Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 3, June 2025, pp. 2107~2119
ISSN: 2302-9285, DOI: 10.11591/eei.v14i3.8963

Multiword target-independent transformer-based model for financial sentiment analysis in colloquial Cantonese

Carlin Chun Fai Chu[1], Raymond So[2], Ernest Kan Lam Kwong[2], Andy Chan[3]
[1]Department of Computer Science, The Hang Seng University of Hong Kong, Hong Kong, China
[2]Big Data Intelligence Centre, The Hang Seng University of Hong Kong, Hong Kong, China
[3]IDDY Ltd, Hong Kong, China

Article Info

Article history:

Received Jul 5, 2024
Revised Nov 22, 2024
Accepted Dec 25, 2024

Keywords:

Aspect-based opinion mining
Natural language inference
Pretrained language model
Sentinel token
Transformer

ABSTRACT

Tokenization process decomposes a multi-word-span instrument name into several tokens and the transformer attention mechanism handles each token individually, thus hindering the treatment of the related tokens as a single entity. The existence of multiple instruments in a single message further exaggerates the complications and results in low predictive performance. This study proposed the use of sequentially tagged target-independent sentinel tokens to encapsulate multiword instrument aspects for natural language inference model fine-tuning. The encapsulation not only facilitated the attention mechanism to handle an instrument name as a single entity but also enabled the model to handle unseen instruments effectively. Our empirical analysis was based on 5,178 manually annotated instrument–sentiment pairs originated from finance discussion board messages that addressed sentiments of one to four instruments in a single post. The proposed approach consistently outperformed the direct bidirectional encoder representations from transformers (BERT) based approach in terms of recall, precision, and F1-score when handling financial commentaries written in colloquial Cantonese. This study demonstrated the potential benefits of target-independent sentinel token encapsulation for natural language inference. The underlying logic of multiword target-independent encapsulation was expected to hold for other languages, including Chinese, Japanese, and Thai.

Corresponding Author:

Carlin Chun Fai Chu
Department of Computer Science, The Hang Seng University of Hong Kong
Hang Shin Link, Siu Lek Yuen, Shatin, N.T., Hong Kong, China
Email: carlinchu@hsu.edu.hk

1. INTRODUCTION

Changes in asset prices are strongly associated with the corresponding market opinions, especially in the case of financial instruments that are heavily traded by individual investors [1]-[3]. It is beneficial to consolidate information extracted from text-based financial materials, in addition to numerical quotes and ratios, to capture the overall market condition and make a well-informed investment decision. A better understanding of text-based market opinions can be obtained by exploring the potential of big data analytics on company reports [4], professional commentaries [5], news articles [6], [7], blogs [8], microblogs [3], [9], and discussions among individual investors on social media [1]. Extracting the opinions of investors from posts on social media is a challenging task as the relevant text invariably involves the use of colloquial language, which cannot be effectively handled by typical off-the-shelf machine learning models [10]-[14]. Moreover, the relevant expressions sometimes address more than one instrument [15] and/or contain different sentiments in a single sentence [16]. Natural language inference (NLI) aspect-based opinion mining (ABOM) [12], [17], [18] has been demonstrated to be a workable approach to extract fine-grained sentiments from social media messages. The 'subject of interest' is treated as 'aspect' and its relationships with particular sentiment directions are testified via the use of premise-hypothesis structure. For messages written in English, individual aspect is usually denoted by a single word which, in turn, is subsequently represented as a single token under WordPiece or byte-pair encoding (BPE) tokenization process. As an aspect is represented by a single token, the transformer self-attention mechanism can pinpoint the particular token and model its relationships effectively. However, when an aspect is tokenized as multiple tokens, the model performance may be hampered due to the complication involved in the weight aggregation for the multiple-token aspect. There is not any built-in mechanism to ensure that the multi-tokens are conceptually encapsulated as a single aspect in the bidirectional encoder representations from transformers (BERT) model [19], [20]. Sophisticated nonlinear functions are required to aggregate the content attentions from multiple tokens [21], [22]. Furthermore, a single discussion post may cover sentiments on several instruments. The presence of multiple instruments leads to the aggregation of noise, thus further deteriorating the performance resulted from transformer-based language models. In addition, adaptation of newly released financial instruments poses another challenge to the model. Typical ABOM approach operates on a set of pre-defined aspects and a direct conversion of instruments as aspects limits its extensibility to newly released financial instruments, which are unknown during the model fine-tuning stage.

This study investigated an alternative approach to extract financial sentiments on multiple multiword instruments from code-mixed colloquial English–Chinese language (Cantonese) discussion board messages. Apart from developing a sophisticated nonlinear function for attention weight aggregation, this study extended the sentinel token concept to solve the problem. The sentinel token concept was discussed and proposed to act as a replacement for a consecutive span of word tokens that can facilitate the formulation of various natural language processing (NLP) tasks [23]. This study proposed using sentinel token as a neutral, target-independent placeholder to represent both known and unknown instruments. Our approach models sentinel tokens as a limited number of sequentially labelled new vocabularies. The new numbered vocabularies are used to sequentially encapsulate multiword instruments in social media messages. The sentinel token is target-independent because it serves as a numbered placeholder indicating a sequence of occurrences, instead of the actual underlying instrument. The independence property also facilitates handling of newly released financial instruments as aspects in ABOM analysis where their names are not covered in training data.

To the best of our knowledge, this study is the first empirical analysis of the use of a BERT-based NLI model in finance that focuses on code-mixed colloquial English–Chinese (Cantonese). The major contributions of this research are as follows:

– A simple effective multiword encapsulation NLI approach to extract financial sentiments from multiple unseen instruments in a single social media message is proposed. Our target-independent sentinel token method is designed to handle multiword instruments that may be seen or unseen in the training corpus.
– We compare the direct application of the BERT-based model with the proposed approach in situations involving variations in the number of instruments mentioned in a single discussion post message. The empirical results reflect the strengths and weaknesses of the two approaches, which in turn indicates their virtues for the analysis of financial sentiments in code-mixed colloquial language.
– A manually labelled English–Chinese code-mixed (Cantonese) dataset of financial discussion board messages containing sentiment labels of one to four multiword span instruments is constructed for research purpose. Cantonese is a low-resource language, and the availability of a labelled dataset can thus facilitate future research in related fields.

The remainder of this paper is structured as follows: the literature review is provided in section 2 and the proposed method is detailed in section 3. Section 4 reports the empirical results and highlights the pros and cons of the proposed approach. Section 5 summarizes the findings of this study.

2. RELATED WORKS

Financial sentiment extraction has been a popular topic in machine learning discipline. Early attempts relied on dictionary-based approach to extract useful features to model the sentiment polarities. For example, using bag of words features with machine learning techniques, including naive bayes model, support vector machine and multilayer perceptron model, to extract Brazilian stock sentiments from Portuguese twitter messages [1]; using hand crafted lexical features, semantic features and their combinations to extract sentiments on English financial microblogs and news [3]. In addition, some studies utilized the extracted sentiment information for subsequent financial applications such as price trend prediction and construction of trading strategies. A study extracted overall investors' sentiment from a large amount of press releases by large financial news companies via aggregation of the occurrence of words appeared in a predefined vocabulary list for stock price trend prediction [15]. Other studies employed dictionary-based approach to extract financial sentiments from English news articles [6] and Germany twitter messages [2] for the construction of trading strategies respectively.

Apart from dictionary-based approach, the BERT family of models [20], [24] demonstrated to be an effective approach due to its simplicity and impressive predictive performance [7]. ABOM has been shown to be a practical approach for sentiment extraction [12], [17], [18]. Pairs of premises and hypotheses are used to extract the sentiment direction of an instrument. An instrument name is treated as an aspect and the sentiment direction is determined via an NLI process [9], [25]. BERT-based models showed to be effective to extract financial sentiment directions on English text [9], [25], [26], and its adaptability in handling informal languages and slang in English was demonstrated [11], [27]. Studies on Indonesian [28] and Spanish [26] sentiment analysis also supported the competence of BERT models.

Tokenization is a key component associated with BERT models. A sequence of input text is processed into tokens before it is put into a model. Asian languages such as Chinese, Japanese, and Thai, make use of a multiword term to represent a single financial instrument aspect. In contrast to a financial instrument written in English which is usually denoted by a single word, a Chinese instrument name is usually denoted by a longer, multiword span term. When an instrument is represented by a single token, the transformer self-attention mechanism can pinpoint the particular token and model its relationships effectively. However, when a multiword aspect is tokenized as multiple tokens, the model performance may be hampered due to the complication involved in the weight aggregation for the multiple-token aspect. There is not any built-in mechanism to ensure that the multi-tokens are conceptually encapsulated as a single aspect in BERT model [19], [20]. It is speculated that the performance may deteriorate with the increase of number of tokens involved in an aspect in typical BERT-based model. To better aggregate the content attentions from multiple tokens, sophisticated nonlinear functions were suggested [21], [22].

On the other hand, a single financial discussion post may cover sentiments on several instruments at a time. The presence of multiple instruments leads to the aggregation of noise, thus further deteriorating the performance. BERT models were employed on Chinese financial sentiment analysis. Studies on finance commentary competition datasets [8], financial post titles [29], short investor reviews [5] and individual company specific news headlines [30] showed satisfactory sentiment extraction performance. However, these studies worked on short input text sequences and did not emphasize cases involving multiple instruments. In addition, adaptation of newly released financial instruments poses another challenge to the model. Typical ABOM approach operates on a set of pre-defined aspects and direct conversion of instruments as aspects limits its extensibility to newly released financial instruments, which are unknown during the fine-tuning stage of the model.

Other than deriving a complicated function to handle instruments expressed in multiple tokens, this study proposed an alternative approach to solve attention weight aggregation problem. Literature in neural machine translation area demonstrated that grouping of multiword expressions as a single unit [23], [30] can improve performance on NLP tasks. This study extends this concept by constructing a sentinel token as a target independent sequentially tagged placeholder for encapsulating a financial instrument which is expressed as multiple tokens. Our approach models sentinel tokens as a limited number of sequentially labelled new vocabularies which are used to sequentially encapsulate multiword instruments in a long input text sequence. The sentinel token is target independent because it serves as a placeholder indicating a sequence of occurrences, instead of the actual underlying instrument. The independence property also facilitates handling of newly released financial instruments as aspects in ABOM analysis as their names are not covered in training data.

This paper proposes an effective approach to encapsulate multiword instrument aspects for sentiment extraction from multiple seen/unseen instruments mentioned in a single social media message. The overall system architecture, data modelling, and sentiment classification model construction are detailed in the subsequent sections.

3. METHOD

The sections below highlight the key methodological components of this research project. To the best of our knowledge, no off-the-shelf pre-trained aspect-based BERT model is available in the market for handling code-mixed colloquial English–Chinese (Cantonese) in finance domain. We first provide the overall design of the aspect-based system of sentiment analysis and then address the data structure for NLI and the models of sentiment classification.

3.1. Overall system design

The overall design of our aspect-based system of sentiment classification is inspired by several past studies [16], [31]. Three main stages including data preparation, NLI modelling and sentiment prediction, together with the key processes are presented in Figure 1. The system architecture is an extension of the previous studies carried out by the research team for ABOM analysis [10], [32]. The key innovation of this study is about the proposal of using sentinel tokens as a method for multi-token encapsulation in the NLI modelling stage.

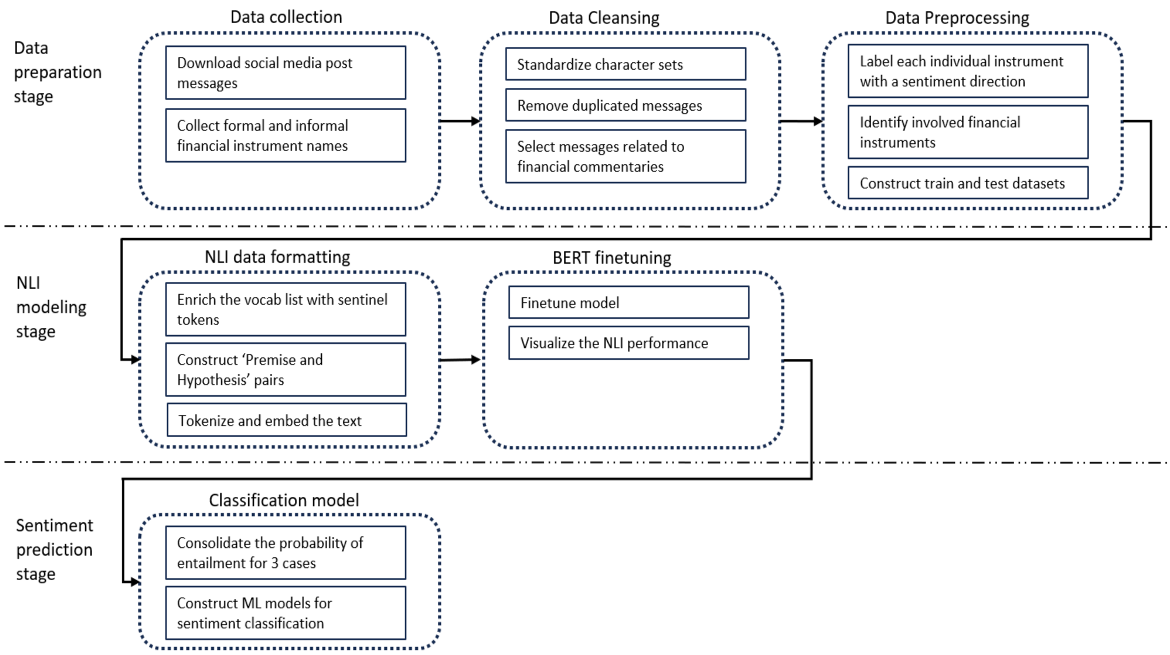

Figure 1. Design of the ABOM system

A list of formal names of instruments was obtained from the official website of the Hong Kong Exchanges and Clearing (HKEX) Ltd., and the informal instrument names were provided by our research partner company iDDY.AI. Character set standardization, removal of duplicated messages and data selection is performed in the data preparing stage as well. In the NLI modelling stage, the input texts were tokenized and appropriately embedded before being used to fine-tune the BERT model. The sentinel tokens proposed in this study were implemented in the NLI stage and the detailed procedures were listed in the subsequent subsections. The hugging face Chinese BERT pre-trained model with a next-sentence prediction (NSP)-based classification head was implemented to model the entailment probabilities of three sentiments, including neutral, bullish and bearish directions respectively. The three probability values are consolidated in the final stage using a rule-based model and different machine learning models to form a multi-class classifier.

3.2. Data acquisition and annotation

Our research team collected 102,786 social media messages from several discussion forums that are popular in Hong Kong (e.g., dicuss.com.hk, investing.com, and likhg.com). The pre-processing task was designed to standardize the given text and remove irrelevant data. The input characters were first standardized into traditional Chinese and lower-case English text, and out-of-scope messages (i.e., messages that did not contain any formal or informal instrument names) were filtered out. 12,521 messages containing financial commentaries on instruments traded in the Hong Kong Stock Exchange were obtained. Messages addressing with one to four investment sentiments were selected for manual annotation. Each message was annotated by three trained annotators and monitored by a senior researcher. The annotators were required to identify all the financial instruments (aspects) together with the directions of their corresponding sentiments: bullish, bearish, or neutral. The senior researcher made the final decision on annotation in case of a conflict among the annotators. The research team managed to annotate 5,178 distinct instrument–sentiment pairs that covered bullish, bearish, and natural sentiments. The distribution of the annotated data is tabulated in Table 1. The research data associated with this study is available on request.

Table 1. Distribution of annotated instrument–sentiment pairs

Num of instruments in a discussion message	Neutral	Bullish	Bearish	Total
1	334	334	344	1002
2	667	667	667	2001
3	683	424	408	1515
4	293	234	133	660

3.3. Data structure for natural language inference processing

This study proposes encapsulating the names of multiword instruments by using target-independent sentinel tokens for BERT-based NLI modelling. The idea of consolidating a consecutive span of tokens was developed as a framework to model any problem in a text-based language as a problem in the text-to-text format [23]. The original paper mentioned that the sentinel token can act as a replacement for a consecutive span of word tokens to facilitate the formulation of various NLP tasks. In addition, research in machine learning translation task also demonstrated that the use of multiword aware modelling approach helps to improve the modelling performance [33].

This study further extends the use of a sentinel token to act as a target-independent neutral sentiment placeholder representing both seen and unseen instruments. It is proposed to model the set of sentinel tokens as a limited number of sequentially labelled new vocabularies. These numbered new vocabularies are used to sequentially encapsulate the multiword instruments in a social media message, and to replace the occurrences of repeated instruments with the corresponding identical vocabularies. In other words, if an instrument occurs in a message more than once, it will be modelled by the same numbered sentinel token, the number of which represents the sequence of its first occurrence in the message. The sentinel token is "target independent" as it serves as a sequential placeholder indicating the sequence of occurrence instead of the actual underlying instrument. Their virtue of target independence enables the sentinel tokens to effectively handle newly released financial instruments (with unseen names). Moreover, as new vocabularies are used in the process of tokenization, the sentinel tokens are sentiment neutral. They can help conceal the potential positivity bias observed in the names of multiword Chinese instruments that affects the subsequent NLI process. The steps for constructing the NLI premise and hypothesis pair can be summarized as follows:

- Aspect identification (identify instrument names)
- Substitution of sentinel tokens (for the sentinel token-based approach)
- Construction of the NLI premise and hypothesis pair

The main purpose of NLI is to determine whether the "premise" and "hypothesis" are logically linked. The "premise" is formulated as a concatenation of post tile and message with the separator symbol "。," while the "hypothesis" is formulated to present the potential direction of the market sentiment. This study wants to handle three possible sentiment directions (i.e., bullish, bearish, and neutral). The RTE4 method of labelling (i.e., using the entailment/contradiction/unknown labels) is not suitable here as the unknown label cannot be logically mapped to a single market direction. Therefore, we labelled our experimental data with the terms "entailment" and "non-entailment" instead.

For example, the following message mentioning that the sentiment of the first instrument as neutral and the second instrument as bearish. The corresponding English translations are enclosed in brackets.

> 微盟集團廢股次次都大回。恆生指數最新跌左二百幾點，跌破二萬五。沽期啦，連支持位都穿埋。
>
> (The scrapped shares of Weimob Group comes back again. The latest Hang Seng Index falls by more than 200 points, falling below 25,000. Sell the options, falling below the support level.)

The corresponding hypothesis-premise pairs for the traditional approach and the proposed sentinel token approach are tabulated in Table 2. Note that the instrument names are replaced by sentiment tokens in the sentinel token-based approach. Each instrument (aspect) is checked for the three sentiment directions in the process of NLI modelling, with the label of "entailment" as representative of its true direction. As NLI involves modelling the probability of entailment of a given input "premise and hypothesis" pair, three input pairs that cover the possibilities of bullish, bearish, and neutral directions are required to predict the direction of the sentiment of a single instrument. In other words, three related probability values are output as the result of prediction of the sentiment of a single instrument.

Table 2. Example of premise, hypothesis, relationship and aspect (English translations are enclosed in brackets)

Method	Premise	Hypothesis	Relationship	Aspect
Traditional	微盟集團廢股次次都大回。恆生指數最新跌左二百幾點，跌破二萬五。沽期啦，連支持位都穿埋。 (The scrapped shares of Weimob Group comes back again. The latest Hang Seng Index falls by more than 200 points, falling below 25,000. Sell the options, falling below the support level.)	微盟集團是中 (Weimob Group is neutral)	Entailment	微盟集團 (Weimob Group)
		微盟集團是正 (Weimob Group is positive)	Non-entailment	
		微盟集團是負 (Weimob Group is negative)	Non-entailment	
		恆生指數是中 (Hang Seng Index is neutral)	Non-entailment	恆生指數 (Hang Seng Index)
		恆生指數是正 (Hang Seng Index is positive)	Non-entailment	
		恆生指數是負 (Hang Seng Index is negative)	Entailment	
Sentinel token	<***投資工具1***>廢股次次都大回。<***投資工具2***>最新跌左二百幾點，跌破二萬五。沽期啦，連支持位都穿埋。 (The scrapped shares of <token1> comes back again. The latest <token2> falls by more than 200 points, falling below 25,000. Sell the options, falling below the support level.)	<***投資工具1***>是中 (<token1> is neutral)	Entailment	<***投資工具1***> (<token2>)
		<***投資工具1***>是正 (<token1> is positive)	Non-entailment	
		<***投資工具1***>是負 (<token1> is negative)	Non-entailment	
		<***投資工具2***>是中 (<token2> is neutral)	Non-entailment	<***投資工具2***> (<token2>)
		<***投資工具2***>是正 (<token2> is positive)	Non-entailment	
		<***投資工具2***>是負 (<token2> is negative)	Entailment	

3.4. Bidirectional encoder representations from transformers model fine tuning

The hugging face BERT pre-trained model for Chinese text with 12 layers, 768 hidden units per layer, 12 attention heads, and 110 million parameters was used along with a NSP-based classification head to model the polarities of the sentiments. Models with and without the use of sentinel token encapsulation were finetuned for 20 epochs separately using the same set of training data. The configurations with the highest in-sample F1 value were selected for out-of-sample performance comparisons. The predicted entailment probabilities of three sentiments are consolidated in the final stage. The use of bounded-interval label space for binary classification [34] is extended and adopted to multi-class classification in this study. Probabilities of entailment of neutral, bullish and bearish sentiments are aggregated by a rule-based model and several machine learning models to form a multi-class classifier.

3.5. Models of sentiment classification

Using the classification modelling approach with the consideration of bounded-interval label space [34], the predicted probabilities of three sentiment labels can be systemically consolidated. For example, to model a sentiment class of a given premise, 3 corresponding hypotheses are involved, and a machine learning model is employed to consolidate the probabilities. Figure 2 depicts the relationship.

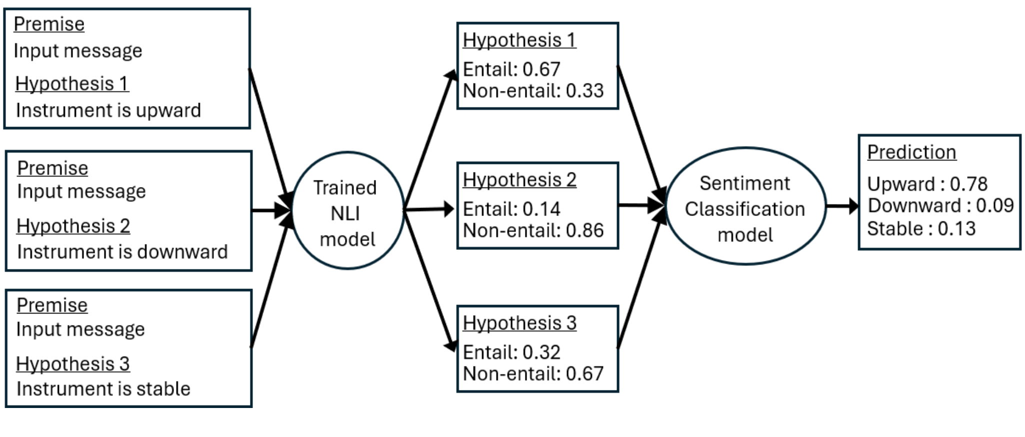

Figure 2. NLI based sentiment classification model

Several sentiment classification models including simple rule-based, logistic regression, extreme gradient boosting (XGBoost), and the support vector machines (SVM) are used to investigate the performance of our proposed method. The simple rule-based model is defined as:
- If P(entailment of bullish hypothesis) ≥0.5 and P(entailment of bearish hypothesis) <0.5, then class=bullish
- If P(entailment of bearish hypothesis) ≥0.5 and P(entailment of bullish hypothesis) <0.5, then class=bearish
- Else, class=neutral

Furthermore, the logistic regression, XGBoost and SVM models are implemented using python scikit-learn and xgboost package. The configuration of the multi-class models are listed below:
- Logistic regression: L2 penalty, regularization parameter C=1
- XGBoost: L2 regularization parameter lambda=1, shrinkage parameter eta=0.3, gamma=0, max_depth=9
- SVM: radial basis function kernel, kernel coefficient=scale, regularization parameter C=1

3.6. Evaluation methods

First, it is expected that the proposed sentinel token encapsulation approach can provide benefits for BERT-based NLI process. For multiclass sentiment extraction, a good NLI model should provide probability predictions on distinct sentiments with a strong discriminating property. Given a set of labelled data on three sentiment polarities, a good model should produce predicted probabilities with close clusters in a 3D probability plot, meaning that the sentiments are discriminated effectively. The resultant probabilities of fine-tuned model will be presented in the experiment result section.

Secondly, the influence on predicted performance is evaluated via recall, precision and F1 on the three sentiments including neutral, bullish and bearish directions respectively. Performance on handling posts with one to four instruments are investigated separately to explore the potential trends. And a unified model which is trained by data with one to four instruments altogether is investigated as well. Prediction performance on unseen multiword instruments (aspects) is investigated. The distribution of the training and testing data is tabulated in Tables 3 and 4.

Table 3. Distribution of training and testing data for 1-4 instruments

Num of instruments in a discussion message	Training (sample per class)	Testing (sample per class)	Total number of samples
1	234	100	1,002
2	467	200	2,001
3	700	300	3,000
4	934	400	4,002

Table 4. Distribution of training and testing data for a unified model

Num of instruments in a discussion message	Instrument 1 (training/testing sample per class)	Instrument 2 (training/testing sample per class)	Instrument 3 (training/testing sample per class)	Instrument 4 (training/testing sample per class)
1	58/25			
2	58/25	78/33		
3	58/25	78/33	117/50	
4	58/25	78/33	117/50	233/100
Total number of samples	996	999	1002	999

The size of the training and testing data is around 70:30. The size of training and testing data increase with the number of instruments involved in a discussion post. Due to the limited labelled data for posts with 3 and 4 instruments, data augmentation is carried out to construct additional data by merging posts with fewer instruments randomly. For unified model, data is prepared to ensure that the total number of samples per sequentially labelled instrument is almost equal while maintaining proportionally more samples for posts with a greater number of involved instruments. The sampling method for training the unified model aims to avoid the potential basis on having a greater occurrence frequency of earlier labelled sentinel tokens. In other words, the procedure helps to avoid the basis of having a better performance on the shorter posts than longer posts. The number of samples with neutral, bullish and bearish sentiments are in the ratio of 1:1:1.

4. RESULTS AND DISCUSSION

4.1. Performance of predicted probabilities from finetuned models

The empirical study aims to investigate the potential benefits contributed by using sentinel tokens for encapsulation of multiword aspect terms. Four BERT models were finetuned for handling 1 to 4 instruments individually and one unified model was finetuned for handling 1 to 4 instruments collectively. All models exhibited a similar outcome and the result from the unified model was reported as the illustration. Figure 3 covers six scatterplots to illustrate the resultant predicted probabilities from the unified model with and without the use of sentinel tokens. The entailment probabilities estimated by the sentinel approach were more concentrated in the vertices whereas the probabilities estimated by the direct BERT approach were more scattered. Under the same limited number of fine-tuning epochs (i.e., 20 epochs), sentinel token encapsulation approach gave predicted probability values with more extreme numbers, implying that the corresponding model was more confident about the predicted outcomes. Based on the experiment result, the entailment probability values of neutral sentiment cases were much less concentrated in the direct BERT approach (model without the use of sentinel tokens), indicating a relatively higher uncertainty about the predicted outcomes. Subsequent subsection examined its influence on the multi-class sentiment classification models.

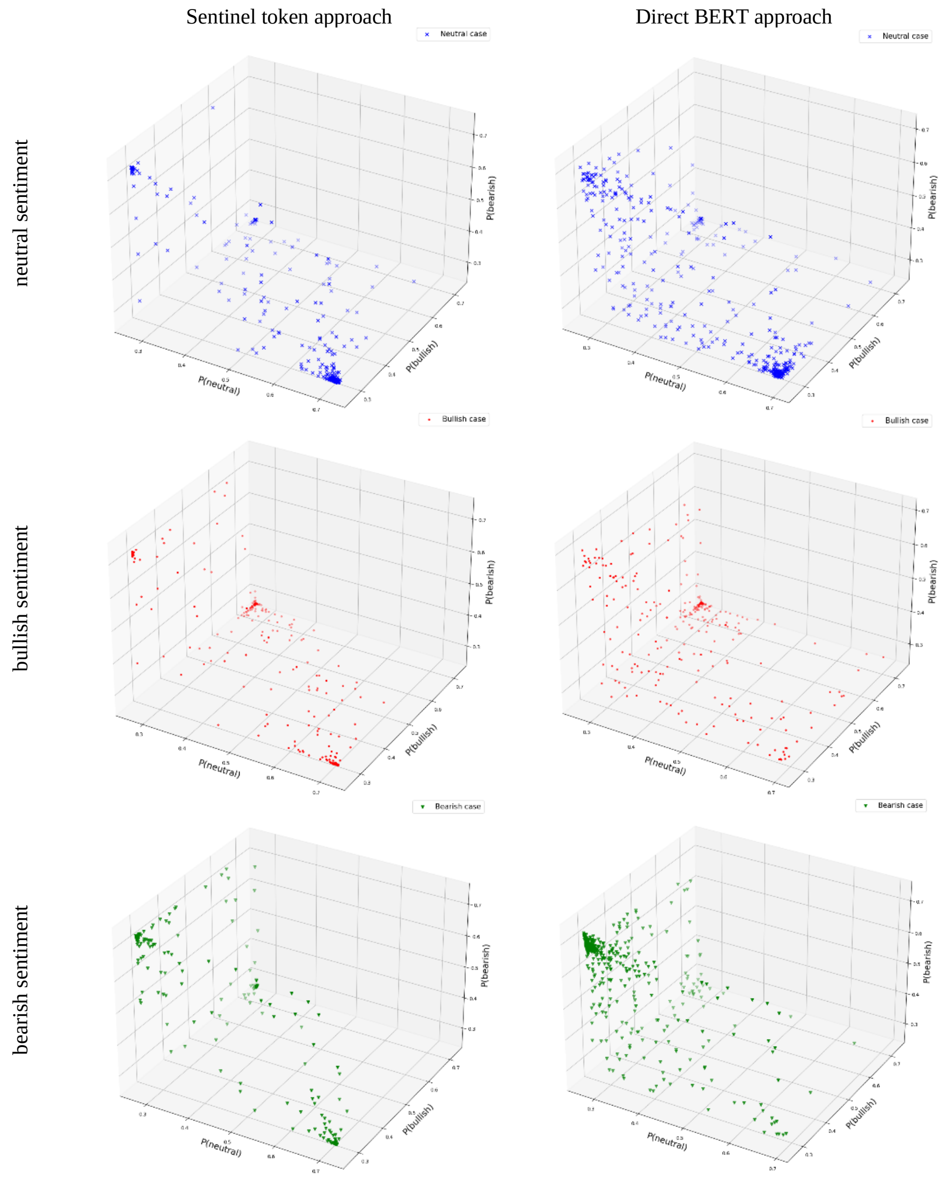

Figure 3. Scatterplot of probabilities of three market sentiments (left: sentinel token approach, right: direct BERT)

4.2. Prediction performance on unseen instrument aspects

The empirical study covered the use of rule-based, logistic regression, XGBoost, and SVM models to assess the performance of the proposed approach. The F1-scores of different methods of sentiment classification on messages containing varying numbers of instruments were given in Table 5. The sentinel token-based approach yielded an improvement of 12.74%–25.73% over the direct BERT approach when handling one to three instrument names. However, it provided an improvement of only 1.18%–13.20% on messages containing four instrument names. It was expected that the lexical and structural complexity of a four-aspect post rendered the underlying language model ineffective, and thus reduced the benefit derived from the sentinel token encapsulation. For the unified model which was capable to handle with one to four instruments, the rates of improvement in F1-score ranged from 8.68% to 11.40%.

Table 5. F1-scores of predictions by different models of sentiment classification on out-of-sample data

Num of instruments	Approach	Basic threshold	Logistic regression	XGBoost	SVM (RBF)
1	BERT	0.611	0.636	0.61	0.626
	Sentinel	0.728	0.717	0.675	0.722
	% increase	19.15%	12.74%	10.66%	15.34%
2	BERT	0.478	0.512	0.502	0.509
	Sentinel	0.601	0.597	0.586	0.603
	% increase	25.73%	16.60%	16.73%	18.47%
3	BERT	0.548	0.559	0.549	0.558
	Sentinel	0.675	0.660	0.663	0.663
	% increase	23.18%	18.07%	20.77%	18.82%
4	BERT	0.553	0.594	0.592	0.592
	Sentinel	0.626	0.626	0.599	0.631
	% increase	13.20%	5.39%	1.18%	6.59%
Unified	BERT	0.614	0.630	0.618	0.634
	Sentinel	0.684	0.688	0.686	0.689
	% increase	11.40%	9.21%	11.00%	8.68%

In-depth results of the unified model was tabulated in Table 6. The sentinel token-based approach consistently improved the predictive performance in terms of the recall, precision, and F1-score. The SVM with the radial basis function yielded the best performance, with a recall rate of 0.689, precision of 0.688, and F1-score of 0.689. For the cases of handling 1 to 4 instruments individually, the results exhibited a similar pattern, but with different rates of improvements.

Table 6. Performance of the unified model in terms of sentiment classification on a dataset of unseen messages and instrument names

	Approach	Basic threshold	Logistic regression	XGBoost	SVM (RBF)
F1	BERT	0.614	0.630	0.618	0.634
	Sentinel	0.684	0.688	0.686	0.689
	% increase	***11.40%***	***9.21%***	***11.00%***	***8.68%***
Recall	*BERT*	*0.613*	*0.632*	*0.619*	*0.635*
	Sentinel	*0.684*	*0.688*	*0.686*	*0.689*
	% increase	***11.58%***	***8.86%***	***10.82%***	***8.50%***
Precision	*BERT*	*0.614*	*0.630*	*0.618*	*0.634*
	Sentinel	*0.685*	*0.688*	*0.686*	*0.688*
	% increase	***11.56%***	***9.21%***	***11.00%***	***8.68%***

The previous subsection mentioned that sentinel token encapsulation helps to form closely packed clusters when handling posts with neutral sentiments. It is interesting to know whether the improved probability estimates give different influences for respective sentiment classification predictions. Table 7 shows that not only the prediction of neutral sentiment was improved, the classifications of bullish and bearish sentiments were also involved. The F1-scores of bullish and bearish sentiments improved, except for a slight decrease of 2.13% when using XGBoost to handle cases involving bearish sentiments. The bold values highlighted the highest rates among all types of classification models. The sentinel token-based approach outperformed the direct BERT-based approach in most scenarios. The improved handling of bullish and bearish cases can be attributed to the more closely packed probability values of neutral sentiment in the probability vectors of aspects. This indirectly helped develop a better model for determining the direction of the sentiment of an individual aspect.

Table 7. Performance in terms of the classification of neutral, bullish, and bearish sentiments (unified model)

	Coverage	Approach	Basic threshold	Logistic regression	XGBoost	SVM (RBF)
F1	Neutral	BERT	0.461	0.526	0.524	0.530
		Sentinel	0.573	0.570	0.546	0.573
		% increase	***24.30%***	***8.37%***	***4.20%***	***8.11%***
	Bullish	BERT	0.593	0.603	0.596	0.590
		Sentinel	0.649	0.647	0.608	0.657
		% increase	***9.44%***	***7.30%***	***2.01%***	***11.36%***
	Bearish	BERT	0.606	0.652	0.657	0.656
		Sentinel	0.655	0.662	0.643	0.663
		% increase	***8.09%***	***1.53%***	***−2.13%***	***1.07%***
Recall	Neutral	BERT	0.468	0.485	0.500	0.517
		Sentinel	0.64	0.583	0.568	0.565
		% increase	36.75%	20.21%	13.60%	9.28%
	Bullish	BERT	0.635	0.637	0.608	0.588
		Sentinel	0.608	0.618	0.570	0.642
		% increase	−4.25%	−2.98%	−6.25%	9.18%
	Bearish	BERT	0.555	0.665	0.675	0.675
		Sentinel	0.620	0.677	0.657	0.688
		% increase	11.71%	1.80%	−2.67%	1.93%
Precision	Neutral	BERT	0.455	0.574	0.551	0.543
		Sentinel	0.518	0.557	0.525	0.581
		% increase	13.85%	−2.96%	−4.72%	7.00%
	Bullish	BERT	0.557	0.572	0.586	0.593
		Sentinel	0.696	0.680	0.651	0.673
		% increase	24.96%	18.88%	11.09%	13.49%
	Bearish	BERT	0.667	0.639	0.640	0.638
		Sentinel	0.695	0.647	0.629	0.641
		% increase	4.20%	1.25%	−1.72%	0.47%

5. CONCLUSION

This paper proposed an efficient way for ABOM to extract sentiments of multi-word-span financial instruments from discussion board messages written in a colloquial English-Chinese code-mixed language (Cantonese). The proposed approach improved on the original sentinel token concept by treating it as a sequentially tagged target-independent placeholder to encapsulate multi-word-span instrument names appeared in the discussion board posts. The encapsulation aimed to facilitate the transformer attention mechanism to handle a multi-word-span instrument aspect as a single entity and enable the NLI model to handle unseen aspects effectively. The empirical analysis was based on 5,178 manually annotated instrument–sentiment pairs originated from finance discussion board messages that addressed sentiments of one to four instruments in a single post. The proposed approach consistently outperformed in terms of recall, precision, and F1-score. Similar as machine learning translation tasks, the use of multi-token aware modelling approach demonstrated to improve the modelling performance. Although the empirical analysis was based on messages in colloquial Cantonese, the logic of encapsulation of a multiword term to reduce noise should hold for other languages (e.g., Chinese, Japanese, and Thai) in which a single instrument name is tokenized into multiple tokens. Apart from focusing on multi-word-span instruments, it is possible to extend the encapsulation process to cover common multi-word-span colloquial phrases. However, it will take some effort to construct a tailored lexicon accordingly. It is expected that the overall performance will be further improved. The findings of this study shed light on the benefits of sentinel token-based encapsulation for BERT-based NLI modelling.

FUNDING INFORMATION

This work was supported by the Faculty Development Grant of the Hang Seng University of Hong Kong (SDSC-SRG010), and the Research Matching Grant Scheme of the University Grants Committee of the HKSAR Government (RMGS700029).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Carlin Chun Fai Chu	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Raymond So	✓	✓	✓	✓	✓	✓			✓		✓	✓	✓	
Ernest Kan Lam Kwong	✓		✓			✓		✓	✓					
Andy Chan	✓									✓				✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

No human subjects are involved in the empirical experiment

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, Carlin Chun Fai Chu. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] A. E. O. Carosia, G. P. Coelho and A. E. A. Silva, "Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media," *Applied Artificial Intelligence*, vol. 34, no. 1, 2020, doi: 10.1080/08839514.2019.1673037.

[2] M. Nofer and O. Hinz, "Using Twitter to Predict the Stock Market: Where is the Mood Effect?" *Business and Information Systems Engineering*, vol. 57, no. 4, 2015, doi: 10.1007/s12599-015-0390-4.

[3] A. Dridi, M. Atzeni and D. Reforgiato Recupero, "FineNews: fine-grained semantic sentiment analysis on financial microblogs and news," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 8, 2019, doi: 10.1007/s13042-018-0805-x.

[4] T. Daudert, "Exploiting textual and relationship information for fine-grained financial sentiment analysis," *Knowledge-Based Systems*, vol. 230, 2021, doi: 10.1016/j.knosys.2021.107389.

[5] M. Li, L. Chen, J. Zhao, and Q. Li, "Sentiment analysis of Chinese stock reviews based on BERT model," *Applied Intelligence*, vol. 51, pp. 5016–5024, 2021, doi: 10.1007/s10489-020-02101-8.

[6] S. Kelly and K. Ahmad, "Estimating the impact of domain-specific news sentiment on financial assets," *Knowledge-Based Systems*, vol. 150, 2018, doi: 10.1016/j.knosys.2018.03.004.

[7] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," *IEEE Access*, vol. 8, pp. 131662-131682, 2020, doi: 10.1109/ACCESS.2020.3009626.

[8] L. Zhao, L. Li, X. Zheng, and J. Zhang, "A BERT based sentiment analysis and key entity detection approach for online financial texts," *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2021, pp. 1233-1238, doi: 10.1109/CSCWD49262.2021.9437616.

[9] K. Du, F. Xing, and E. Cambria, "Incorporating Multiple Knowledge Sources for Targeted Aspect-based Financial Sentiment Analysis," *ACM Transactions on Management Information Systems*, vol. 14, no. 3, 2023, doi: 10.1145/3580480.

[10] C. C.-F. Chu, R. So, E. K.-L. Kwong, and A. Chan, "Leveraging rule-based model and machine learning transformer for mining aspect-based financial opinions in colloquial language," *Proceedings of the 7th International Conference on Software and e-Business*, 2023, pp. 71-78, doi: 10.1145/3641067.3641075.

[11] K. S. Kalyan and S. Sangeetha, "BertMCN: Mapping colloquial phrases to standard medical concepts using BERT and highway network," *Artificial Intelligence in Medicine*, vol. 112, 2021, doi: 10.1016/j.artmed.2021.102008.

[12] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Information Fusion*, vol. 36, 2017, doi: 10.1016/j.inffus.2016.10.004.

[13] K. W. Trisna and H. J. Jie, "Deep Learning Approach for Aspect-Based Sentiment Classification: A Comparative Review," *Applied Artificial Intelligence*, vol. 36, no. 1. 2022, doi: 10.1080/08839514.2021.2014186.

[14] T. A. Rana and Y. N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artificial Intelligence Review*, vol. 46, no. 4, 2016, doi: 10.1007/s10462-016-9472-z.

[15] S. G. Chowdhury, S. Routh, and S. Chakrabarti, "News Analytics and Sentiment Analysis to Predict Stock Price Trends," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, 2014.

[16] R. Kumar, H. S. Pannu, and A. K. Malhi, "Aspect-based sentiment analysis using deep networks and stochastic optimization," *Neural Computing and Applications*, vol. 32, no. 8, 2020, doi: 10.1007/s00521-019-04105-z.

[17] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, 2015, doi: 10.1016/j.knosys.2015.06.015.

[18] S. Welleck, J. Weston, A. Szlam, and K. Cho, "Dialogue natural language inference," *arXiv*, 2018, doi: 10.48550/arXiv.1811.00671.

[19] A. Vaswani *et. al*, "Attention is all you need," *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 6000-6010, doi: 10.5555/3295222.3295349.

[20] K. Clar, U. Khandelwal, O. Levy, and C. D. Manning, "What does BERT look at? an analysis of BERT's attention," *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 2019, pp. 276–286, doi: 10.18653/v1/w19-4828.

[21] H. j. Park, M. Song and K. S. Shin, "Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories," *Knowledge-Based Systems*, vol. 187, 2020, doi: 10.1016/j.knosys.2019.06.033.

[22] A. Kumar, A. S. Veerubhotla, V. T. Narapareddy, V. Aruru, L. B. M. Neti, A. Malapati, "Aspect term extraction for opinion mining using a Hierarchical Self-Attention Network," *Neurocomputing*, vol. 465, 2021, doi: 10.1016/j.neucom.2021.08.133.

[23] C. Raffel *et. al*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Researc*h, vol. 21, no. 1, pp. 5485-555, 2020, doi: 10.5555/3455716.3455856.

[24] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of naacL-HLT*, 2019, pp. 4171–4186.

[25] B. Huang and K. M. Carley, "Syntax-aware aspect level sentiment classification with graph attention networks," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 5469–5477, doi: 10.18653/v1/D19-1549.

[26] J. A. Garcia-Diaz, F. Garcia-Sanchez, and R. Valencia-Garcia, "Smart Analysis of Economics Sentiment in Spanish Based on Linguistic Features and Transformers," *IEEE Access*, vol. 11, pp. 14211-14224, 2023, doi: 10.1109/ACCESS.2023.3244065.

[27] K. N. Prasanthi, R. E. Madhavi, D. N. S. Sabarinadh, and B. Sravani, "A novel approach for sentiment analysis on social media using BERT & ROBERTA transformer-based models," *Proceedings of IEEE 8th International Conference for Convergence in Technology (I2CT)*, 2023, pp. 1-6, doi: 10.1109/I2CT57861.2023.10126206.

[28] D. G. Mandhasiya, H. Murfi and A. Bustamam, "The hybrid of BERT and deep learning models for Indonesian sentiment analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, pp. 591-602, 2024, doi: 10.11591/ijeecs.v33.i1.pp591-602.

[29] X. Weng, X. Lin, and S. Zhao, "Stock price prediction based on lstm and bert," *Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC)*, pp. 12–17, 2022, doi: 10.1109/ICMLC56445.2022.9941293.

[30] W. Li, C. Hu, and Y. Luo, "A Deep Learning Approach with Extensive Sentiment Analysis for Quantitative Investment," *Electronics (Switzerland)*, vol. 12, no. 18, 2023, doi: 10.3390/electronics12183960.

[31] T. A. Rana and Y. N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert Systems with Applications*, vol. 89, 2017 doi: 10.1016/j.eswa.2017.07.047.

[32] R. So, C. F. C. Chu, and C. W. J. Lee, "Extract aspect-based financial opinion using natural language inference," *Proceedings of the International Conference on E-business and Mobile Commerce*, 2022, pp. 83-87, doi: 10.1145/3543106.3543120.

[33] A. Zaninello and A. Birch, "Multiword expression aware neural machine translation," *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 3816-3825.

[34] T. Gao, A. Fisch and D. Chen, "Making pre-trained language models better few-shot learners," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2020, pp. 3816–3830, doi: 10.18653/v1/2021.acl-long.295.

BIOGRAPHIES OF AUTHORS

Carlin Chun Fai Chu received the Ph.D. degree from The Chinese University of Hong Kong in 2011. His research interests include large language models, data mining, financial modelling, and statistical computations. He is a Chartered Statistician (CStat) of the Royal Statistical Society. He can be contacted at email: carlinchu@hsu.edu.hk.

Raymond So received the Ph.D. degree from The University of Melbourne in 2008. His research interest includes machine learning, cloud computing, and software engineering. Proven ability to convey and explain highly complex and technical subject matter to an audience of different levels. He can be contacted at email: raymond.wc.so@gmail.com.

Ernest Kan Lam Kwong received the M.Sc. in Information Technology with Internet Applications and M.Sc. in Quantitative Analysis and Computational Mathematics in 2012 and 2015 respectively. He is a seasoned analyst and award winner in IBM programming contest in 2012. He can be contacted at email: ernestklkwong@gmail.com.

Andy Chan is a Co-Founder at iDDY.ai, a Co-Founder and Product Director at Play More, and a Managing Director at Quamnet. He is an experienced CTO and received HSBC Youth Business Award in 2014. He can be contacted at email: andy.chan@iddy.ai.