International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

An Efficient Model In Online Education Data Analysis Using Accelerated Algorithm

R Ragunathan, S G Balakrishnan

PG Student, Department of Coputer Science and Engineering, Mahendra Engineering College, Namakkal
Associate Professor, Department of Coputer Science and Engineering, Mahendra Engineering College, Namakkal

ABSTACT: The main idea of our approach is to analyze sentiment analysis is an issue that can be solved by using learning management systems to enhance learning via social media. Existing sentiment categorization has long been recognised as a domain-specific issue. As a result, it produces a very low estimation depression rate of a person and accuracy is low. The problem can be rectified by using the classifier that is collaborative multi domain sentiment classifier, advantage of this is to find the depression state of a person more accurately. To do this, we will utilise multi-task learning to collaboratively train sentiment classifiers for different domains in a collaborative way. When it comes to brand messaging, political campaigns, marketing research, and customer feedback, it is often utilised. The analysis of the subject The use of a Bag of word is important technique (BODW). The essential words in a text reflect either a fact or a feeling about the subject. The objective labels are represented by fact, whereas the subjective labels are represented by emotion. Based on objective and subjective variable selection, the system extracts a bag of discriminative words from a text. The discriminative words in a text are filtered using LDA and regression methods. The SA is also done by one of those methods named support vector machine, and the classification algorithm of Nave Bayes without the need for language resources, it is possible to categorise emotion words into positive and negative categories while simultaneously recording changes. This result will tell us who is in depression by calculating the score. This can be done with the help of government approval to access the data from socialmedia.

Index Terms : Sentiment Analysis , Point wise mutual information , Latent Dirichlet Allocation

1. INTRODUCTION

Mining sentiment data from large amounts of user-generated material may aid in detecting public sentiment on a variety number of applications and may be used to a wide range of topics, including products, businesses, disasters, events, personalities, and so on. Taking the example of tweets, researchers have found that analysing emotions in tweets has the ability to forecast variations in stock market values as well as presidential election results. In addition, classifying the emotions expressed in a large number of microblog posts may be used to replace [1] Sentiment analysis of product reviews may help businesses improve their goods and services while also assisting consumers in making better informed choices. Client interest removal, customised suggestion, social publicity, buyer relationship management, and crisis management have all been shown to benefit from analysing consumer produced pleased emotions. As a consequence, sentiment categorization has become a popular study subject in both the academic and industrial worlds.

Sentiment organisation is considered as a passage classification issue in certain majority sentiment research techniques. It has been possible to utilise these methods to investigate the emotions expressed in product reviews, microblogs, and other locations. However, classification of emotions is widely recognised to be a domain-dependent problem in most cases. [3]This is due to the fact that different response phrases are accessible in different domains, and that the same phrase may indicate different emotions in different domains when used together. The word "easy" is, for example, usually considered favourable in electronic product assessments, as in "this digital camera is easy to use." In the context of film reviews, on the other hand, the word "easy" is often used as a descriptive term. Take, for example, the statement, "This movie's ending is easy to predict." [4] An unstructured approach to this issue would be to utilise the labelled samples from this field to drive a domain detailed sentiment classifier for both domains, which would be an unstructured solution to this problem.Nonetheless, labelled data is often sparse in many areas. Because there are so many domains inhabited by online consumer produced content, explaining adequate examples for them is extremely expensive and time consuming. It is challenging to construct a strong and effective domain-specific sentiment classifier when the training data is lacking.Because a domain has its own unique collection of sentiment expressions, yet a number of sentiment terms are common across domains, our study is required, which provides a starting point for our investigation. In this method, each domain has a sentiment classifier that has two distinct parts: a domain-specific classifier and a global classifier. Using labelled samples from a single domain, is capable of identifying domain-specific attitude expressions. [5] It has the capability of collecting broad sentiment information over a wide range of topics. It's also important to acquire broad sentiment information that's been gathered before for sentiment lexicons that have a general purpose and include that data into our overall strategy so the model can learn the general sentiment classifier. In addition, domain-specific sentiment information is best collected from limited labelled data, small unlabeled samples, and large unlabeled samples. To increase the effectiveness of a domain-specific sentiment classifier, it is feasible to use domain-specific sentiment information and make use of domain-specific sentiment features, as discussed in [6]. To measure the similarities across different domains, both textual content and emotion word distribution similarity metrics are being investigated, with one focusing on text and the other on emotion word distribution..

Related Work

Trends in Distance Education: Using New Technologies to Encourage Student Collaboration and Interaction

Disruption is occurring in the area of distant education. Advances in new technologies point to an emphasis on student engagement in learner-centered constructivist learning settings. This article will investigate the advantages of using modern technological tools like wikis, blogs, and podcasts to help enhance student involvement in online courses. Write Board™, InstaColl™, and Imeem™ are also used for social media applications.Distance education programmes across the globe are confronted with challenges that may limit or prohibit their use[8], despite all the new contemporary technology that make it easier to promote cooperation in both synchronous and asynchronous learning settings, there are still situations when students choose to study in the classroom alone. In this article, researchers are looking at the possible effect of technology on theory and the consequences that this impact may have.

A Learning Analytics approach is used in student-centered learning processes where Business Intelligence is used

Learning Analitycs (LA) are being used in Higher Education to evaluate student performance and attributes such as profile, interactions in a virtual learning environment learning - for this, the business intelligence paradigm is being used to explore and exploit data from one of the actors in the educational process is being examined, test scores, and others, which will contribute to their educationa This article is intended to discuss the specific objectives listed below in more detail: To identify the factors that influence a college student's decision to drop out of distant learning and to develop a profile of people who may be at risk of becoming susceptible as a result of their university studies[4]. Our strategy is to create two learning analysis tasks and execute them using a business intelligence methodology in order to accomplish this objective.

An Investigation into the Use of a Data Mining Method in LMS for the Improvement of Network Engineering Programs

In spite of the significant disadvantages, education is nevertheless feasible despite the merger. Students must recognise and deal with the issues that have emerged, but it is essential to use adaptable models of education that integrate new and improved technology, which enable students to keep learning in even the most difficult situations. To achieve this goal, some previously overlooked ideas and techniques must be reviewed.

An educational strategy based on knowledge discovery in databases (KDD) that use categorization techniques to predict learner behaviour is known as predictive analytics.

Researchers investigated how students obtained information from the university's Learning Management System (LMS) in the past (LMS). Classification techniques are used to build an educational model based on Knowledge Discovery in Databases (KDD)[6] in order to predict student behaviour and to predict student performance. It determined the most important factor influencing learners' learning outcomes; it developed prediction models using the J48 decision tree algorithm and multiple linear regression; and it determined the likelihood that Distance Education (DE) students will receive a “Passed” grade in a given course, which could be useful information for teachers and university administrators in course planning. In order to predict the students' final grade based on their prior access to data in the university's learning management system (LMS), the proponents carried out tests[3] [9]. It was discovered that the score obtained via participation in online activities had the most impact on DE learners' learning outcomes, based on the model that had been established. Therefore, the manner in which students interact with the activities provided in the LMS affects whether or not the programme is effectively completed. Thus, the proposed approach may be utilised to identify DE students who need early intervention in order to enhance academic performance and to offer a suitable online learning environment for these students.

Creating an Efficient Cloud Management Architecture for Long-Term Online Learning.

There is a critical component to attaining success in education: efficient and reliable data storage and dissemination systems, as well as the ability to deliver convenient, on-demand education services [3]. To improve the system architecture of lifelong education, we create a cloud-based information system architecture. In order to accommodate the virtualization technology that underpins cloud computing, we present a virtual resource management scheme that includes virtual machine allocation and monitoring node assignment. An e-learning information system for online lifetime education may be constructed and operated using the cloud-based architecture that has been proposed, which meets the requirements for efficiency, reliability, and persistence[7] [10]. In comparison to the existing method, Our suggested approach is able to handle a greater number of e-learning activities (e.g. requesting an LMS navigation, uploading and processing text contents, and processing and delivering text and media contents, as well as video contents) without incurring 48 SLA breaches. According to the evaluation results, our proposed method causes 48 fewer SLA violations than the existing method.

Drivers, Trends, and Challenges in Learning Analytics

Learning analytics is a significant area of technology-assisted learning that has emerged in the last decade. It is comprised of many subfields. [5] The technological, pedagogical, and political factors that have impacted the development of analytics in educational contexts are discussed first in this review of the topic. Next, it describes how learning analytics came to be, covering its origins in the twentieth century, the emergence of data-driven analytics, the development of learning-focused perspectives, and the influence of national economic concerns[7]. In the next section, the connection between learning analytics, educational data mining, and academic analytics is explored. Finally, it examines new areas of research in learning analytics and makes a number of recommendations about possible issues.

Existing System

Our society is still fragile, as shown by the events of 2020, and is susceptible to events that change the assumptions that govern it in a short period of time. This has been shown by a pandemic such as the Corona virus disease of 2019; this global disaster has changed the way people interact, communicate, learn, and work together. Briefly stated, the manner in which society performs all of its operations has changed. This includes education, which has put a bet on the use of information and communications technologies (ICTs) to reach students. As an example, consider the use of learning management systems, which have developed into ideal environments for resource management and activity development. In this project, learning management systems were integrated with artificial intelligence and data analysis technologies in order to improve student performance in the classroom. Specifically, this aim is expressed in the context of a new normal that seeks solid educational paradigms in which certain activities are completed online and Virtual assistants help students study better.

Limitations:

- Society is in charge of everything.
- The nature of the activities has altered.
- Has altered the manner in which individuals engage with one another.

Proposed System

In addition to emotional and factual information, each domain has data that has unique emotional and factual data added to provide domain-specific sentiment information. Domain-specific sentiment expressions may be inferred directly from labelled samples, which have been linked to sentiment labels.In the field of sentiment analysis, it is a common discovery that words that occur more often in happy samples than in sad samples usually represent positive sentiment orientations, and vice versa, according to the findings. This allows us to derive domain-specific sentiment expressions from documents/sentences by propagating sentiment labels from documents/sentences to individual words. Many steps of preprocessing had been accomplished prior to the trials' start. Stop words were removed, and all words were converted to lower case for readability. In this article, we propose to use variations in word distribution both in happy and sad samples to extract sentiment ratings of words, and to do an analysis.

Advantage:

- Ability to generate output from various task learning
- It takes less time to complete.
- It makes use of fewer hardware and software.
- By using several domains, you can determine a person's depressive condition.

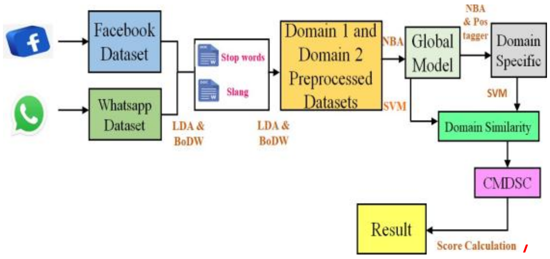


Figure 1. System Architecture

2. METHODOLOGY

- Domain-Specific Sentiment Knowledge
- Domain Similarity Determined by Textual Content
- Domain Similarity Determined by Sentiment Expression
- Sentiment Classification in a Collaborative Multi-Domain Environment

Domain-Specific Sentiment Knowledge

Figure 1 depicts our suggested system design. Positive sentiment orientations are transmitted by words that be seen more often in favourable data samples across all domains, and vice versa for negative sentiment orientations, according to a common finding in the sentiment analysis area. More precisely, denote $s^m \epsilon R^{D*1}$as the sentiment word distribution of domain m, where $s_w^m$is the sentiment score of the word w is important. Then smwis calculated by subtracting the association of word w with a likes of domain m, the correlation of word w with a negative sentiment label in domain m. Using point wise mutual information (PMI) in this research, the connections between words and sentiment labels are assessed, and the sentiment score of word w domain w is defined as

$$s_w^m = \mathrm{PMI}(w, posLabel_m) - \mathrm{PMI}(w, negLabel_m)$$

$$= \log \frac{n(w,posLabel_m)N_m}{n(w)n((w,posLabel_m)} - \log \frac{n(w,negLabel_m)N_m}{n(w)n((w,Label_m)}$$

=lognw,posLabel_mN_m/nwn(w,posLabel_m-lognw,negLabel_mN_m/nwn(w,Label_m

$$= \log \frac{n(w, posLabel_m) n(w, negLabel_m)}{n(w, negLabel_m) n(w, posLabel_m)},$$

Domain Similarity Determined by Textual Content

Content-based domain similarity is thought to be driven by the fact that although distinct domains focus on quite diverse subjects and aim to achieve various goals, domains that are connected have many words in common. and phrases with one another. In the Smart Phone and Digital Camera industries, for example, terms such as "screen," "battery," and "image" are often used and utilised. In contrast, the likelihood of two domains as different as Smart Phone and Book having many common terms is very low. As a consequence, we have developed a technique for comparing domains based on the linguistic information included inside them. According to the work in, we utiliseJensenShannon divergence to determine the similarity of two domains based on their textual word distributions. Denote $\mathrm{d}^m \epsilon \mathbb{R}^{D*1}$ and $\mathrm{d}^n \epsilon \mathbb{R}^{D*1}$ as the term distribution vectors of domains m and n respectively, where D represents the dictionary size. $d_t^m \epsilon$ [0,1]stands for the probability of term t occurring in domain m.

Then the textual content based domain similarity
between domains *m* and *n* is formulated as

$$ContentSim_{(m,n)} = 1 - D_{JS}(\mathrm{d}^m \| \mathrm{d}^n)$$
$$= 1 - \frac{1}{2}\left(D_{KL}\left(\mathrm{d}^m \| \bar{\mathrm{d}}\right) + D_{KL}\left(\mathrm{d}^n \| \bar{\mathrm{d}}\right)\right),$$

Domain Similarity Determined by Sentiment Expression

Based on the textual material provided in the previous section, it may be possible to establish Based on domain similarity, are two domains comparable in their word use patterns. But while textual content similarity often indicates the presence of emotion words, emotion terms do not necessarily appear similarly across domains. Both the CPU and the battery, for example, are examples of electronic hardware. The term "quick" is generally associated with a favourable connotation in the CPU area. “Intel Core i7 is very fast,” for example. In the battery industry, the phrase "rapid" is often used as a derogatory epithet (e.g., "This battery goes out too quickly"). A better match for multi-domain sentiment classification may be found in measuring domain similarity using sentiment expressions as a consequence of this finding. The distributions of emotion words in domains m and n, which are derived from both labelled and unlabeled data in accordance with the previous section, are represented by the symbols pmand$p^m$and pmand p^n. The cosine similarity of their sentiment word distributions is then utilised to determine if domains m and n are comparable in terms of sentiment expressions based on their sentiment expressions.

$$Sentisim_{(m,n)} = \frac{p^m \cdot p^n}{\|p^m\|_2 \cdot \|p^n\|_2}.$$

Sentiment Classification in a Collaborative Multi-Domain Environment

Our model's objective function is convex, which means that in our approach, finding the best sentiment classifiers is the same as solving a convex optimization problem. However, despite the smoothness of the loss function f, the optimization problem is still not smooth due to the l_1--norm regularisation used by the global sentiment classification model as well as the $l_{1,1}$--norm regularisation used by the domain-specific sentiment classification models. Therefore, addressing the optimization problem in our approach effectively and efficiently is a tough task. Our solution to the model of our approach is offered as an accelerated FISTA method that we go into depth on in this article. Also presented is a parallel method that is based on ADMM and which may increase the overall efficiency of our approach when investigating a larger number of domains.

3. RESULTS AND DISCUSSION

DATASETS AND EXPERIMENTAL SETTINGS

In our research, we utilised two benchmark multi-domain sentiment datasets. The first is the well-known sentiment dataset1 (also known as Facebook), which was compiled by people. Facebook and Whatsapp are two of the domains included. It's extensively utilised in sentiment analysis, both multi-domain and cross-domain. There are 1,000 positive and 1,000 negative reviews in each domain.

MEASUREMENTS OF DOMAIN SIMILARITY IN COMPARISON

We performed tests in this section to determine which similarity between the two domains metrics presented is better suited for multi-domain sentiment categorization. Figure 2 shows the experimental findings on the Facebook dataset, while Figure 3 shows similar patterns on the Whatsapp dataset. In these trials, we utilised hinge loss as part of our strategy.

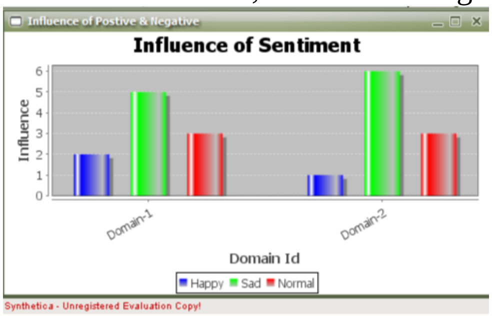


Figure 2. Influence of Sentiment

The effectiveness of our method was evaluated by using various types of domain similarities Using NoSim, ContentSim, and SentiSim, our method' performance is shown by the coloursNoSim (red), ContentSim (blue), and SentiSim (green) (green). SentiSim-Initial and SentiSim-Prop have different start points: SentiSim-Initial is based on sentiment scores that are developed after propagation, while SentiSim-Prop is based on sentiment ratings that are initially collected from labelled samples.

In contrast, we find that our collaborative multi-domain sentiment classification method, which utilises sentiment expressions as the foundation for domain similarity, outperforms text-based domain similarity when using sentiment expressions. It may be concluded that, in a multi-domain sentiment classification job, the likelihood of discovering sentiment similarity based on sentiment expressions is greater than that of discovering sentiment similarity based on textual content..

INTERACTION WITH THE AMOUNT OF TRAINING DATA

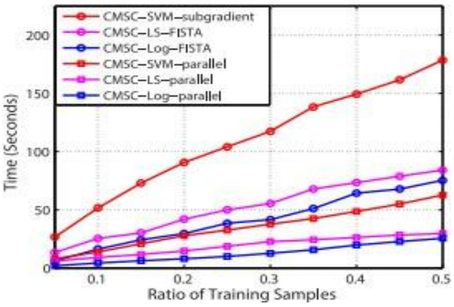
We performed tests in this section to discover whether or not the amount of training data we utilised had an effect on our approach's performance. To finish our task, we'll have removed the need for tagged data. We raised the training set count from 100 to 1,000 while using a 100-sample increment per domain. To determine the loss, we utilise the log loss function.

Our technique, on the other hand, is more than six times better than the LR-all method. For our method, the domain-specific sentiment classifiers it produces are several times better than those generated by LR-single, even though it is trained on labelled data from that domain. With sparse labelled data, the results obtained using our technique are much more substantial, as shown in Figure 2.

TIME EFFICIENCY

We ran many tests to see how complicated our method was in terms of time. Matlab 2014a was used to implement the algorithms. All of the tests were carried out on a desktop PC with a 3.4 GHz Intel Core i7 CPU and 16 GB RAM. The ADMM-based parallel technique was distributed over the four processor cores of this computer, while the single-node version of the FISTA-based accelerated algorithm was executed on a single processor core of this computer. A dataset called Amazon-21 was used for the testing, and the results were published.

Figure 3. Different numbers of training samples



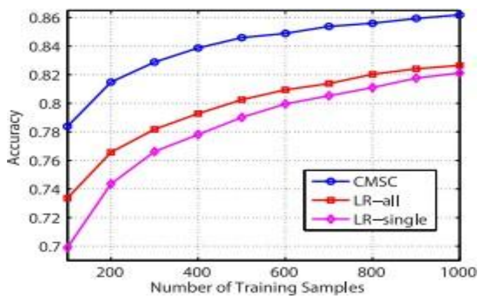


Figure 4. Different ratio of training samples

On two datasets with differing amounts of training samples, the average performance of our approach and baseline methods were evaluated. We refer to our collaborative multi-domain sentiment classification method as CMSC, which stands for Content-based Multidomain Sentiment Classification. These LR-all and LR-single trained classifiers look for a positive and negative sentiment in all the data and in single-domain samples, respectively.

Using various loss functions, we find that our approach's running time varies linearly with the quantity of training data. The results of our previous study on temporal complexity concur with this conclusion. This methodology, along with the simpler hinge loss formula, is noticeably quicker (CMSC-I B). FISTA-based accelerated algorithm delivers significant benefit by increasing overall efficiency. Running the parallel technique also saves time, because it uses a methodology that takes less time to execute than the single-node version optimization methodology. We demonstrate in this article how our parallel method, in which we train sentiment classifiers for various domains concurrently at numerous compute nodes across the network, can accelerate the learning process.

4. CONCLUSION

The multi-domain sentiment categorization system exhibits collaborative multi-domain sentiment processing that is multi-domain in nature. This technique has the potential to train sentiment classifiers that are very efficient for various domains while being done collaboratively, and it can also be used to tackle the issue of under-labelled data because of its applicability across domains. These classifiers have two separate categories: a classifier that operates on the domain, and a classifier that has domain-specific rules. Incorporating domain-specific models for each domain provides for the collection of individual expressions of sentiment that go along with that domain, while using a global model allows for the gathering of sentiment that is shared across all domains. One possible way to better train sentiment classifiers is to include domain-specific sentiment information from both labelled and unlabeled data. Finally, a way to support the global sentiment classifier's learning is to use information about the general sentiment in general-purpose sentiment lexicons. It is also suggested that commonalities across various domains be incorporated in a technique for discovering sentiment associations, as a regularisation over domain-specific sentiment classifiers. Use the convex optimization technique to convert the approach model into a convex optimization problem. Additionally, we see a parallel method to speed up the model of our approach, and an accelerated algorithm to solve it in a shorter period of time. As experimental findings show, the method's improved multi-domain sentiment classification performance, as well as being better than baseline methods, is theoretically possible.

5. REFERENCES

[1] Li, H.; Liu, S.M.; Yu, X.H.; Tang, S.L.; Tang, C.K. Coronavirus disease 2019 (COVID-19): Current status and future perspectives. Int. J. Antimicrob. Agents 2020, 55, 105951. [CrossRef]

[2] 2.Riofrio, G.; Encalada, E.; Guaman, D.; Aguilar, J. Business intelligence applied to learning analytics in student-centered learning processes. In Proceedings of the Latin American Computing Conference, Arequipa, Peru, 19−23 October 2015; pp. 1–10.

[3] 3.Beldarrain, Y. Distance education trends: Integrating new technologies to foster student interaction and collaboration. Distance Educ. 2006, 27, 139–153.

[4] 4.Hssina, B.; Bouikhalene, B.; Merbouha, A. Europe and MENA Cooperation Advances in Information and Communication Technologies; Rocha, A., Mohammed, S., Felgueiras, C., Eds.; Springer International Publishing: Saidia, Marocco, 2017; Volume 520, ISBN 978-3-319-46567-8.

[5] 5.Villegas-Ch, W.; Lujan-Mora, S.; Buenano-Fernandez, D. Application of a Data Mining Method in to LMS for the Improvement of Engineering Courses in Networks. In Proceedings of the 10th International Conference of Education, Research and Innovation (Iceri2017), Seville, Spain, 16–18 November 2017; pp. 6374–6381.

[6] 6.Comendador, B.E.V.; Rabago, L.W.; Tanguilig, B.T. An educational model based on Knowledge Discovery in Databases (KDD) to predict learner's behavior using classification techniques. In Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing, Shanghai, China, 20–25 March 2016; pp. 1–6.

[7] 7.Kim, T.; Lim, J. Designing an Efficient Cloud Management Architecture for Sustainable Online Lifelong Education. Sustainability 2019, 11, 1523. [CrossRef]

[8] 8.Ferguson, R. Learning analytics: Drivers, developments and challenges. Int. J. Technol. Enhanc. Learn. 2013, 4, 304–317.

[9] 9.Lee, S.J.; Lee, H.; Kim, T.T. A study on the instructor role in dealing with mixed contents: How it affects learner satisfaction and retention in e-learning. Sustainability 2018, 10, 850.

[10] 10.Lee, J.; Song, H.D.; Hong, A.J. Exploring factors, and indicators for measuring students' sustainable engagement in e-learning. Sustainability 2019, 11, 985.

[11] 11.Villegas-Ch, W.; Palacios-Pacheco, X.; Buenaño-Fernandez, D.; Luján-Mora, S. Comprehensive learning system based on the analysis of data and the recommendation of activities in a distance education environment. Int. J. Eng. Educ. 2019, 35, 1316–1325.