

Research Methods and Professional Development 2203588 dongbo

by Bo Dong

Submission date: 25-Apr-2023 02:48PM (UTC+0100)

Submission ID: 203371738

File name:

42017_Bo_Dong_Research_Methods_and_Professional_Development_2203588_dongbo_493226_388816847.docx
(149.88K)

Word count: 4238

Character count: 24023



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

University of Wales Trinity Saint David

School of Computing

2023.4.25

Research on network public opinion prediction based on machine learning

Directory

1. Research background	1
2. Research significance	2
3. Purpose of the study	2
4. Literature review	3
4.1 Emotional tendency analysis	3
4.2 Public opinion trend forecast	4
4.3 Introduction to the analysis method of emotional tendency	5
4.4 Introduction to the public opinion trend prediction model	6
4.5 Predictive model based on fuzzy neural network	6
4.6 Relevant predictive models based on gray theory	7
5. Methodology	8
6. Results	8
7. Evaluation	8
8. References	9

1. Research background

The Internet platform has the characteristics of fast transmission speed, wide dissemination and strong interaction. According to the 49th Statistical Report on Internet Development in China, the size of China's Internet users reached 10.32 billion, an increase of 42.96 million from December 2020, and the internet penetration rate reached 73.0%. Internet platforms provide users with a diversified and open discussion environment, and these subjective information contain a large number of potential public opinion orientations, and network contradictions and conflicts have a potential impact on social stability.

Network public opinion is mainly divided into enterprise-related public opinion and official-related public opinion. The crisis of online public opinion has an important impact on the promotion and expansion of the enterprise crisis. For example, negative information on

enterprise operations, judicial disputes and labor disputes form a dispersive network communication structure through media platforms[1]. The public opinion of government affairs officials has attracted widespread attention from society because of its unique topicality [2]. Negative and mixed with false news will have a negative impact on society, thereby reducing the credibility of the government and damaging the image of the government[3]. Therefore, predicting and dealing with online public opinion crises in advance can quickly answer people's doubts, maintain social order, and improve government governance systems and governance capabilities[4].

Traditional online public opinion monitoring relies on manual visual information recognition, and judges the development trend of public opinion after sorting and summarizing, which consumes a lot of manpower and time. In recent years, the method of predicting online public opinion trends based on machine learning methods has become a new type of online public opinion monitoring method. The nonlinear forecasting model based on machine learning method combines artificial intelligence technology and time series analysis to achieve public opinion trend prediction[5]. Machine learning methods can automatically and quickly analyze public opinion from massive amounts of information, and make timely predictions of public opinion development, helping local governments implement corresponding processing measures, in order to improve the credibility of the government. Therefore, comprehensively analyzing the main methods of machine learning in network public opinion prediction and actively exploring its key technologies is of great practical significance to the healthy development of society.

2. Research significance

Due to the subjectivity and freedom of online comment information and the popularity of online media, more and more netizens will express their views and opinions on hot events on the Internet. In the past ten years, Chinese netizens have become increasingly involved in online public opinion. Looking at home and abroad, many events are derived from the fierce debate on the Internet, forming a public opinion force in the society, and eventually triggering public opinion warning, and the public opinion generated by hot events itself will also have a huge impact on the figures, companies or brands that trigger public opinion. It can be seen from this that timely judgment and prediction of the development trend of public opinion is of great significance to public opinion monitors. However, due to the increasing number of Internet users and its membership structure, which includes almost all age groups, this also brings challenges to the analysis of online public opinion. Therefore, the analysis of netizens' emotional tendencies and the study of the trend prediction of public opinion has become a step that cannot be ignored in the process of public opinion analysis, followed by the effective analysis of netizens' emotions in hot public opinion events and the timely prediction of the development trend of public opinion events, which is of great significance to promote social harmony while ensuring a harmonious network atmosphere.

3. Purpose of the study

Through the network public opinion trend prediction method based on emotional tendency analysis, the network public opinion trend prediction trend is more accurate.

4. Literature review

4.1 Emotional tendency analysis

In the study of emotion classification based on emotional dictionaries, research is mostly based on the constructed emotion dictionaries [6], and the analysis of sentence structure can also be improved on the basis of the original emotion dictionaries, in order to improve the accuracy of emotion classification. Among them, Xu Linhong and Lin Hongfei divided the sentences into simple sentences with single emotional words, simple sentences with multiple emotional words, and complex sentences with multiple emotional words according to the complexity of expressing emotions. Then, from the aspects of sentence vocabulary and structure, 9 semantic features affecting the emotional color of the sentence are extracted, and the semantic features are added to the conditional random domain on the basis of the existing emotional vocabulary ontology and sentence-level key emotional vocabulary recognition, and then the article-level sentiment analysis is carried out [7]. Zhang Yang et al. proposed a method based on the combination of emotional description items and improved mutual information calculation, extracting a number of matching patterns that can obtain text emotional description items through syntactic analysis, and training a classification model to obtain the praise and disapproval of text according to pattern matching and calculating the mutual information of emotional description items as feature values. By analyzing the results of the hotel and mobile phone corpus set experiments, the method has good effect [8]. Liang Xiaomin et al. used dependency syntactic analysis to propose a public opinion object analysis model, identify and extract public opinion object-emotional word pairs, conduct sentiment analysis, and study the relationship network of public opinion objects, which can effectively identify the main public opinion objects and their emotional words, and intuitively display the emotional expression and relationship cognition of netizens on the evolution of public opinion objects over time [9].

In the research of sentiment classification based on machine learning, the required data is extracted and trained by related technologies in machine learning, and then an analytical model based on the combination of machine learning methods and corpus data is proposed. Among them, Li Tianchen et al. used the LDA (latent Dirichlet allocation) theme model, first clustered the text, and then applied the recurrent neural network method on each topic subclass to establish a theme model for positive and negative emotional samples, using this method, by dividing the subcategories, the expression of the text under different themes was standardized, and the problem that the meaning of words under different themes was easy to change. Moreover, the method of training language model is used to avoid the difficulty of directly extracting features, and finally joint judgment is made based on the probability of belonging to the theme and the emotion [10]. Sun Yan et al. proposed a sentiment analysis model that integrates language model-based word embedding (ELMo) and multi-scale convolutional neural network (MSCNN). The model uses ELMo to learn pre-trained corpus and generate context-related word vectors. Compared with traditional word embedding technology, ELMo uses the LSTM network to fuse the characteristics of words themselves and the context characteristics of words, which can accurately represent multiple different semantics of polysemy words. This method can speed up the training speed of the model and improve the training accuracy [11]. Yin Guanghua et al.

proposed a method for extracting emotional elements of Internet short text based on long and short time memory network model. The Encode-Decoder sequence annotation framework model for the extraction task of Internet short text emotional elements is mainly constructed by using the long and short time memory network model, and the three-membered window emotional feature selection is integrated on this basis[12].

In summary, whether it is based on emotional dictionary or based on machine learning, the emotional tendency research is focused on the content of the commentary, and mainly studies the text itself, and the changes about the emotion itself are diverse, and the emotional tendency contained in it also needs deeper research and discussion, only focusing on the emotional tendency displayed by the text is lacking in multi-type factors, and the accuracy of judgment based on the above consideration of emotional tendency needs to be improved. Based on the previous research only focusing on the data of comment Chinese, this paper considers the influence of expression factors in comment information on the emotional tendency of public opinion events, and at the same time uses the self-attention mechanism to effectively pay attention to the characteristics of important emotional words in the data, and proposes to study the emotional tendency of Weibo hot topics based on the Emo-ATT model, which improves the accuracy of the entire emotional tendency classification compared with other models that only consider text information.

4.2 Public opinion trend forecast

The existing network public opinion prediction analysis is mainly based on the analysis of time series models, using mathematical methods to establish forecasting models, which is a quantitative analysis method [13]. At present, most network public opinion prediction models are mainly divided into two categories, one is a linear prediction model based on traditional statistics, and the other is a nonlinear prediction model based on intelligent machine algorithms.

Linear prediction models mostly predict public opinion trends based on the general law of the evolution process of public opinion, combined with mathematical formulas. Among them, Shi Rui et al. used particle swarm algorithm to optimize the parameters of the GM(1,1) power model, and combined with the GM(1,N) model to build a combined gray model for predicting the development trend of network public opinion heat [14]. Luo Taiye's research on the prediction of the development of microblog public opinion hotspots based on the logistic model can effectively predict the development of microblog public opinion hotspots in the self-organization state[15]. Zheng Buqing proposed a prediction method based on inflection point on the basis of ARIMA and gray prediction on the basis of ARIMA and gray prediction, and established a mathematical model of segmentation and image processing [16]. In the analysis of the development law of Weibo public opinion data, Xu introduced the mathematical prediction method of exponential smoothing method, which can effectively exclude the impact of data fluctuations and fit the general trend of public opinion trend based on past data[17].

The nonlinear prediction model based on intelligent machine algorithm mainly combines artificial intelligence technology and time series prediction to achieve public opinion trend prediction. Based on the Markov model, Wang Yamin et al. solved the state transition matrix by using the combined model of genetics and quadratic programming, and used the dynamic error compensation formula to correct the experimental results to improve the prediction accuracy [18]. Huang Yaju, Chen Fuji and others successively constructed a sequence prediction model and

a back propagation (BP) neural network prediction model for fuzzy logic events, and finally improved the overall accuracy of event prediction by combining them, and completed the early prediction of public opinion events [19-20]. Wang Ning et al. used the EGM(1,1) model to predict the development trend of public opinion events, and based on the prediction data, they proposed a grading scheme for online public opinion events by using the gray correlation analysis method [21]. Sun Jingchao et al. designed a network public opinion model based on recurrent neural network adaptive learning rate, selected a variety of features according to the characteristics of public opinion data to construct a recurrent neural network sequence generation model, aiming at the problem of difficult convergence of recurrent neural network model, adaptive adjustment of learning rate through continuous optimal coin throwing strategy to improve training speed and prediction accuracy [22].

In summary, although there are many research algorithms on public opinion prediction, there are still some shortcomings that need to be improved. First, most of the online public opinion trend prediction models are mainly linear models, mainly based on the characteristics of public opinion data itself as the research starting point, but with the growing Internet population, public opinion data changes rapidly, making linear prediction accuracy low. Second, the use of nonlinear models to predict public opinion trends is the focus of current research, mostly through artificial intelligence technology and neural networks to achieve dynamic prediction, but due to many uncertainties in network public opinion events, it will lead to high computational complexity and gradient explosion in the calculation process, which makes the prediction effect poor. In order to improve the accuracy and timeliness of network public opinion event prediction, this paper proposes to integrate SDZ (suprisal-driven zoneout) method in long-short term memory (LSTM) network to improve the accuracy of public opinion trend prediction.

4.3 Introduction to the analysis method of emotional tendency

The emotional tendency analysis method based on emotional dictionary is relatively simple, using unsupervised learning, and its mathematical model is bag-of-words model [23]. This method judges the emotional polarity of the whole sentence by traversing emotional words and matching according to the dictionary matching, so whether the emotional polarity judgment can be benchmarked depends very much on the construction of the emotional dictionary. If the emotional word divided in the sample cannot be accurately located in the emotional dictionary, the word does not have emotional polarity and will affect the emotional tendency of the entire sentence. Then it is required that when building an emotional dictionary, it should strive to include all emotional words as completely as possible. Among the emotional tendency analysis algorithms based on emotional dictionaries, the more commonly used ones are Dalian Polytechnic Emotional Thesaurus, CNKI's HowNet Dictionary, TEDA's Chinese Simplified Dictionary, etc., while domestic research related to emotional discrimination is an extended dictionary and a dictionary-based application. Yuan Haixia et al. used word2vec and How net to expand the existing emotional dictionary, which had a significant effect in the analysis of emotional tendency, the method first used word2vec to train word vectors, and then calculated the similarity with candidate words to obtain the most similar 10 words, and used these 10 words to analyze emotional tendency; At the same time, the hownet method is used to complete the content that word2vec cannot judge, and finally obtains the emotional tendency of the sentence

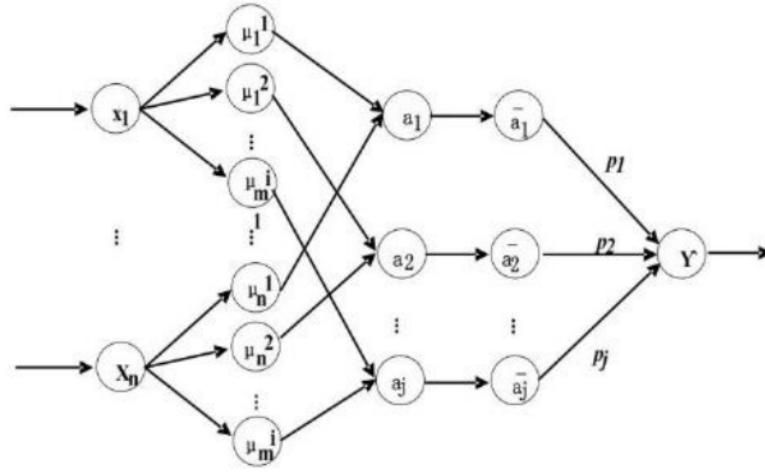
[23]. Wang Lidong et al. used the original function of the hownet thesaurus to propose a semantic retrieval method, which matched the similarity between the word vectors extracted in the Weibo text and the words in the hownet thesaurus system, selected the most similar statements according to their matching degree and returned the results, providing users with relatively satisfactory search results [24]. Yang Zhihao et al. proposed an emotional tendency analysis method based on word similarity through the semantic tendency calculation method. Using CNKI Dictionary, they pioneered a method of polar emotional tendency analysis based on the establishment of positive and negative reference words, which focused on the average semantic similarity between candidate emotions and positive and negative emotional words, and greatly improved the accuracy of sentiment classification by calculating the similarity between words [25]. Wang Shihua et al. analyzed the emotional tendency of Weibo text on the basis of the synthetic emotional dictionary with good word frequency statistics, used the coefficient of variation method to calculate the weight of blog posts, calculated the influence of Weibo, and proposed a microblog leader identification method based on sentiment analysis on this basis, which effectively used the sentiment analysis technology. Finally, the PageRank algorithm is used to predict the diffusion process of user influence in the Weibo user forwarding relationship network, and calculate the size of the user's final influence [26].

4.4 Introduction to the public opinion trend prediction model

At present, the research on public opinion trend prediction is roughly divided into models based on traditional statistics and prediction models based on machine learning. Based on traditional statistics, the more common ones are exponential smoothing model, moving average model, exponential smoothing model and ARIMA model. At this stage, most of the mainstream research models based on machine learning are used in the research on public opinion trend prediction, and this article will briefly introduce the more common models based on gray theory, Markov model and fuzzy neural network model.

4.5 Predictive model based on fuzzy neural network

Fuzzy Neural Network (FNN) is a fusion of fuzzy theory into neural networks, which is conducive to improving the speed of information processing and improving the adaptive ability and fault tolerance of the network [32,33]. Fuzzy neural networks have nonlinear mapping capabilities and self-learning capabilities, making them research hotspots in many fields. The structure of the fuzzy neural network is shown in the figure, and the specific functions are described as follows:



First layer: Input layer. The node represents the input of the variable, where x_i represents the i th input variable, and the number of input layer nodes n is equal to the number of input variables;

Second layer: blur layer. This layer first fuzzes the input variables, and then uses the membership function to calculate the membership degree μ_i ($i=1,2,\dots,n; j=1,2,\dots,m$), the number of nodes of i is equal to the sum of the number of fuzzy gradings m_i for each variable;

The third layer: the fuzzy inference layer. Each node has certain fuzzy rules, and the number of nodes is equal to the number of fuzzy rules. a_j represents the incentive strength of the current input for each rule, and the calculation formula is shown in Equation (2.13):

$$a_j = \prod \mu_i^j$$

The fourth layer: the normalization layer. This layer needs to normalize the incentive strength a_j of each rule, and the number of nodes is equal to the fuzzy inference layer. The normalization process is calculated as shown in Equation (2.14):

$$\bar{a}_j = \frac{a_j}{\sum_{j=1}^m a_j}$$

Fifth layer: Output layer. The final inference value Y of the neural network model is calculated, and the number of nodes in this layer is calculated as shown in Equation (2.15):

$$Y = \sum_{j=1}^m p_j \bar{a}_j$$

4.6 Relevant predictive models based on gray theory

It mainly studies the uncertain system, and predicts the system with uncertain information, and the effect is better. The gray prediction method is to find the law affecting the change of the system according to the regularity displayed by the original time series data, and finally produce a regular system, and establish the corresponding differential equation to realize the trend

prediction of the public opinion data [31]. In order to reduce the impact of time series randomness on the forecasting system, it is often necessary to accumulate or subtract the data in the early stage of model building.

Additive generation: The generated column is calculated by the accumulation of the original sequence. Its rule is that the first data remains unchanged as the first data of the generated column, and the second data is accumulated to the first and second data of the original sequence, and then so on to obtain the generated sequence. Note that the original time series is $X(0)$, as shown in equation (2.10):

$$X(0) = X(0)(1), X(0)(2), \dots, X(0)(n) \quad (2.10)$$

The resulting column is shown in Equation (2.11):

$$X(1) = X(1)(1), X(1)(2), \dots, X(1)(n) \quad (2.11)$$

Progressive generation: Subtract the adjacent before and after data in the original sequence to obtain the generated column. The calculation

The method is the inverse of accumulation, and the accumulated generated column can be reduced to non-generated column, and incremental information is obtained to make the model construction more perfect.

The method of one decrement is shown in Equation (2.12):

$$X(1)(k) = X(0)(k) - X(0)(k - 1) \quad (2.12)$$

Grey prediction has the characteristics of not requiring a large amount of data, can solve the problems of lack of historical data, incomplete series and low data reliability, and is an effective way to deal with the problem of small sample prediction, and has the characteristics of simple operation, easy to test and no need to consider the distribution law, but its prediction results for series with large volatility are poor.

5. Methodology

Mainly through the network public opinion trend prediction method based on emotional tendency analysis, the implementation of the method mainly includes two models. Firstly, the emotional tendency of Weibo hot topics is analyzed, and secondly, according to the emotional tendency of the hot topics analyzed above, the trend prediction based on LSTM public opinion trend prediction model is used to use the proposed LSTM public opinion trend prediction model for hot topics with strong emotional performance and easy to generate public opinion events.

6. Results

Pick hot event commentary information, and the study will collect data through auxiliary tools. Organize data into tables for basic text preprocessing. The data is divided into a training set and a prediction set, and then sentiment analysis is carried out to bring into the public opinion trend prediction model.

7. Evaluation

It can have a deeper understanding of the emotional tendency of netizens to public opinion hot events, timely grasp the dynamic trend of public opinion, so that the regulatory authorities can effectively guide the attitude and tendency of netizens correctly, timely sort out public opinion information, this paper on the basis of emotional tendency analysis, the hot events that

are easy to cause public opinion alarm to predict its public opinion development trend. At the same time, with the advantage of sudden public opinion events from the long and short-term memory network, the development trend of hot events is predicted, so that the entire public opinion analysis process can provide the network supervision department with a timely and accurate grasp of the attitude of netizens and the trend of public opinion.

8. References

- [1] China Internet Network Information Center (CNNIC). The 47th Statistical Report on China's Internet Development Status [R]. 2020 April 28.
- [2] Ding Zhaoyun, Jia Yan, Zhou Bin. A Review of Microblog Data Mining Research [J]. Computer Research and Development, 2014, 51(04):691-706.
- [3] Jennifer P. Sharing and querying content through mobile phones and social participation[J]. Middle East Journal of Scientific Research, 2013, 16(12): 1923-1925.
- [4] Bojar U, Breslin, J G, Peristeras, V, et al. Interlinking the social web with semantics[J]. Intelligent System. 2008, 23(4): 3-3.
- [5] Jiang Hong. A Review of Sentiment Analysis Research [J]. Intelligent Computer and Application, 2018, 8(05): 103-105.
- [6] Li Jianhua, Liu Gongshen, Lin Xiang. A Review of Emotional Tendency Analysis and Application Research [J]. Journal of Information Security, 2017, 2(02): 48-62.
- [7] Xu L. Constructing the Affective Lexicon Ontology[J]. Journal of the China Society for Scientific and Technical Information, 2008, 2(27): 180-185.
- [8] Zhang Yang, Liu Xiaoxia, Sun Kailong, Li Hongan. Research on text orientation recognition based on emotional description items [J]. Computer Engineering and Application, 2015, 51(04): 158-161.
- [9] Liang Xiaomin, Xu Jian. Research on Sentiment Analysis and Relationship Network of Comment Objects in Public Opinion Events [J]. Information Science, 2018, 36(02): 37-42.
- [10] Li Tianchen, Yin Jianping. Method for Discriminating Sentiment Polarity Based on Topic Clustering [J]. Computer Science and Exploration, 2016, 10(07): 989-994.
- [11] Li Yongyong, Zhou Xueguang, Sun Yan, Zhang Huanguo. Research and Implementation of Chinese Microblog Sentiment Analysis [J]. Journal of Software, 2017, 28(12): 3183-3205.
- [12] Yin Guanghua, Liu Xiaoming, Zhang Lu, etc. Analysis and Research on Sentiment Elements of Short Texts Based on LSTM Feature Templates [J]. Computer Research and Development, 2020, 31(11): 42-45.
- [13] You Dandan, Chen Fuji. A Review of Research on Internet Public Opinion Prediction in my country [J]. Information Science, 2016, 34(12): 156-160.
- [14] Shi Rui, Chen Fuji, Zhang Jinhua. Research on Network Public Opinion Prediction Based on Combined Gray Model [J]. Journal of Intelligence, 2018, 037(007): 101-106.
- [15] Luo Taiye. Research on the Development and Prediction of Microblog Public Opinion Hotspots Based on the Logistic Model [J]. Statistics and Information Forum, 2017, 32(10): 91-95.

- [16] Zheng Buqing, Zou Hongxia, Hu Xinjie. Research on Internet Public Opinion Prediction Based on Inflection Point [J]. Computer Science, 2018, 45(S2): 539-541 575.
- [17] Xu Minmin. Research on Microblog Public Opinion Prediction Model Based on Exponential Smoothing Method [J]. China Public Security (Academic Edition), 2016(01): 80-84.
- [18] Wang Yamin, Ning Jing, Ma Xubu. Research on Prediction of Public Opinion in Public Policy Based on Social Media[J]. Information Theory and Practice, 2019, 42(01): 87-93.
- [19] Huang Yaju, Chen Fuji, You Dandan. Research on Network Public Opinion Prediction Based on Hybrid Algorithm and BP Neural Network [J]. Information Science, 2018, 36(02): 24-29.
- [20] Huang Yaju, Chen Fuji. Research on Combination Prediction of Internet Public Opinion Based on Entropy Method[J]. Information Science, 2018, 319(03): 70-74.
- [21] Wang Ning, Zhao Shengyang, Shan Xiaohong. Research on Network Public Opinion Prediction and Grading Method Based on Gray System Theory [J]. Information Theory and Practice, 2019, 42(02): 124-130.
- [22] Sun Jingchao, Zhou Rui, Li Peiyue, Lu Tianliang. Research on Network Public Opinion Trend Prediction Based on Recurrent Neural Network [J]. Information Science, 2018, 36(08): 118-122.
- [23] Yuan Haixia. Sentiment Dictionary Construction Method Based on Word2Vec and HowNet [J]. Modern Computer (Professional Edition), 2018, 04(02): 9-13.
- [24] Wang Lidong, Zhang Huixi. Research on Microblog Text Semantic Retrieval Based on HowNet [J]. Information Science, 2016, 34(09): 134-137.
- [25] Xu Linhong, Lin Hongfei, Yang Zhihao. Text Tendency Recognition Mechanism Based on Semantic Understanding[J]. Chinese Journal of Information, 2007, 21(01): 96-100.
- [26] Chen Zhixiong, Wang Shihui, Gao Rong. Microblog Opinion Leader Identification Model Based on Sentiment Analysis [J]. Computer Science, 2018, 45(05): 168-175.

Research Methods and Professional Development 2203588 dongbo

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

Instructor

57 /100

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

GRADING FORM: ACFA7004_SEMB_2023

BO DONG

57

ENGAGEMENT WITH LITERATURE

(20 marks)



Fair appraisal of networked public opinion supported by credible references

13

KNOWLEDGE AND UNDERSTANDING

(20 marks)



Trend forecast of perceived public opinions introduced and explained sufficiently well.

14

COGNITIVE/ INTELLECTUAL SKILL

(30 marks)



Overall research methodology is not made explicit.

11

PRACTICAL/ APPLICATION SKILLS

(10 marks)



Little implementation and project managerial elements shown in proposal.

9

TRANSFERRABLE SFL/ PP

(20 marks)



Fairly well organised document but does not include all core elements of proposal as indicated in assignment brief.

10