

A Comprehensive Study on Social Network Mental Disorders Detection Via Online Social Media Mining

C. Rukumanikhandhan

Assistant Professor
Dept. of Computer Science and
Engineering
K.S.Rangasamy College of
Technology
Tiruchengode, India

B. Megala

Dept. of Computer Science
and Engineering
K.S.Rangasamy College of
Technology
Tiruchengode, India

K. Ragul

Dept. of Computer Science
and Engineering
K.S.Rangasamy College of
Technology
Tiruchengode, India

R. Sabitha

Dept. of Computer
Science and Engineering
K. S. Rangasamy College
of Technology
Tiruchengode, India

Abstract:- The explosive growth in popularity of social networking leads to the problematic usage. An increasing number of social network mental disorders, such as Cyber-Relationship Addiction, Information Overload, and Net Compulsion, have been recently noted. Symptoms of these mental disorders are usually observed passively today, resulting in delayed clinical intervention. In this Project, we argue that mining online social behavior provides an opportunity to actively identify Social Network Mental Disorders at an early stage. It is challenging to detect Social Network Mental Disorders because the mental status cannot be directly observed from online social activity logs. Our approach, new and innovative to the practice of Social Network Mental Disorders detection, does not rely on self-revealing of those mental factors via questionnaires in Psychology. Instead, we propose a machine learning framework, namely, Social Network Mental Disorder Detection, that exploits features extracted from social network data to accurately identify potential cases of Social Network Mental Disorders. We also exploit multi-source learning in Social Network Mental Disorders Detection and propose a new Social Network Mental Disorders -based Tensor Model to improve the accuracy. To increase the scalability of Social Network Mental Disorders based Tensor Model, we further improve the efficiency with performance guarantee. Our framework is evaluated via a user study with 3126 online social network users. We conduct a feature analysis, and also apply Social Network Mental Disorders Detection on large-scale datasets and analyze the characteristics of the three Social Network Mental Disorder types. The results manifest that Social Network Mental Disorders Detection is promising for identifying online social network users with potential Social Network Mental Disorders.

1 INTRODUCTION

TENSOR FACTORIZATION ACCELERATION

Low-rank sparse tensor factorization is a popular tool for analyzing multi-way data and is used in domains such as recommender systems, precision healthcare, and cyber security. Imposing constraints on a factorization, such as non-negativity or sparsity, is a natural way of encoding prior knowledge of the multi-way data. While constrained factorizations are useful

for practitioners, they can greatly increase factorization time due to slower convergence and computational overheads. Recently, a hybrid of alternating optimization and alternating direction method of multipliers (AO-ADMM) was shown to have both a high convergence rate and the ability naturally incorporate a variety of popular constraints.

In this work, we present a parallelization strategy and two approaches for accelerating AO-ADMM. By redefining the convergence criteria of the inner ADMM iterations, we are able to split the data in a way that not only accelerates the per-iteration convergence, but also speeds up the execution of the ADMM iterations due to efficient use of cache resources. Secondly, we develop a method of exploiting dynamic sparsity in the factors to speed up tensor-matrix kernels. These combined advancements achieve up to 8x speedup over the state-of-the-art on a variety of real-world sparse tensors.

Tensors are the generalization of matrices to higher orders. Tensor factorization is a powerful tool for approximating and analyzing multi-way data, and is popular in many domains

across machine learning and signal processing, including recommender systems, precision healthcare, and cybersecurity. These domains produce sparse tensors with millions to billions of non-zeros. Oftentimes, a domain expert wishes to encode some prior knowledge of the data in order to obtain a more interpretable factorization. Prior knowledge is typically incorporated by either forcing the solution to take some form (i.e., imposing a constraint), or penalizing unwanted solutions (i.e., adding a regularization). For example, imposing a nonnegativity constraint on a factorization allows one to better model data whose values are additive. Similarly, adding a regularization term which encourages sparsity can help model data whose interactions are sparse. While valuable to practitioners, constrained and regularized factorizations change the underlying computations and can significantly increase the computational cost of factorization. There is a growing body of research dedicated to efficient optimization algorithms for

constrained and regularized tensor factorization, especially non negative factorization. Huang introduced AO-ADMM, a hybridization of alternating optimization (AO) with the alternating direction method of multipliers (ADMM). The combination of the two frameworks allows AO-ADMM to have both a fast convergence rate and the flexibility to incorporate new constraints and regularizations with minimal effort.

However, alongside the growing body of research is an increasing disparity between efficient optimization algorithms and the available implementations for large-scale tensors. Likewise, there are few available tools which flexibly support a variety of constraints, and to the best of our knowledge none of them are parallel or handle large-scale data. Domain experts must currently go through a major implementation effort to explore the application of a new constraint or regularization, and likely will not easily be able to analyze the full amount of available data due to computational complexity.

To that end, we present a parallelization strategy and high performance implementation of the AO-ADMM framework for shared-memory systems. Our algorithm features two optimizations: (i) a blockwise reformulation of ADMM to improve convergence rate, parallelism, and cache efficiency; and (ii) a method of exploiting the sparsity which dynamically evolves in the factorization. The blockwise reformulation is applicable to any constraint or regularization which is row separable (e.g., non-negativity or row simple constraints), and factor sparsity naturally occurs in many constraints and regularizations including non-negativity. In summary, our contributions include: 1) A block wise reformulation of the AO-ADMM algorithm which improves convergence and execution rate while eliminating parallel synchronization overheads. 2) A method of leveraging sparsity in the factors as they dynamically evolve. 3) An open source, high performance implementation of AO-ADMM which flexibly handles new constraint and regularizations. The rest of this paper is organized as follows. Section I introduces notation and details the AO-ADMM algorithm. Section III reviews existing work on matrix and tensor factorization.

FEATURE EXTRACTION

Feature extraction addresses the problem of finding the most compact and informative set of features, to improve the efficiency of data storage and processing. Defining feature vectors remains the most common and convenient means of data representation for classification and regression problems. Data can then be stored in simple tables (lines representing “entries”, “data points”, “samples”, or “patterns”, and columns representing “features”). Each feature results from a quantitative or qualitative measurement, it is an “attribute” or a “variable”. Modern feature extraction methodology is driven by the size of the data tables, which is ever increasing as data storage becomes more

and more efficient. After many years of parallel efforts, researchers in Soft-Computing, Statistics, Machine Learning, and Knowledge Discovery, who are interested in predictive modeling are uniting their effort to advance the problem of feature extraction.

The recent advances made in both sensor technologies and machine learning techniques make it possible to design recognition systems, which are capable of performing tasks that could not be performed in the past. Feature extraction lies at the center of these advances with applications in the pharmaco-medical industry, oil industry, industrial inspection and diagnosis systems, speech recognition, biotechnology, Internet, targeted marketing and many of other emerging applications. Dozens of research groups competed on five large feature selection problems from various application domains: medical diagnosis, text processing, drug discovery, and handwriting recognition.

2 PROPOSED SYSTEM

In this project, we aim to explore data mining techniques to detect three types of SNMDs [1]: 1) Cyber-Relationship (CR) Addiction, which includes the addiction to social networking, checking and messaging to the point where social relationships to virtual and online friends become more important than real-life ones with friends and families; 2) Net Compulsion (NC), which includes compulsive online social gaming or gambling, often resulting in financial and job-related problems; and 3) Information Overload (IO), which includes addictive surfing of user status and news feeds, leading to lower work productivity and fewer social interactions with families and friends offline. Accordingly, we formulate the detection of SNMD cases as a classification problem. We detect each type of SNMDs with a binary SVM. In this study, we propose a two phase framework, called Social Network Mental Disorder Detection (SNMDD), as shown in Figure 1. The first phase extracts various discriminative features of users, while the second phase presents a new SNMD-based tensor model to derive latent factors for training and use of classifiers built upon Transductive SVM (TSVM). Two key challenges exist in design of SNMDD: i) we are not able to directly extract mental factors like what have been done via questionnaires in Psychology and thus need new features for learning the classification models; ii) we aim to exploit user data logs from multiple OSNs and thus need new techniques for integrating multi-source data based on SNMD characteristics. We address these two challenges in Sections 3.1 and 4, respectively.

ADVANTAGES:

- ✓ Better performance.
- ✓ Improved social media data analysis classification.
- ✓ High in accuracy.
- ✓ SNMD approach that provide better accuracy compare with previous work.

3 MODULE DESCRIPTION

3.1 PREPROCESSING MISSING VALUE IMPUTATION:

In this module the SNMD social media datasets contain missing values, often encoded as blanks, NaNs or other placeholders. Such datasets however are incompatible with process estimators which assume that all values in an array are numerical, and that all have and hold meaning. A basic strategy to use incomplete datasets is to discard entire rows and/or columns containing missing values. However, this comes at the price of losing data which may be valuable (even though incomplete).

A better strategy is to impute the missing values, i.e., to infer them from the known part of the data. See the Glossary of Common Terms and API Elements entry on imputation.

The SimpleImputer class provides basic strategies for imputing missing values. Missing values can be imputed with a provided constant value, or using the statistics (mean, median or most frequent) of each column in which the missing values are located. This class also allows for different missing values encodings.

3.2 FEATURE EXTRACTION:

In this module that focus on extracting discriminative and informative features for design of SNMDD. This task is nontrivial for the following three reasons. 1. Lack of mental features. Psychological studies have shown that many mental factors are related to SNMDs, e.g., low self-esteem, loneliness. To detect SNMDs, an intuitive idea is to simply extract the usage (time) of a user as a feature for training SNMDD. However, this feature is not sufficient because i) the status of a user may be shown as "online" if she does not log out or close the social network applications on mobile phones, and ii) heavy users and addictive users all stay online for a long period, but heavy users do not show symptoms of anxiety or depression when they are not using social apps. How to distinguish them by extracting discriminative features is critical. 3. Multi-source learning with the SNMD characteristics.

3.3 MULTI-SOURCE LEARNING WITH TENSOR DECOMPOSITION ACCELERATION

Many users are inclined to use different OSNs, and it is expected that data logs of these OSNs could provide enriched and complementary information about the user behavior. Thus, we aim to explore multiple data sources (i.e., OSNs) in SNMDD, in order to derive a more complete portrait of users' behavior and effectively deal with the data sparsity problem.

To exploit multi source learning in SNMDD, one simple way is to directly concatenate the features of each person derived from different OSNs as a huge vector. However, the above approach tends to miss the correlation of a feature in different OSNs and introduce interference. Thus, we explore tensor techniques which

have been used increasingly to model multiple data sources because a tensor can naturally represent multi-source data.

We aim to employ tensor decomposition to extract common latent factors from different sources and objects. Based on tensor decomposition on T , we present a SNMD-based Tensor Model (STM) in previous work, which enables U to incorporate important characteristics of SNMDs, such as the correlation of the same SNMD sharing among close friends.⁸ Finally, equipped with the new tensor model, we conduct semi supervised learning to classify each user by exploiting Transductive Support Vector Machines (TSVM) in Appendix B. In the following, the problem definition, notation explanation, and brief introduction are first presented for better reading.

4 CONCLUSION

In this Project, we make an attempt to automatically identify potential online users with SNMDs. We propose an SNMDD framework that explores various features from data logs of OSNs and a new tensor technique for deriving latent features from multiple OSNs for SNMD detection.

This work represents a collaborative effort between computer scientists and mental healthcare researchers to address emerging issues in SNMDs. As for the next step, we plan to study the features extracted from multimedia contents by techniques on NLP and computer vision.

We also plan to further explore new issues from the perspective of a social network service provider, e.g., Facebook or Instagram, to improve the well-beings of OSN users without compromising the user engagement.

ACKNOWLEDGEMENTS

We Acknowledge DST- File No.368. DST – FIST (SR/FIST/College-235/2014 dated 21-11-2014) for financial support and DBT – STAR College – Scheme - ref.no: BT/HRD/11/09/2018 for providing infrastructure support.

REFERENCES

- [1] Archana S, Dr. K. Elangovan, "Survey of Classification Techniques in Data Mining", International Journal of Computer Science and Mobile Applications vol2, Issue 2,, February 2015, pp. 65-71
- [2] Harwatia, Ardita Permata Alfiania, Febriana Ayu Wulandaria, "Mapping Student's Performance Based on Data Mining Approach," Science Direct, Agriculture and Agricultural Science Procedia 3 (2014) pp:173 – 177.
- [3] Harwatia, Ardita Permata Alfiania, Febriana Ayu Wulandaria, "Mapping Student's Performance Based on Data Mining Approach," Science Direct, Agriculture and Agricultural Science Procedia 3 (2015), ICoA.
- [4] Hijazi, Naqvi, "Factors Affecting Students

- [5] Performance-A Case of Private Colleges”, Bangladesh e-Journal of Sociology. Vol. 3, Number 1, Jan 2006.
- [5] Kavipriya .P,”A Review on Predicting Students’ Academic Performance Earlier, Using Data Mining Techniques,” Int. Jour. Of Adv Res. In Comp. sci. and Software Engg., Volume 6, Issue 12, December 2016 ,ISSN . 2277 128X ,pp. 101-105.
- [6] Kalyani M Raval,”Data Mining Techniques for Students’ Performance Prediction,” Int. Jour. Of Adv. Res. In comp sci. and software Engg.,Vol. 2,Issue. 10,Oct 2012, pp. 439-442.
- [7] Rajni Jindal , Malaya Dutta Borah India, “A Survey On Educational Datamining And Research Trends”, International Journal of Database Management Systems (IJDM) Vol.5, No.3, June 2013.