paper 3

by Hanafi Hanafi

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Generate Contextual Insight of Product Review Using Deep LSTM and Word Embedding

To cite this article: Hanafi *et al* 2020 *J. Phys.: Conf. Ser.* **1577** 012006

2nd 2019 ICERA IOP Publishing
Journal of Physics: Conference Series **1577** (2020) 012006 doi:10.1088/1742-6596/1577/1/012006

Generate Contextual Insight of Product Review Using Deep LSTM and Word Embedding

Hanafi[1,2], N Suryana[2] and A S H Basari[2]

[1]Department of Computer Science, Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta 55283 Indonesia.
[2]Faculty of Information and Communication Technology, University Teknikal Malaysia Malaka, Malaka 76100 Malaysia.

Abstract. Nowadays, in every day live, majority people face in many internet options. For example, what meal to eat, what news to read, what vehicle to ride, what the best path to travelling, what the best group in social network to joint, what the best video to watch, what the best video in YouTube to watch and etc. The best way to recommend the internet content to customer is by using recommender system. Recommender system calculate product recommendation by detecting user behaviour in the past. The user behaviour in the past was being variable to compute similarity between many customers. One of the majority user behaviour is in the term of document. Most of document interpret model in recommender system use traditional NLP model such as TF-IDF, LDA model. According to NLP point of view, traditional NLP face the weakness in contextual understanding. Aims to handle the problem on above, we proposed novel model to generate contextual understanding by involve two important aspect considered subtle word and word sequence. We implemented word embedding based on GLOVE and detecting word sequential using RNN-LSTM. According to qualitative evaluation report, our model successful to capture contextual insight of the document of movie review by IMDB. This model suitable to integrated with latent factor based on matrix factorization to generate product recommendation in Collaborative filtering model.

1. Introduction

E-commerce have been widely using in around the world. E-commerce plays important role to make economic growth in worldwide. The successful of e-commerce service influenced of recommender system work. Recommender system responsible to produce product recommendation to customer. According to customer perspective, recommender system has specific work to provide relevant product information. In e-commerce companies, recommender system aims to increase marketing target. Recommender system have been adopted and have successful in many large ecommerce companies for example, 60% YouTube user watching video following to recommendation machine, 80% movie sold on Netflix are also role of recommendation machine[1] [2].

Collaborative filtering is the most and widely use recommender system technique to develop recommendation machine. Collaborative filtering has several advantages where it has ability to provide product information which not only relevant information but the characteristic serendipity information. Serendipity refer to customer feeling surprise to receive product information [3]. This is essential factor to make customer enjoy, fun and happy to find the product in e-commerce portal territory. Different with another method, collaborative filtering

Journal of Physics: Conference Series **1577** (2020) 012006 doi:10.1088/1742-6596/1577/1/012006

exploits user behaviour in the past to calculate recommendation. Rating is the major element as user behaviour record. Rating is representation of user expression about product or service satisfaction. However, the number of rating is too small. This is not easy to collect quite rating from customer. According several literature review, number of rating less than 3 percent and another one said there are less than 1 percent for example Amazon information video (AIV). This problem become major problem in recommender system territory. This problem also popular as sparse data caused minimum rating. Minimum rating influenced the output of recommendation inaccurate or in extreme condition no recommendation result.

Recommender system firstly introduce in the middle 90s. Majority of them implemented recommendation based on statistical approach such as Spearman rank, Cosine similarity, Adjust Cosine similarity and etc aims to calculate behaviour activity in the past. This algorithm method famous called memory based or nearest neighbours. Indeed, memory based having serious problem according to modern recommendation point of view, for example it have scalability issue problem where needed high computation to generate recommendation even if for small population [4] [5]. However, it also has benefit in simplicity to implemented. They have no need training data and develop model or algorithm.

In early 2006, it was beginning novel model that involve latent factor to generate recommendation. This model popular called matrix factorization or model based. Model based rely on matrix factorization to estimated relationship between user and product. One of famous matrix factorization model named SVD (singular value decomposition) where this is model use low rank dimensional to generate correspondent among columns and rows as representation of users and products. Memory based proven better performance over memory based include in accuracy and scalability. However, SVD model face the problem when faced in sparse data and gigantic dataset. So, there are several enhancements for SVD model such as SVD++, TimeSVD++ where there is enhancement in several side information like timestamp information [6]. However, the enhancement based on mathematical approach cannot improve with high impact. Several efforts have been made by many researchers, most of study use auxiliary information of the product also user information to make effective rating prediction. Document belong to product is the most majority involvement to integrated with matrix factorization. The famous approach to interpreted document is latent Dirichlet Allocation (LDA), frequency–inverse document frequency (TF-IDF), Autoencoder (AE) model. Document terminology in this section refer to product description, synopsis, user feedback, abstract, product review and etc. Following to several study, they declared when the involvement of product document proven to increase better performance over traditional model based significantly. However, traditional natural language processing (NLP) such as LDA, TF-IDF and Autoencoder categorical Bag of Word (BOW) mechanism where this approach faces the weakness in contextual insight [7]. For example, traditional BOW cannot distinguish between "the lion chases the Gazelle" and "the Gazelle chase the lion", whereas in the real-world perspective, the second sentence is abnormal. According to traditional point of view, both of the sentence having similar in weight that mean they calculate with similar meaning. The ability to capturing contextual insight believed to make robustness of matrix factorization to correspondent among users and items representation. This research to develop hybridisation between word embedding and LSTM to capture contextual meaning by considering word sequential aspect and subtle word surrounding. This study implemented on movie review that produce by IMDB movie review use more than 10 thousand review. Figure 1 on below is example of movie opinion provide by IMDB portal.

The contribution in this study is to develop novel method by use LSTM and word embedding based on Global Vector for Word Representation (GLOVE) to capture of contextual meaning of a product review, so that it can interpret the consumers opinions more understandable.

Journal of Physics: Conference Series **1577** (2020) 012006 doi:10.1088/1742-6596/1577/1/012006

Figure 1. Example of movie review by customer.

2. Previous Work

Interpreting of document become essential NLP issue recently. It also become important issue in recommender system. Document is the most valuable asset in the internet and big data era. It was caused majority information reserved on document and text. However, the big problem is how to translate and interpreted the document become important and essential document context. Several previous works have been made to increase effectiveness product recommendation. A study [8] first introduce topic modelling as representation of LDA model. They combined review with collaborative filtering and the result of involving text document success to improve rating prediction. Similar with previous study, another study [9] proposed novel model use Topic Modelling to increase ability of interpretation meaning. Both of them categorical LDA model to generate interpretation of document meaning.

Second generation to extract document understanding involve neural language proposed by [10], the researcher exploits autoencoder to extract "tags" document to generate document meaning to combine with matrix factorization, another study [11][12] proposed novel method consider to exploit deep learning method based on autoencoder to extract contextual meaning. Both of model become first collaborative filtering model reach better performance over LDA model. However, autoencoder generally not consider NLP concept to capture contextual meaning. So, both of the model fails to capture contextual understanding because they ignore word surrounding aspect and word sequence aspect to detect document review understanding. The surprise model consider novel deep learning based on Convolutional neural network (CNN) proposed by [13] successful beats another competitor in previous work. CNN model inspired by CNN in sentence classification by [14]. Indeed, CNN is not specific to generate sequential aspect, but they have a bit ability to detect subtle word and surrounding word. This believe the CNN success to improve traditional LDA and topic modelling in recommender system field. Aims to improve the weakness in previous work, we proposed the novel method expected to improve in interpreting of product movie document.

3. Proposed Method

According to previous explanation, this model adopted novel neural language method based on GLOVE. It has objective to produce word vector representation where a word with similar meaning will have located in similar vector. This model also adopted neural model with sequential aspect, it has goal to detect contextual understanding of the document. This neural method famous called Long Short-Term Memory (LSTM) in which it consists several hidden states following time base series.

3.1. Word embedding based on GLOVE

Generally, there are several approaches have been employed by researchers such as word2vec, Glove, Random walk, Star Space, and Fastext. The most popular approach is word2vec and

Glove. The basic principle of both techniques can be observed by geometrical representation in terms of vector of word based on co-occurrence information concerning how often the word rises together in the big corpus data. Both have a distinct concept that word2vec uses the predictive approach and Glove uses by count based technique. The basic principle of predictive model trains the vector aim to increase the prediction perform of loss function that means (goal word, context word, vector representation), the loss value from prediction goal of context word obtained vector value. In case of word2vec, it produces using neural network and optimized as such using SGD (Stochastic Gradient Descent), etc.

3.2. LSTM design
The basic idea behind the emergence of LSTM is to develop a sequential information model mechanism. In the conventional neural network, all input variables are independent towards another. However, for several computational tasks that is not a good idea. RNN was dubbed recurrent due to its work to every layer of sequence in which the result was obtained from the previous computation process. On the other hand, LSTM has memory where it catches information related to the output which has been computed.

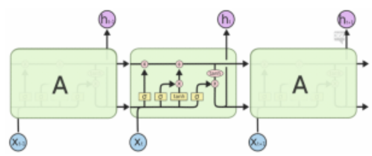

Figure 2. Basic concept of LSTM work.

The generating contextual process consist 300 hidden state LSTM for every product. This scenario following to previous study that including text document of product. The design of contextual extraction use LSTM model shows on figure 3 below.

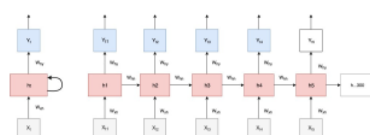

Figure 3. Design of LSTM to capture contextual insight.

3.3. GLOVE-LSTM model
Hybridization between GLOVE and LSTM show on figure 4. This model consists 4 process to gain word vector 2D space with contextual ability. Generally, preprocessing required to transform text document into numeric. The numeric can be processed GLOVE training. There is some option to transform text document into 50, 100, 300-dimension 2D space. The output of GLOVE will be process into generating contextual understanding process by LSTM

3.4. Text Document Pre-processing
Text processing is needed to fulfill several things namely standardization so that the text document is recognized by the neural network to be computed. In this preprocessing the Author uses the NLTK module which is available in python. Some important things in this preprocessing

Journal of Physics: Conference Series **1577** (2020) 012006 doi:10.1088/1742-6596/1577/1/012006

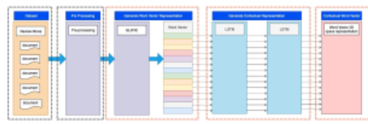

Figure 4. Combining GLOVE and LSTM

are tokenizer, removing stop word, punctuation, lemmatization. Detail pre-processing stage in this experiment show on Table 1

Table 1. Document pre-processing scenario

No	Text Pre-processing Strategy
1	Schematic maximum length of sentences to 300.
	This step very important to grasp the context of sentences in the document review of a product. Indeed, the longer number of words will increasingly accommodate the meaning of words that can be captured, but it will burden the computational stage in the training set process. A maximum 300-word decision is a fairly conservative number to represent the meaning of the sentence in a fairly conservative document
2	Remove stop words in the document.
	This step is intended to eliminate unnecessary characters and also the role of characters that are not important so that it does not affect the meaning and burden the computational process.
3	Compute TF-IDF value for every sentence.
	This step aims to eliminate the number of words that often appear in overall document review collections from Amazon. The function of this step will be explained in the next step as the goal for this step.
4	Delete corpus data word collection of special stop words belonging to the document with frequency of higher than 0.5.
	Based on the passage mentioned in the previous step, this step needs to be done to avoid words that appear too often and are very dominant, because it will cause a single meaning that is also very dominant.
5	Delete all non-vocabulary words from table documents.
	As a step created above, removing non-vocabulary words also means deleting the variable that is less meaningful. It also aims to reduce the computational burden on processing the meaning of documents in product reviews.

4. Result and analysis

This section contains 3 particulars of essential test include GLOVE training result in the term of word vector representation diagram, qualitative test to detected contextual understanding of product review where specifically having ability to distinguish normal context and abnormal context of the sentence and transform contextual document into 50-dimensional 2D space for every product. The detail our experiment result show in section below.

4.1. Result of GLOVE output vector
The output obtains by GLOVE in the term of blue dot as representation of word vector. The distance among dot show the distance of similarity meaning. The output result from movie review document involve 10 thousand review by customer show figure 5. The further process, the word vector representation compute to generating contextual insight using LSTM model.

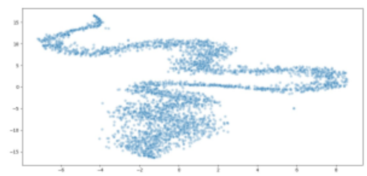

Figure 5. Output of GLOVE word vector

4.2. Test of similarity word vector
This section tries to detect the closest similarity word when raises word inputting to the model. Table 2 below show the example list of similarity closest word gain by GLOVE. For example, the output when typing Bill Gates then the output is identic with Microsoft. As we know Bill Gate is the owner or founding father of Microsoft company. And second example when we tried to input Messi, the output is Barcelona. Indeed, Messi is one of the biggest football stars in Barcelona.

Table 2. Similarity word vector identification test

Relationship	Example 1	Example 2	Example 3
Mr. Bean-Comedy	Rambo-Actio	n Tom & Jerry-Cartoon	Romantic-Titanic
Bill Gate-Microsoft	Steve Job-Apple	Samsung-Korea	China-Great Wall
France-Paris	Italy-Rome	Japan-Tokyo	Indonesia-Jakarta
Tesla-Electric	Messi-Barcelona	Ronaldo-Madrid	Hitler-Nazi
Italy-Roma	Picasso-Painter	Obama-Barack	Uranium-Nuclear
Japan-Sushi	Samsung-Handphone	PS2-Game	s Natal-Sinterklaas
World Cup-Frances	Honda-Automobile	Iraq-war	Google-Android
Einstein-Scientist	James Bond-007	DiCaprio-Leonardo	Rambo-Stallone
Spiderman-Return	Dracula-Horror	Ferrari-Speed Car	Oscar-Award

4.3. Qualitative test of contextual insight

The evaluation is to ensure and confirm the ability of GLOVE-LSTM in capturing contextual meaning. One parameter is that the model can give a high score if an expression has a contextual meaning and vice versa will give a low score if an expression does not have a contextual meaning. Some examples of input results on the GLOVE-LSTM training results as the table below. For example, in the story of Romeo and Juliet, normally the story is a romantic story. If in the context of the story is converted into an expression that gives the opposite meaning as an example of hatred, then the machine will give a low score. Because high scores will have more influence on the matrix calculation. Some examples of testing can be seen in the Table 3.

Table 3. Qualitative test of phrase

Phrase capture use	**weight**	**Phrase capture use**	**weight**
Romeo love Juliet	0.0678	Rambo loss fighting	0.0056
Juliet miss Romeo	0.0391	Rambo so romantic	0.0089
Juliet dear Romeo	0.0374	Micky chase Donald	0.0109
Juliet hate Romeo	0.0054	Donald chase Micky	0.0781
Romeo betray Juliet	0.0047	Hulk strong man	0.0541
Romeo believe Juliet	0.0618	Hulk weak man	0.0051

4.4. Output of word vector 2D representation

We choose two outputs of document interpreted that generated from the GLOVE-LSTM model can be seen in Table 4 on below. In the table can be seen that a product document can be transformed into 50 dimensional spaces. This 2D vector space will be hybridized by matrix vector in collaborative filtering model. This is made to support the matrix in dealing with sparse data issues. This experiment will be carried out in subsequent study.

Journal of Physics: Conference Series **1577** (2020) 012006 doi:10.1088/1742-6596/1577/1/012006

Table 4: Qualitative test of phrase

No.	Movie Review Document	Output Word Vector
1	Sentence: criminal pleads insanity after getting into trouble again once mental institution rebels against oppressive nurse up scared criminal past once again gotten himself into trouble sentenced escape labor duties pleads insanity sent ward mentally both stands witness abuse oppressive gains power through other other inmates band together make rebellious against few months left year old convict serving time several just been transferred labor camp associated psychiatric been able use acting having attitude behavior benefit not having do any at hospital authorities at want him psychiatric prove not believing this all act get out believes this stint will get him out any more work while serves out remainder placed ward group men have control their mental continues same manner always get what using other patients either or things own adds list goals	Last RNN State: [0.00084321, 0.00144706,-0.00401044, 0.00257065, 0.00225992,- 0.00073185, -0.00043757, 0.00269554, 0.00116859,-0.00396964,-0.00154141, 0.00306911, 0.00363265, 0.00084947,-0.00007796, 0.00240284,-0.00161516, 0.00081176, -0.00228911, 0.00423318, 0.00229056, 0.00006444, 0.00725766,-0.00197187, 0.00206384, 0.00568174, 0.0001365 ,- 0.00154586,-0.00441508,-0.00533816, 0.00196574, 0.00324589, 0.00375766, 0.00469626, 0.00416555,-0.00060936, 0.00483586,-0.00046192, 0.00283102,- 0.00127622,-0.00195044, 0.00425937, 0.00420238, 0.00051502, 0.00319298, 0.00017351, 0.0027331 , 0.00112974, -0.00506582,-0.00692325, 0.00008129, 0.00001952,-0.00037403,-0.00152063, 0.00279147,-0.00101742,-0.00162607, 0.00498988,-0.00526247,-0.00395839, -0.00594708,-0.00580164,-0.00043284, 0.00224017, 0.00209852,-0.00112896, 0.00158129,-0.0033648 ,-0.00024615,- 0.00739116, 0.0006688 , 0.00177063, 0.00143779, 0.00642913,-0.0039397 ,- 0.0033999 , 0.00128422,-0.00144704, -0.00166969,-0.00243981, 0.00022438, 0.00475973, 0.00146994,-0.00059578, -0.00237243, 0.00010416,-0.00467012,- 0.00235825, 0.00054439, 0.00439096, 0.00143782, 0.00132802, 0.00580835, 0.0012416 ,-0.00224719,-0.00016983, 0.00249277,-0.00083794,-0.000737 ,- 0.00126232]

2	Sentence: lives two cruel befriends bugs live inside giant they embark journey happy life at seaside ended parents are killed goes live two saving life spider comes into possession magic crocodile after which enormous starts grow inside meets not only spider number new friends including help him plan try get	Last RNN State: [0.00107973, 0.00508082, 0.00355901,-0.00392048, 0.00611115,- 0.00105035, 0.00354371,- 0.00966446, 0.00331892, 0.0020839 ,- 0.00112833,-0.00262576, - 0.00150739, 0.00158827,- 0.00109174, 0.00498265, 0.00619777, 0.00736556, 0.00465961,-0.00887842,-0.00232945,- 0.00288069, 0.00865848, 0.00265919, -0.00288546,-0.00359717,- 0.00173427, 0.00039289,-0.00433664, 0.00141917, 0.00131208,-0.00463233, 0.00023525, 0.00008202,- 0.00039931,-0.00463334, -0.00153508,-0.0030069 ,- 0.00524354, 0.00532046, 0.00161104,-0.00190012, 0.00397158,-0.0018483 ,-0.00439018,- 0.00081055, 0.00060273,-0.00473418, 0.00695798,-0.01014331,- 0.00304296, 0.00392223, 0.00128475,-0.00272423, -0.00094031,-0.00424525,-0.00291265, 0.00359614, 0.00520504, 0.0126545 , 0.00013376, 0.00400501,- 0.00169136, 0.00419567,-0.00145644, 0.00102336, - 0.00849879, 0.00193416,-0.00341029, 0.00583275,- 0.00072109, 0.00453854, -0.00224958,-0.00063251,- 0.00175551,- 0.0014703 ,-0.0057299 , 0.00210403, 0.00053354,-0.00335856, 0.00210629,- 0.00142628, 0.00077934, 0.00337293, 0.00050787,-0.00156855,- 0.00167821, 0.01166623,-0.00419289,-0.00184923, - 0.00190834, 0.00172968, 0.00561418,- 0.00109293, 0.00209859,-0.00465574, -0.00192821,-0.00100691, 0.00093099, 0.00264098]

5. Conclusion

Based on several evaluation tests that the model we developed with the named GLOVE-LSTM successful to capture the contextual meaning of product document. It is characterized by being given a high score for expressions that have a contextual meaning while giving low scores to expressions that are not common. It can be concluded that the model involve GLOVE and LSTM successfully captures contextual semantic insight. The deep understanding of user opinion expects to increase robustness of matrix factorization to handle sparse data

6. References

[1] Carlos A Gomez-Uribe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19, 2015.

[2] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296, 2010.

[3] Charu C Aggarwal et al. *Recommender systems*, volume 1. Springer, 2016.

[4] Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[5] J Ben Schafer, Joseph A Konstan, and John Riedl. E-commerce recommendation applications. *Data mining and knowledge discovery*, 5(1-2):115–153, 2001.

[6] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[7] Charu C Aggarwal. *Machine learning for text*. Springer, 2018.

[8] Guang Ling, Michael R Lyu, and Irwin King. Ratings meet reviews, a combined approach to recommend. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 105–112, 2014.

[9] Chong Wang and David M Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 448–456, 2011.

[10] Hao Wang, Xingjian Shi, and Dit-Yan Yeung. Relational stacked denoising autoencoder for tag recommendation. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[11] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1235–1244, 2015.

[12] Jian Wei, Jianhua He, Kai Chen, Yi Zhou, and Zuoyin Tang. Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69:29–39, 2017.

[13] Donghyun Kim, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 233–240, 2016.

[14] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

Acknowledgments

We would to thank for institution who always supporting us to funding our research. Our great thank Time Excelindo company, Universitas Amikom Yogyakarta and University Teknikal Malaysia Malaka (UTeM).

paper 3

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11