

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

Similarity Based and Transfer Learning Based Data Fusion Using Classifier

Yogesh Shinde¹, Nikhil Ghule¹, Rubina Shaikh¹, Ashish Ramdasi²

B. E Students, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India¹

Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India²

ABSTRACT- We combination the data fusion methodologies, grouping them into three categories: the stage-based, the feature level-based, and the semantic meaning-based data fusion methods. The last category of data fusion methods is further divided into four groups: multi-view learning-based, similarity-based, probabilistic dependency-based, and transfer learning-based methods. These techniques concentrate on knowledge fusion rather than schema mapping and data merge, significantly distinguishing between cross-domain data fusion and traditional data fusion studied in the database group. We does not only abnormal high-level principles of each category of techniques, but also give examples in which these techniques are used to handle genuine big data problems. In addition, this paper positions existing works in a system, investigating the relationship and distinction between different data fusion methods. This paper will enable a wide range of communities answer a solution for a data fusion problem in big data projects.

KEYWORDS: Big Data, cross-domain data mining, data fusion, multi-modality data representation, deep neural networks, multi-view learning, matrix factorization, probabilistic graphical models, transfer learning, urban computing.

I. INTRODUCTION

Over couple of years back data fusion/information fusion was seen as sensor data fusion method, where different sensor would produce data, information retrieved from those sensor would be used to sense the other environment situation, so to take decision accordingly.[5] Fusion consists in touching or merging information those branches from several sources and exploiting, merged information in various tasks such as answering questions, making decisions, numerical estimations for further predictive analysis, for gaining insights of the data pattern and used for retrieving useful inferences. Information fusion is process dealing with association, correlation, and combination of data and information from multiple sources to accomplish refined estimates of parameters, characteristics, events, and behaviors for observed objects in an observed field of view.[3] It is sometimes implemented for automated decisions support system. Integrated information systems provide users with a unified view of various heterogeneous sources. Querying the underlying data sources, combing the results, and presenting them to user is performed by the integration system. [2] With more and more information sources easily available via cheap network connection, either over the internet or in company intranet, the desired to access all these source through a consistent interface has been the driving force behind much research in the field of information integration. During the last three decades numerous systems that try to accomplish this goal have been developed, with varying degrees of success.[4]

1. Motivation

Data Mining generally works on the single domain. Recently embrace data is a new technology to works on large amount of information. [5]There are different methodologies which are stage based, feature level based and Semantic meaning based data fusion method. In semantic meaning based there are four type they are multi view learning based, similarity based probabilistic dependency based and transfer learning based method. [6]

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

2. Objectives

This paper condenses existing data fusion methods, characterizing them into three major categories and giving genuine examples in each sub-category of methods. This paper investigates the relationship and differences between different methods, helping people find proper data fusion methods to solve hug data set problem .Some open multi-modality datasets have been shared to facilitate further research into data fusion problems with different Classifications as K-NN and Clustering Classification are utilized.

3. Problem Definitions

Cross-domain recommendation gives best result and more varied recommendations leading to the satisfaction of user and solving cold start and sparsely problems but it is less accurate than single domain recommendation system. Also another issue is the lack of contextual features in cross domain recommender system, which when used can be very efficient in recommendation. Thus this paper talks about different methodologies such as the recommendation methods are POI and Collaborative filtering.

II. RELATED WORK

Sr No	Title of Paper	Author	Description
1	“Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data,”	J.BaoY.Zheng, and M.F.Mokbel	Geo Social networking Data
2	“Using collaborative filtering to weave an information tapestry,”	D. Goldberg, N. David, M. O. Brain, T. Douglas,	Collaborative filtering to weave an information is used for Fusion Process
3	“Collabrative Filtering based on user attributes and user rating for restaurant recommendation”	Ling Li, Ya Zhou, Han Xiong, Cailin Hu, Xiafei Wei	Users preference about restaurant affected by many factors

III. EXISTING SYSTEM APPROACH

This document summarizes methods that can merge multiple data sets into three categories. The first category of data merging methods utilizes several data groups in different phases of a data mining activity. The second category of methods learns a new representation of the original features extracted from different data sets using deep neural networks (DNNs). The new feature representation will then be introduced into a classification or prediction model.

The third category combines data based on their semantic meanings, which can be categorized into two groups:

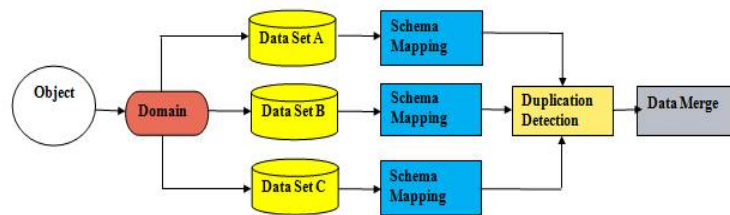
1. Similarity based methods
2. Transfer leaning based Methods

International Journal of Innovative Research in Science, Engineering and Technology

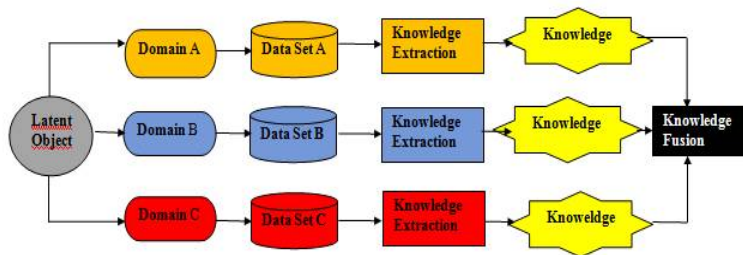
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018



A) Paradigm of the conventional data fusion



B) Paradigm of data fusion for (cross domain) big data

Fig 1. Existing System Architecture

IV. PROPOSED SYSTEM ARCHITECTURE

This Document summarizes the methods that can fuse multiple datasets by 2 categories. The first type of data fusion methods uses similarity data sets. Second type uses transfer of data set from source to domain.

These categories are sub category of semantic meanings, which can be further arranged as:

1. Similarity based methods
2. Transfer leaning based Methods

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

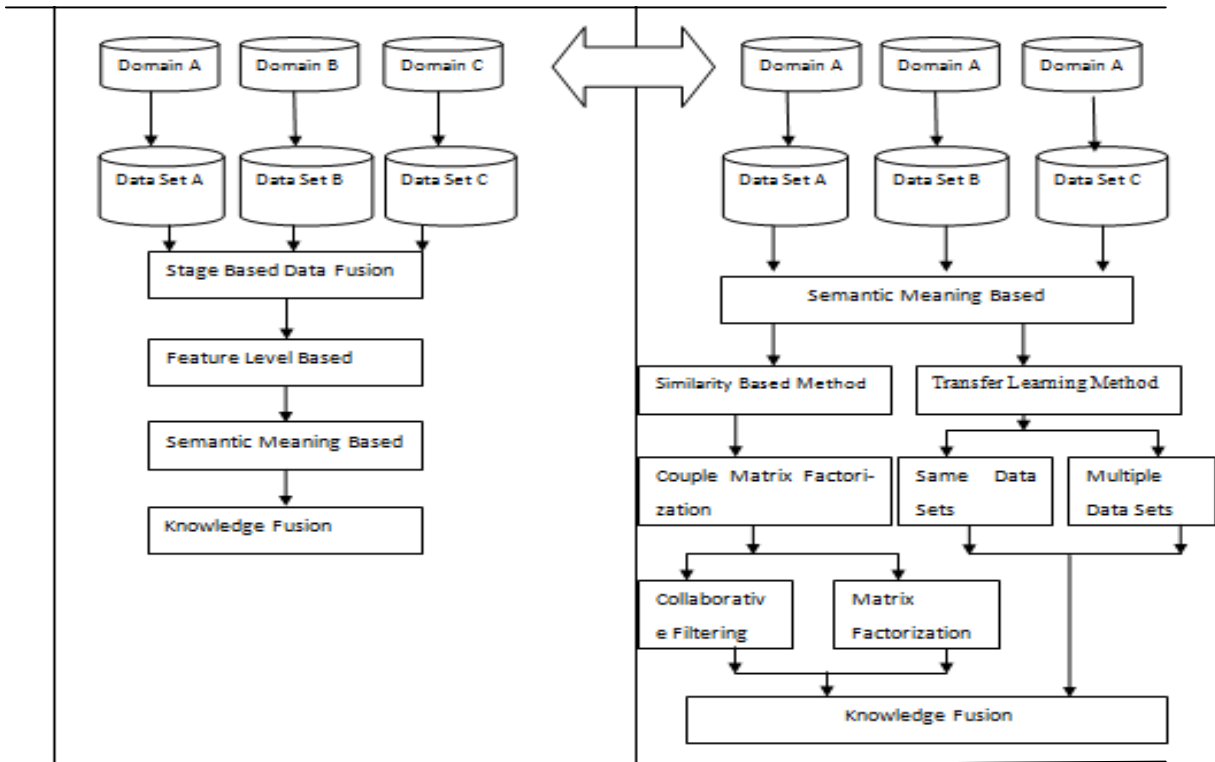


Fig 2. Proposed System Architecture

V. CONCLUSION

The methods that can find knowledge from multiple disparate datasets with underlying connections. This paper compresses existing data fusion methods, arranging them into three major categories and giving real examples in each sub-classification of strategies. This paper explores the relationship and differences between different methods, helping individuals find proper data fusion methods to care big data problems. Some public multi-modality datasets have been shared to encourage further research into data fusion problems.

REFERENCES

- 1) Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data," J. Bao, Y. Zheng, and M. F. Mokbel.,
- 2) A model of inductive bias learning," J. Baxter.
- 3) Representation learning: A re-view and new perspectives, Y. Bengio, A. Courville, and P. Vincent.
- 4) Christopher M. Bishop, "Chapter 8. Graphical Models," A. Blum and T. Mitchell.
- 5) Co-em support vector learning," U. Brefeld and T. Scheffer,
- 6) Predictive subspace learning for multi-view data: A large margin approach," N. Chen, J. Zhu, and E.P. Xing,
- 7) Context-Dependent Pre Trained Deep Neural Networks for Large Vocabulary Speech Recognition, E. Dahl, D. Yu, L. Deng, and A. Acero,
- 8) Using collaborative filtering to weave an information tapestry D. Goldberg, N. David, M. O. Brain, T. Douglas.
- 9) "Multiple kernel learning algorithms" M. Gonen and E. Alpaydn,
- 10) "Investigating causal relations by econometric models and cross-spectral methods" C. W. Granger