Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

Movie Recommender System Based on K-Means Dynamic Collaborative Filtering

DR.S.ANANDAMURUGAN[1], V.PRAMOTH[2*], P.M.PRASANNTH RAJA[3], S.RAGUPATHI[4]
[1]Associate Professor, Information Technology, Kongu Engineering College, Perundurai, India
[2]Information Technology, Kongu Engineering College, Perundurai, India
[3]Information Technology, Kongu Engineering College, Perundurai, India
[4]Information Technology, Kongu Engineering College, Perundurai, India
*Corresponding author: pramothv.17it@kongu.edu

ABSTRACT: In today's world, the internet has become an indispensable tool for everyday life. Machine learning models that learn from themselves without being explicitly programmed are becoming increasingly important as the internet grows. Recommendation plays a crucial role in thisworld of computers. One can get lost in a sea of information without guidance. Movies are a popular medium of entertainment as well. Most people watch movies that are suggested by others. Every individual likes different type of movies. So a movie recommender system can apparantly increase viewers count for the movies and also reduces time for searching the movies. Traditional movie recommendation system are based on either Content Based (CB) Filtering technique or Collaborative Filtering (CF) technique that produces result in a minimum number. A combination of the Content Based (CB) and Collaborative Filtering (CF) approach is used here to recommend the most similar movies based on the user's choice. A K-means Dynamic Collaborative Filtering (DCF) approach is used to predict the most relevant recommendation for the user. Many websites provide dataset for the research purpose on machine learning models. Here, MovieLens dataset is used which contains millions of data of movie along with user ratings.

KEY WORDS: *recommender system;content-based;collaborative; hybrid*

1. INTRODUCTION

A recommendation system is a class of information filtering system that predicts a customer's "rank" or "preference" for a particular object. A recommendation system collects information regarding a user's tastes for various things such as movies, shopping, tourism, television, and so on, either indirectly or explicitly.

The use of different online suggestion attributes, such as user, object, and feedback, for recommendation generation is evaluated using DCF-K-Means. To test the efficiency of the proposed system, experiments are carried out using the Movie Lens, Film Trust data collection. This project investigates the possibility of allowing recommendation system designers to interact with each other and exchange reviews of specific monitoring and control algorithms, allowing each designer to receive expected reviews for the algorithms that has not been personally tried by him. Specific algorithms can be replaced with a DCF-K-Means method to release the coupling on web mining user recommendations. With the exponential

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

growth in the field of Web 2.0, an avalanche of reviews of products has gathered on the internet. Customers can get firsthand product details and direct supervision of their buying decisions by reading these articles. In the meantime, manufacturers will get direct reviews and opportunities to improve the efficiency of their lenses. As a result, mining online feedback for movie opinions has become a more urgent practise that has drawn a lot of attention from researchers. Obtaining the overall feeling about a product is not sufficient in order to extract and evaluate movie opinions from online reviews. People, in many cases, prefer to obtainfine-grained sentiments of afeatureof an object that is analyzed.

Customersanticipate to learn not only the overall sentiment of the reviewer but also specific reviews. For example, the reviewer has a favourable thought on the mobile phone‘sdisplay and a negative thoughton the resolution of the display. This goal is achieved by identifying both target opinion and its terms. However, before fine-grained opinion mining can begin, extraction and creationof a list of opinion target and a lexicon for opinion word should be done, which gives information needed for the mining of fine-grained opinion and are the subject of this paper.

Output fromthe existing systems for a new class of data analysis applications known as "recommender systems." Information discovery methods are used in recommender systems to solve the problem of making customised product suggestions during a live consumer engagement. Customer and product growth have exploded in recent years, posing some significant hurdles for recommender systems. For millions of customers and goods, these are delivering recommendations with high-quality and with increased delivery count for every second. Singular Value Decomposition (SVD)-based recommendation algorithms can generate recommendations with high-quality faster, but they require time-consuming matrix factorization.

A method is proposed and experimentally tested for incrementally building SVD-based models, with the potential to make recommender systems highly scalable.

K-Means Dynamic Collaborative Filtering, as suggested, suggests things by matching users with others who share common interests. This method collects feedback from the user in the form of reviews for specific items and looks for patterns in rating habits among users to identify groups of viewers with sameliking. A customer profile is a set of preferences that the viewer has given either implicitly or explicitly. For example, Amazon uses the CF method to recommend products on the basis of its customers' purchasing patterns as well as customer feedback. Each consumer has a variety of things that are rated implicitly or explicitly. By this way, a user-item rating matrix ‘R' is developed, which reflects user choices about products. Different methods are used to locate missing feedback, such as identifying “nearest neighbour” for a newly signed up user in recommending products to them based on reviews given by their nearest neighbours.

2. RELATED WORKS

In the paper work [1] Extracting subject lexicons and sentiment is critical for mining opinions, according to Fangtao Li. For this mission, previous research has shown that supervised learning approaches are superior. On the other hand, the output of supervised approaches, is heavily relied on manually labelled training data. A domain adaptation

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

framework is proposed in this paper for combined extraction of sentiment along with topic-lexicon in an area where labelled data isn‘t required but there is plenty of labelled data in another relevant domain. There are two sections of the system. The first step is to produce a little of sentiment and a few topic seeds with high-confidence in the requiredarea. Secondly, novel Relational Adaptive bootstrapping (RAP) algorithm is proposed owing to lengthen the seeds inrequiredarea using labelled origin data and correlationamong subject as well as sentiment terms. The proposed domain adaptation system can draw out accurate lexicons in the requiredarea with no annotation, according to experimental results.

In the paper work [2] Extracting opinions of the people on specifications of an object, according to Lei Zhang, is an essential task of opinion mining. The emphasis of the paper is on mining features. Double propagation is a cutting-edge method for resolving the problem. It's perfect for medium-sized enterprises. However, this can conclude in low accuracy and retract for both large and small corpora. To address these two issues, two changes focused on part-whole and "no" trends have been implemented to enhance recall. The extracted feature candidates are then subjected to feature ranking to increase the accuracy of candidates with the high ranks. Such candidates are rated according to the importance of their features, which is measured using two properties: frequency and feature relevance. The issue is expressed as a bipartite graph, and the famous HITS page ranking algorithm is employed to search and rate important properties. Experiments on a number of real-world datasets have yielded positive results.

In the paper work [3] Kang Liu has suggested an method of extracting opinion goals on the basis of a word-based translation model (WTM). WTM is, at beginning, used in monolingual setting to explore relationships among opinion goals and opinion terms. Then, using a graph-based algorithm, opinion targets are extracted, with opinion relevance of candidate calculated from mined union combined along candidate significance for producing a common metric. The proposed method can more precisely capture opinion relations by using WTM, particularly for long-span relationships. When observing informal texts in wide Web corpora, the proposed approach will effectively prevent noises from parsing errors when compared to previous syntax-based approaches. Opinion targets are derived in a global process using a graph-based algorithm, which essentially solves the error propagation problem in conventional bootstrap-based approaches like Double Propagation. The proposed methodology outperforms state-of-the-art approaches on dataset of various language and sizes obtained fromthe real-world, according to the experimental results.

In this paper work [4] Guang Qiu has made a proposal. Opinion mining, otherwise called as sentiment analysis, is emerged lately because of its practical uses in difficult research works. In this paper, we look at two notable issues: expansion of opinion lexicon and extraction of opinion target. Opinion targets (aka "targets") are individuals and their characteristics upon which people formed opinions. It is discovered that there exists many syntaxrelated links which connect objectives and opinion words in order to complete the tasks.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

3. PROPOSED WORK

3.1 Preprocessing

Creation of database for recommender system, dataset of reviews i.e. actual reviews is used. Validation of final results is on the basis of use of dataset, so creation of database is one important step. Some websites provides the available datasets which include items and userswith noteworthy rating history, which helps to identifyadequatefigure of highly anticipated items for recommendations to each user.

Five star scale reviews are made for the ratings. 1-star indicates very poor rating. 2-star indicates poor rating. 3-star indicates ok rating. 4-star indicates good rating. 5-star indicates very good rating.

3.2 Rating Prediction

The K-Means DCF recommender system methods proposed in this module include Collaborative Filtering (CF),Content-Based (CB) Filtering, and hybrid approaches. A content-based tactic suggests requiredproducts that are alike to what the viewer has favored previously.

Dynamic Collaborative filtering tactic recommends items that people with samelikings have liked in the past. K-Means can merge both content based filtering (CB) and collaborative filtering (CF) methods. The projectedmethod uses K-Means DCF tactic. While giving suggestions to each user, recommender system performs the following two tasks. First, based on the available information the reviews of unrated items are predicted using some recommendation algorithm. And second, based on the result of predicted reviews the system finds relevant items and recommends them to the target user.

3.3 DCF Item Based Collaborative Filtering

This module takes a collection of items for which a dynamicviewer has given ratings and determines the resemblance among them and the target item, after which it chooses the N most similar items. The corresponding similarities for each item are also calculated. The forecast is made using the items that are the most similar.The actual retrieval and choice of movies from the database is done by this information cleaning module. The filtering process is done based on the facts collected from the learning module,

3.4 Item Similarity Computation

The resemblance calculation amongitem a (target item) and item b is performed in this module by first locating viewers who rated both items. Similarity can be calculated in a variety of ways. The suggested system employs an adjusted cosine similarity approach, which is more advantageous since each co-rated pair is separated from the corresponding user average. Similarity between items a and b is given.

3.5 Prediction Computation Module

In this moduleweighted sum approach is used to obtain the predictions. The approachtotals the prediction of target item for a user u by toting the sum of reviews rated by the user on the items alike to target item. Prediction on an item i for current user k is given Content Based (CB) technique the utility for user k of i^{th}item is assessed based on the valuesdispensed by user u to collection of all items. Only items with high amount of likeness to user's preferences will be recommended.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

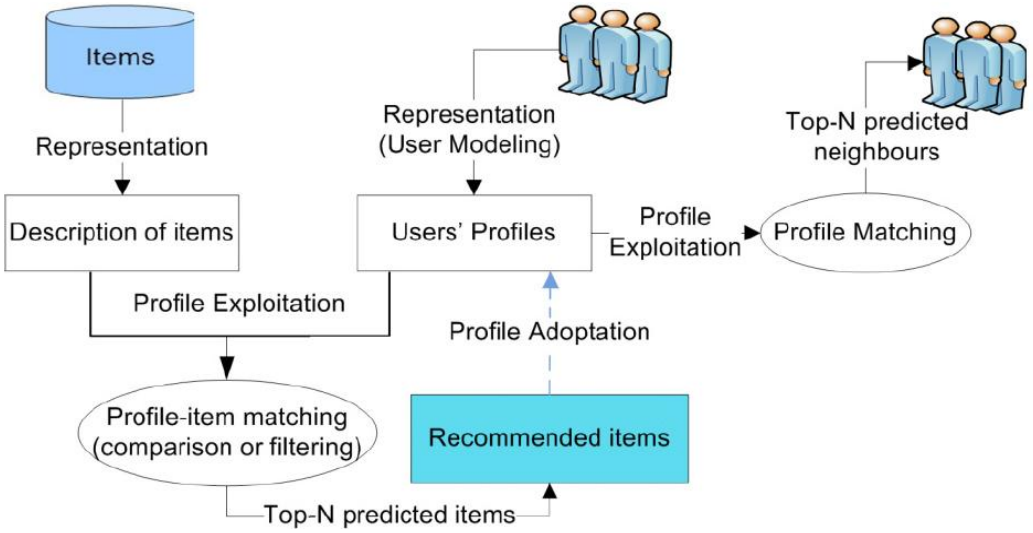

Fig.1 System Architecture Diagram

4. RESULT ANALYSIS

In movie database creation module, information related to user, movies and reviews has been stored in different tables. Thus system can retrieve the data properly from database and also get movie reviews explicitly from the users. In item based collaborative filtering technique, prediction computation and item likeness computation modules have been implemented. Recommended lists are generated on non-purchased movies of login user. So system predicted reviews for all non-purchased movies of login user have been computed. To calculate system predicted rating of target movie, first 5 most similar items are obtained and then used weighted sum approach for rating prediction computation. As per the 5-star scale of rating, predicted value lies between 1 to 5. Mean Absolute Error (MAE) accuracy metric is employedto evaluate accuracy of predicted reviews by this module shown in graph.

The proposed system is tested in Python Language using Jupyter Notebook in system with Intel core i7 processor and 16GB RAM.

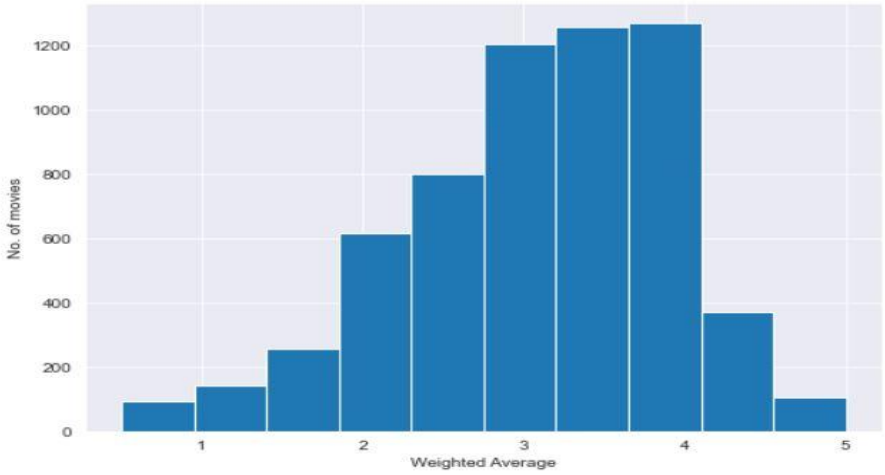

Fig.2 Weighted Average plot

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

The weighted average plot (Fig. 2) shows that the average weight of the ratings is maximum between 3 and 4 indicating that the majority of the movies with rating from 3 to 4 are similar to the target user.

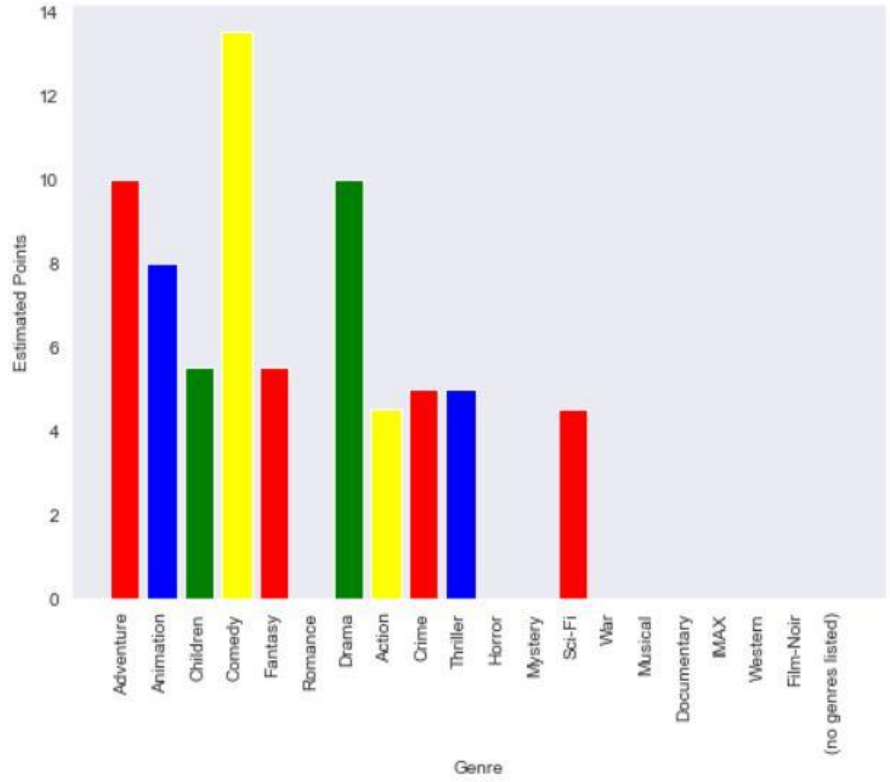

Fig.3 Genre Estimation plot

The genre estimation plot (Fig. 3) shows that the target user prefers movies in the genres as shown in the plot. The maximum bar in the plot shows the most preferred genre for the target user.

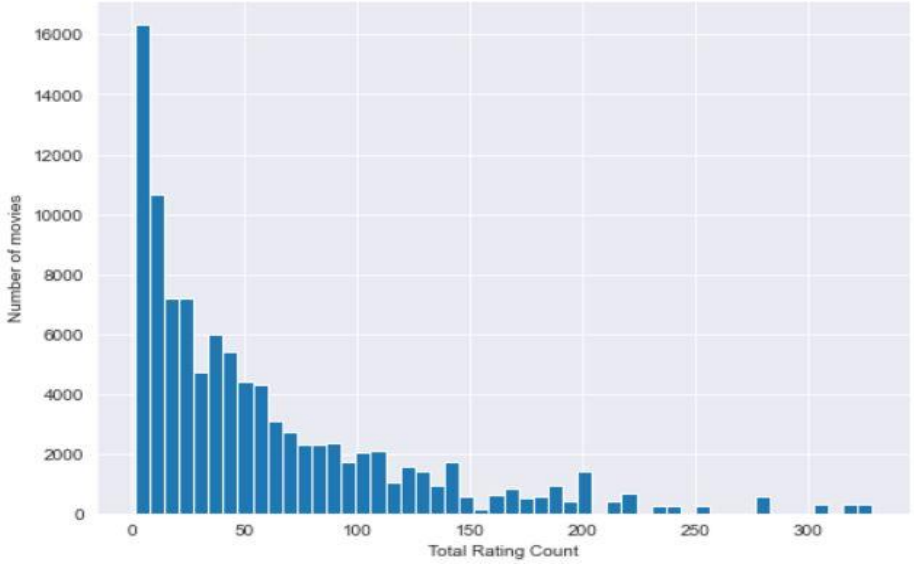

Fig.4 Rating count plot

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 3, 2021, Pages. 6608 - 6615
Received 16 February 2021; Accepted 08 March 2021.

The rating count plot (Fig. 4) shows that the rating count for the majority of the movies is less than 50. This indicates that only minumum number of movies are rated by maximum number of users.

Compared to the traditional approach of content-based filtering and collaborative filtering, it is found that the proposed hybrid approach gives more accurate recommendations than the older approaches. Also the K-means DCF approach is efficient to filter the movies with maximum similarity to the user preferences. The ranking scheme based on the weight sum approach on the maximum watched movies gives result based on not only similarity between the movies, but also the similarity between the users.

5. CONCLUSION AND FUTURE WORK

The project examined the most pressing issues and challenges, as well as what has been done to address them, as well as what should be done in the way of various research works and guidelines for dealing with issues such as sparsity, latency, grey sheep, context-awareness, and the cold-start problem.

In this work, movie recommender system and the internal details of the same is presented. In specific, the mechanism for selection of movies has been stressed on the basis of Collaborative Filtering (CF)along with Content Based (CB) filtering, wherethe stored user choices for different movie dimensions are used. The method approached results in desirable guarantees about the way of recommendations obtained and is also solidly built to variations of user interests.

In multiple ways the movie recommendation system can be refined. The system can be enhanced implementing into social media which gives the feature to recommend a movie to a friend by including friends' data from Facebook. The projectedmethod is not limited to movie recommender systems. In fact, the same can be applied in any domain.

REFERENCES

[1] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain coextraction of sentiment and topic lexicons," in Proceedings of the 50nd Annual Meeting on Association for Computational Linguistics, Jeju, Korea, 2012, pp. 410–419.

[2] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents." in Proceedings of the 23th International Conference on Computational Linguistics, Beijing, China, 2010, pp. 1462–1470.

[3] K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using word-based translation model," in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju, Korea, July 2012, pp. 1346–1356.

[4] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Computational Linguistics, vol. 37, no. 1, pp. 9–27, 2011.

[5] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu," Structure-aware review mining and summarization." in Proceedings of the 23th International Conference on Computational Linguistics, Beijing, China, 2010, pp. 653–661.

[6] T. Ma and X. Wan, "Opinion target extraction in Chinese news comments." in Proceedings of the 23th International Conference on Computational Linguistics, Beijing, China, 2010, pp. 782–790.
[7] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modelling aspects and opinions with a maxent-lda hybrid," in Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, Massachusetts, 2010, pp. 56–65.
[8] A. Mukherjee and B. Liu, "Modelling review comments," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Korea, July 2012, pp. 320–329.
[9] Z. Liu, X. Chen, and M. Sun, "A simple word trigger method for social tag suggestion," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, United Kingdom, 2011, pp. 1577–1588.
[10] Q. Gao, N. Bach, and S. Vogel, "A semi-supervised word alignment algorithm with partial manual alignments," in Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, Uppsala, Sweden, July 2010, pp. 1–10.
[11] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, "Video suggestion and discovery for youtube: taking random walks through the view graph," in Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 2008, pp. 895–904.
[12] P. P. Talukdar, J. Reisinger, M. Pasca, D. Ravichandran, R. Bhagat, and F. Pereira, "Weakly-supervised acquisition of labeled class instances using graph random walks," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Honolulu, Hawaii, 2008, pp. 582–590.
[13] G. Qiu, B. Liu, J. Bu, and C. Che, "Expanding domain sentiment lexicon through double propagation," in Proceedings of the 21st International Jont Conference on Artifical Intelligence, Pasadena, California, USA, 2009, pp. 1199–1204.
[14] R. C. Moore, "A discriminative framework for bilingual word alignment," in Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, 2005, pp. 81–88.
[15] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872
[16] Mr.H.Muthukrishnan "Advent of Disruptive Technologies – Assimilation of Blockchain and IoT and it's Challenges n relevance for the upliftment of Digital Relationship " International Journal of Scientific and Technology Research ISSN : 2277-8616, Volume 9, Issue 4, p.no – 672-676, April 2020
[17] Mr.H.Muthukrishnan, Ms.B.Sunita, Ms.S.Najeera banu, Mr. V.Yasuvanth " Observational study of WPAN and LPWA Technologies for various IoT devices and its applications " International Journal of Advanced Science and Technology, Vol. 29, No. 5, p.no 4231-4243, May 2020