selection: geographical data, social media data, security footage, news reports, satellite imagery (e.g. fires).

Preprocessing, blending, unification and aggregation: data streams are coming in at high and varying rates. Light weight preprocessing is performed to clean noisy data and discard data based on predetermined rules to meet latency constraints. Light weight video processing is applied to security footage to capture representative frames and extract semantic meaning as tags. Data streams from various sources are suitably wrapped and blended using the unifying format.

Analysis: aggregated data now arrive at the conceptual analyzer. As described in Section III.A, data are put through the FCA or fuzzified FCA analysis and the discovered concepts are clustered and stored in a lattice structure for subsequent knowledge retrieval and reasoning.

D. Preliminary Results

Simulation was performed to gauge the influence of social media influence using F1-measure. Without social media data (-SM), the algorithm uses only data from official government reports and newsfeeds from well-known agencies. The experiment was repeated three times over a period of one month. The -SM results were consistently about half of the +SM results. These findings suggest that social media data have a significant influence on the detection results. Likewise, presence of sensor network data (\pm SN) also has a significant influence on the results. Both SM and SN are important.

IV. CONCLUSION

Human social networks and sensor networks are becoming increasingly intertwined, which can lead to an explosive growth of volumes and types of heterogeneous data. Trying to make sense of such data and actually extracting useful knowledge can be both challenging and rewarding. This paper has presented an adaptable framework for performing formal concept analysis informed by social media data, which are typically characterized by containing useful spatiotemporal information.

REFERENCES

- [1] J. G. Breslin, S. Decker, M. Hauswirth, G. Hynes, D. Le Phuoc, A. Passant, A. Polleres, C. Rabsch, V. Reynolds, "Integrating social networks and sensor networks", W3C Workshop on the Future of Social Networking, 15-16 January 2009, Barcelona, Spain.
- [2] A.C.M. Fong, B. Zhou, S.C. Hui, G.Y. Hong and T.A. Do, "Web content recommender system based on consumer behavior modeling", *IEEE Trans. Consumer Electronics*, Vol. 57/2, pp. 962-969, 2011.
- [3] A.C.M. Fong, B. Zhou, S.C. Hui, G.Y. Hong and J. Tang, "Generation of personalized ontology based on consumer emotion and behavior analysis", *IEEE Trans. Affective Computing*, Vol. 3/2, pp. 152-164, April-June, 2012.
- [4] M. Gao, V.K. Singh, and R. Jain, "Eventshop: from heterogeneous web streams to personalized situation detection and control", *Proc. 4th Annual ACM Web Science Conference WebSci '12*, pp. 105-108, 2012.
- [5] B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, 1999.
- [6] R. Kruse, E. Schwecke, J. Heinsohn. *Uncertainty and vagueness in knowledge-based systems: numerical methods*. Springer, 2011.
- [7] Q.T. Tho, S.C. Hui, A.C.M. Fong and T.H. Cao. "Automatic fuzzy ontology generation for semantic web", *IEEE Trans. Knowledge and Data Engineering*, Vol. 18/6, pp. 842-856, June 2006.