

캄보디아 음악 산업 재건을 위한 감정 분석 엔진 연구

삭소니타 코으른 · 김윤선[†]

Sentiment Analysis Engine for Cambodian Music Industry Re-building

Saksonita Khoeurn · Yun Seon Kim[†]

ABSTRACT

During Khmer Rouge Regime, Cambodian pop music was completely forgotten since 90% of artists were killed. After recovering from war since 1979, the music started to grow again in 1990. However, Cambodian popular music dynamic and flows are observably directed by the multifaceted socioeconomic, political and creative forces. The major problems are the plagiarism and piracy which have been prevailing for years in the industry. Recently, the consciousness of the need to preserve Khmer original songs from both fans and artist have been increased and become a new trend for Cambodia young population. Still, the music quality is in the limit state. To increase the mind-set, the feedbacks and inspiration are needed. The study suggested a music ranking website using sentiment analysis which data were collected from Production Companies Facebook Pages' posts and comments. The study proposed an algorithm which translates from Khmer to English, doing sentiment analysis and generate the ranking. The result showed 80% accuracy of translation and sentiment analysis on the proposed system. The songs that rank high in the system are the songs which are original and fit the occasion in Cambodia. With the proposed ranking algorithm, it would help to increase the competitive advantage of the musical productions as well as to encourage the producers to compose the new songs which fit the particular activities and event.

Key words : Cambodian Pop Music, Sentiment Analysis, Ranking Chart, Khmer, Google Cloud Computing

요약

캄보디아의 대중음악은 크메르 루즈 정권 기간 동안 예술가의 90 %가 사망 한 이래로 완전히 잊혀졌다. 1979 년부터 전쟁에서 회복 한 후 1990 년 음악은 다시 성장하기 시작했다. 그러나 캄보디아 대중 음악의 역 동성과 흐름은 다면적 사회 경제적, 정치적, 창조적 세력에 의해 관찰되고 있지만, 표절과 불법 복제로 수년간 대중음악산업에서 널리 퍼져 많은 문제가 되어왔다. 최근에는 크메르(캄보디아어) 전통 음악을 팬과 아티스트 모두에게 보존해야 할 필요성에 대한 의식이 높아져 캄보디아 젊은 인구의 새로운 트렌드가 되었으나, 음악 품질은 여전히 한계상태에 봉착해 있고, 전통 대중 음악의 전문성을 높이기 위해서는 대중의 드백과 영감이 필요하다. 이 연구는 캄보디아에서 가장 많은 대중음악 관련 사이트인 페이스 북 페이지의 게시물과 코멘트에서 수집된 문장들을 감정분석을 사용하여 음악 순위 차트(웹 사이트)를 구현하였다. 크메르어에서 영어로 번역하고 감정 분석을 수행하고 순위를 생성하는 알고리즘 개발하였다. 그 결과로 제안된 시스템에서 번역 및 감정 분석의 정확도가 80 %임을 보여주었다. 순위에서 높이 평가된 노래는 크메르(캄보디아어)로 된 전통대중음악으로 이 논문의 취지와 부합이 되었다. 캄보디아 전통대중음악을 다시 부활하기 위해서 제안된 시스템과 순위 알고리즘을 사용하여 음악 제작의 경쟁 우위를 높이고 제작자가 특정 활동 및 이벤트에 맞는 새 노래를 작곡하는 데 도움이 될 것으로 사료된다.

주요어 : 캄보디아 대중음악, 감정분석, 음악순위차트, 크메르(캄보디아어), 구글 클라우드 컴퓨팅

Received: 16 November 2017, **Revised:** 6 December 2017,

Accepted: 7 December 2017

† Corresponding Author: Yun Seon Kim

E-mail: sean0831@handong.edu

Handong Global University

1. Introduction

During Golden Age (the 1960s and the 1970s), Cambodian culture and art were known as the most

advanced and lively music in Asia. Many genres such as Rock ‘n’ Roll, Rhythms and Blues, and Country music were introduced to South East Asia including Cambodia. Not only western music, Talent Producers, and Singers had created Cambodian music genre which had Khmer color into the music. However, the liberal culture, talented artist and producers were killed during Pol Pot Regime (1975-1979). Anything related to liberal arts and immoral culture was abolished and destroyed. Cambodian music including Electronic and dance music including harmony of guitars began to grow again since 1990. The most significant problems in the Entertainment industry is facing the collapse. According to Ministry of Culture and Fine Arts indicated that there were 67 films and karaoke production films and karaoke album in 2008; however, there are just 31 companies remained due to rampant piracy and the economic downturn led to low-quality products and prices^[1]. Even though the music industry has increased popularity among youth in the country, it has been noticed that most of the music production companies plagiarize the songs from other countries instead of making their own music since composing music is much more difficult than singing and it requires a lot of investment to make an entire song and write lyrics which could not reach the requirement of the market. Moreover, songwriters say they still face major obstacles in gaining a foothold writing original music, not least of which is the problem of how to earn money selling music in the era of digital downloads and YouTube^[2]. Additionally, quality of the music is not good enough because of plagiarism from other foreign countries^[3]. Some companies were bankrupt due to the consistency market of massive copied songs. The copied songs are massively criticized from the public in the social network.

Therefore, to encourage and improve the music creation in Cambodia, the study will encounter the reliable marketing strategies that need to imply for the music production companies. The study suggested a music ranking website using sentiment analysis which data were collected from Production Companies

Facebook Pages' posts and comments. The studied used implementation driven research methodology following XP development method of Agile Development methodology. Mostly, the fans comment on the posts in the Khmer Language; therefore, the study proposed an algorithm which translates from Khmer to English, doing sentiment analysis and generate the ranking. The result showed 80% accuracy of translation and sentiment analysis on the proposed system. The songs that rank high in the system are the songs which are original and fit the occasion in Cambodia. With the finding of the studies, the music trends will create a new culture for the whole country. Moreover, the opinion mining found from the research will increase the competitive advantage of the music productions as well as to encourage the producers to compose the new songs which fit the particular activities and event. According to the music producer said that "creating a new sound will be amplified to the next generation, impacting them and impacting our society in return. If we can gain a more comprehensive awareness of how our art form is making a difference around us, we will undoubtedly become better musicians - musicians with a purpose^[4]. More significantly, through our methodology of research in the social network in the field of Cambodian popular music, it can be applied to any areas in the Entertainment such as variety shows, movie trailers. Not just in the Entertainment industry, but this approach could be applied in another field such as tourism since our country is based on the tourism and also to the food, restaurants, and hotels. Further, our research will mine the data from Khmer; therefore, this also can apply to another language with the purpose of finding the opinion of the customers on their services or products.

The study is divided into six sections. First section talked about the introduction of the study; second section portrayed the literature review which related to the study. The third section explained the research methodology; meanwhile, the proposed system described in fourth section. Fifth section is for interpreting the result found from the research whereas the last section is for conclusion and further research.

2. Literature Review

2.1 Music Ranking Charts

The music ranking chart plays a major role in industry of music. According to Giles, the outcome of music chart represents the information to the market of music^[5]. Some of music charts are based on the CDs sales for example, Billboard and UK Charts^[6]. However, we cannot value the music popularity only by the sales number since there are many ways for fan to listen the music nowadays. Due to the digital download, the most famous music should be not the best-selling one^[7]. To have the album or songs featured in the charts, basically without a doubt, be regarded as the basic achievement of most pop music artists and their labels. By having the song in the chart, it represents as the quality of public recognition for the artists and their record label's talents. Besides, it also indicates the appreciable artists' earning amounts with gathered directly with their pre-recorded musical output's sale and through indirectly related income sources comprising merchandise products, public performance, touring, etc. Additionally, the success in chart will regard as publicity which can gain more sales for current and future work^[8].

There are many well-known charts available in some countries. For instance, in the United States music industry, there is Billboard Hot 100 is a standard popularity charts of singles offered weekly by the Billboard Magazine^[9]. The ranking of the chart is basically relied on sales data and radio play which gathered 10 days before the chart is released. The process of ranking does not involve in file sharing activity of the account. The new chart is collated and released officially to the public every Tuesday. In billboard magazine's websites, top 50 singles currently ranked on the charts are published every week. Meanwhile, the full charts are accessible only for on-line subscribers who paid in a small amount of fee^[10].

Meanwhile, in Korean music industry, there is an official ranking chart owned by the Korea Music Content Industry Association and the Ministry of Culture, Sports and Tourism. With the cooperation between two organizations, they established a local music chart which

is resembled Billboard Charts in 2010, called "Gaon Chart". This local chart is based on offline and online data of music and album sales. The chart is divided into five categories the Gaon Chart, Online Chart, Digital Chart, Album Chart and Mobile Chart^[11].

The Album chart is published with the total number of album sales by six big album distributing companies such as SM Entertainment, LOEN Entertainment, Universal Music, Mnet Media, Sony Music Korea and Mnet Media. The online chart calculates the sales of music on some sources, streaming services and online album sales. For the digital chart, Gaon always release every six months. The chart will calculate Mobile and Online Chart. Besides, the other charts release offline and online on a basis of weekly^[11]. The result online is calculated by the given data from the top music websites in South Korea such as Mnet, Melon, Dosirak, and Cyworld. The online and offline data are based on the results created by precisely 97 percents of Internet Users who buy music online^[11].

2.2 Social Network Analysis

The Social Media sites offer the function that enables the people to make connections with other users and brands and share the information and content such as news, message, photos, and videos. The social network such as Facebook, Twitter, and YouTube have experienced extraordinary rise these years and are some of the most talked about destination on the Web. Social Media provides an opportunity for the business to connect and interact with the customers, promote its brand and lead the traffic in the website. "Using social media is a significant way to make business seem friendlier and to form closer relationships with the customer"^[12]. The companies can directly interact with customers by engaging in social media and by creating useful and interesting content companies can position themselves as an expert in the field^[13]. Even though social media has the significant impact on business, its marketing has received little research attention despite managers and scholars in the travel, hospitality, and tourism industries have urged that social media should become a first priority for future research. Particularly,

research should concentrate on helping managers confront the Web 2.0. hurdles, understanding the behavior of the consumer, using social media in handling customer interactions, and measuring customers' activities and outcomes^{[14], [15]}. It should be marked that there is a need to develop a clear methodology for data mining related purposes^[16]. There are many research that have been done in order to analyze the data from the social network for marketing purpose. There are many of works have been done so far that have dealt with different perspective, mostly sentiment extraction, retrieval, classification, summarization, and presentation^{[17], [18][19]-[23]}. Particularly, Data that they have been used is from Twitter to do the text mining and sentiment analysis to found out the brand popularity, the users' opinion and feedback^{[21], [24]-[27]}.

However, there is not much research for sentiment analysis using Facebook data in the field of Entertainment yet. There are just some works done to classify the opinion in e-learning^[28]. Besides, there are a research using Naive Bayes Classifier to help language learning, by stimulating the educational process and experimental results^[29].

2.3 Sentiment Analysis

The methods for doing sentiment analysis could be categorized into two parts such as language processing based and application oriented methods. In the research of Liu, instead of addressing the sentiment analysis from two difference aspects, lexicon-based, namely, and linguistic analysis, he gave the brief overview on optical analytics approaches in sentiment analysis. The most common and significance indicators of sentiments are the opinion or sentiment words such as good, poor, bad, amazing and also some idioms phrases which are used to express positive or negative opinion. The opinion lexicon or sentiment lexicon is the bag of such those kind of words and phrases and is important; however, it is not enough for sentiment analysis. Moreover, to exploit lexicons, linguistic based methods also use the text grammatical structure for sentiment classification. There are two types of lexicon generation approaches namely, dictionary based^{[17], [22], [30]-[32]} and

corpus-based methods^[25]. With the first approach, they started with a small dataset of opinion words and extends the lexicon by “bootstrapping a certain dictionary” whereas the second approach creates the lexicon of opinion through learning the dataset.

For the world wide real-life applications of sentiment analysis and opinion mining are one of many reasons that sentiment analysis is the common research problem. Because of these provided applications by opinion mining, many activities particularly those industries have increased in recent years which have widespread to almost every possible domain such as services, products, financial services, healthcare, social events and the political elections. By applying sentiment analysis to social network data, there are many applications in social media that have been done to movie review^{[33], [34]} product reviews^{[34], [35]}, App Reviews^[36], the prediction of stock market^[37], and the detection of trend^[38].

There is many research have been done using text mining approach to find the sentiment from the users. The expressed opinions in social networks play a significant role impacting public opinion's behavior throughout areas as various as buying products, seizing the “pulse” of stock markets and voting for the president^[27]. The opinion from online has been analyzed recently using sentiment analysis which is basically a natural language processing application that uses text mining and computational linguistics to identify text sentiment, typically as negative, neutral or positive. some languages might claim that such trend represents poor language use is assisted by technological advancements along with social factors^[39]. This complicated factor regarding common Web texts' sentiment detection has been dispensed with through several techniques, including word sense disambiguation^[40], negation accurate detection^[41], and gathering semantic orientation from the connection^[42]. Successfully dealing with this problem has led to a plethora of online sentiment analyses in texts written in languages such as English^[20], Chinese^[43], Arabic, and multi- languages^[17]. In the field of music and video streaming, there are also many research conducted by using the text mining and sentiment analysis. Particularly, the data mining that use the data

from the major platform such as Amazon, Spotify, and many other platforms. In the paper of^[44], they analyzed the review text sentiment to predict the product ratings by extracting the features from text reviews and the model of regression to predict ratings of review from Amazon Digital Music.

Meanwhile, there are also some research that have done in South Korea to find out the opinion of fans from the microblogging sites. This study carries out an analysis using text mining that extracts portal site blogs, twitter page views and associative terms. In addition, this study derives viewers' response by extracting key words with opinion mining techniques that divide positive words, neutral words and negative words through customer sentiment analysis^[18].

2.4 Cloud Machine Learning

There are many libraries and cloud computing system are available for machine learning. Since Machine learning is a complicated for coding. The major companies such as Google, Yahoo and others have developed Machine Learning with Software-as-a-Service for startups. Using Machine Learning is time consuming, so the developers have to take too much efforts. Therefore, paradigm of Cloud Computing and cloud providers are precious changes to stimulate machine learning platform. In the research from Daniel Pop, he did the survey which found that Machine Learning in circulated environments come in variety of approaches, providing feasible and cost-effective option to traditional Machine Learning and statistical application that are not absorbed on distributed environments^[45]. Meanwhile, Google also officially published its cloud computing for startup called Google Cloud. The technologies provided new cloud and machine learning tool host in a part of the package. There are many businesses are using the google service such as Evernote, Snapchat Inc, and Airbus^[46].

In Google Cloud Machine Learning, there are many services provided. One of them is Google Cloud Natural Language that provide powerful text analysis such as syntax analysis, entity analysis, sentiment analysis, entity sentiment analysis, and multi-language

analysis. Since Google have just offered this service in 2016, there is not much research using this service yet. The study have searched in Google Scholar and Science Direct by using keyword “google cloud machine learning”, but we have found only one article that use video-to-text API and natural language processing of Google to focus on linguistic characteristics of lead user entrepreneur by using 404 videos from crowdfunding^[47].

3. Research Methodology

In the study, the research methodology is a crucial part of the study regarded as the main ingredient in order to address the research questions or objectives. The methodology clarifies how a particular approach would be implemented. The research used the implementation driven research that aims in solving Cambodian popular music problem that use sentiment analysis to classify the opinion from the Social network and make the ranking, the study used implementation driven research^[48].

First, the data we collected are from Facebook. Those data included page posts, comments, likes, and other related information. To gather data from Facebook, the study use Facebook Graph API explorers. The Graph API that Facebook provide is an efficient way to collect the data from Facebook^[49]. Since the data got from Facebook Fan Page are in Khmer, the comments saved in database need to be translated into in English for doing sentiment analysis. To translate the comments from English, the study use GOOGLE CLOUD TRANSLATION API with Java Programming Language. Google Cloud Translation API offer simple interface for programming to translate string into any language by using Neural Machine Translation^[50]. After translation into English, the study did the sentiment analysis which identify positive and negative feedback from the comments. To do sentiment analysis, there is a library from Google Cloud Platform to perform sentiment analysis. The service is Google Cloud Natural Language API which is a powerful text analysis library that can be used with Java Programming. This API show the text meaning by providing machine learning models in

easy way^[51]. Together with this API, the study also used Spring Framework to do the analysis. The sentiment analysis result got from this API provided information such as positive, negative, and neural.

4. KHHits.com

Khhits.com is a music ranking chart website which based on the number of likes and share, and the sentimental analysis of fans comments from music companies Facebook page posts. Khhits is the first website in Cambodia for music ranking chart. Khhits aims to measure the popularity of the music and artists in Cambodia. In order to do the ranking, the likes, comment, share in each post will be counted and calculated to know which song has the most like, comment, and share. Moreover, the system will update the ranking chart every week. Besides, it is the only one website where they can enjoy the music from all the major companies in Cambodia. It ensures the copyright of the music; the system embed the original url from the companies' Facebook pages. Moreover, the website analysed the fans opinion about the music from each company: the research will aim to know the positive and negative opinion from the fans comment on each post of the Facebook pages.

4.1 Marketing of Music

To spread their music to the fans, the production companies and musicians upload the music in YouTube. YouTube is the biggest online video platform which can help the musicians reach and provide additional promotion of music video. Therefore, music production companies in Cambodia upload the full audio and music videos in YouTube. Beside YouTube, they also use other social networks to reach the fans. Meanwhile in Cambodia, Facebook is the most popular social network for Cambodian users. As stated in the report, there are 4.8 million users in 2017^[52]. Among the users, 3,700,000 users are interested in the entertainment including music. Compare to YouTube, Cambodian internet users are more active in Facebook. Therefore, the production companies openly uploaded the music

videos in their official Facebook Pages. It's also supported that Facebook has more data about its users than other platforms that means they can collect the data about who's viewing the video, and have a well targeting audience when it comes to its Facebook advertisement^[53].

Throughout the interview with Mr. Pol Vibo, who is the Director of Town Production Company (Pol, 2017a), to find the popularity of the songs, the company manually check the feedbacks from the fans by the number of likes and shares also what fans commented which called word-of-mouth. However, they could not do it often. He also said that the company need the system which can provide the company the data. Original songs are going viral and more people listen to Khmer songs. But, it is still not yet well acknowledged. Also, the culture of copy the music is still going on. Therefore, the music productions need to always check the opinion and feedback for expanding their marketing plan from the fans. However, the companies just manually checked the feedback from Facebook posts. Moreover, the 59.1% of respondents suggest to have the music ranking chart to know the popular songs.

4.2 Analysis on proposed system

The collected data are from Facebook. Those data included page posts, comments, likes, and other related information. To gather data from Facebook, the study used Facebook Graph API explorers. Figure 1 explained the process of gathering the data. The system connects to the Facebook Graph API through user access token API provided by Facebook. Then it will access to Facebook Page by its ID that is saved in database. After that, it will collect the posts, comments, users from the page. If the post exists in the database, it will

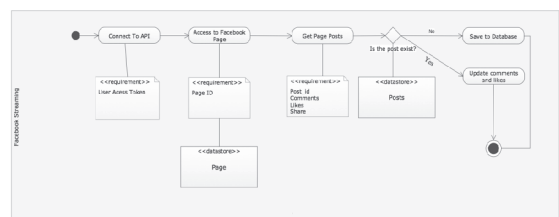


Fig. 1. Facebook Data Streaming Activity Diagram

update the comment and likes. If the post does not exist in database, it will be saved to database.

The figure 2 described the process of translating the comments and analyze the comments. The system first need to connect to the Google Cloud Platform that provide the API for translating and doing sentiment analysis. Before doing these processes, admin need to check the posts that collected from the pages. The posts need to be about the music. If the posts are not related to the music, the posts will be deleted from database. Also, the titles of the music are required to be edited since some pages have “name” field null, so the admin has to manually edit the title. After that, the admin can run the translate process. The process starts from getting all the comments from the database. Each comments are translated to English and save to database. Then, each comments are analyzed into positive and negative then save them to table posts. Also, each posts are also analysed sentiment and save the result to database. The system generates the ranking by calculating the number of likes 50%, number of share 10%, and positive score 40%.

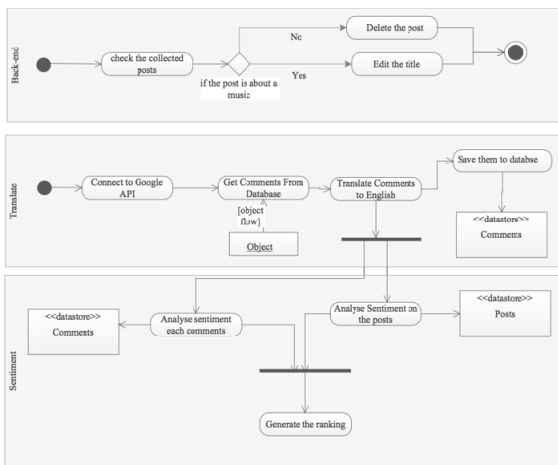


Fig. 2. Process of translate and sentiment

5. Analysis and Result

5.1 Result and Discussion

5.1.1 Facebook Data Streaming

The study collected the posts from seven Facebook Pages of the production companies which always upload the music into their pages. Those companies are Rasmey Hangmeas, Town Production, Sunday Production, M Production, Tena, Kmeng Khmer, and Small Word Small Band. As the result, The posts and comments from the pages are collected and saved to databases with the developed system using Spring Framework.

5.1.2 Sentiment Analysis

To do the Sentiment Analysis from all the comments, the research developed a system that atomically analyse the comments by using Spring Framework with Google Cloud Natural Language API. Before doing the sentiment analysis, the system firstly translates the comments into English then start to analyze sentiment in each comments. There are four columns that get from the database. The first comment could not be translated properly since the user commented in English which spell as Khmer. The second comment which is in Khmer means “The famous song writer, Son Ratha, has come to Sunday” also similar to the translated text from API “That Famous writer, Sonar, Came to Sande”. It reflects as the positive. Whereas the score from the API also offered 0.6 which means positive. However, the fourth comment translation is wrong. The comment meant “It's the song that I am waiting, it's really nice. But it lacks of flute sound”. But, the message translated to English from API means “the song I am awaiting is almost gone”. Also, the comment contains positive and also negative feedback, but the score that provided by the system is -0.2 which mean negative. Out of 10 comments, there are two comments could not be considered as accurate while the other 8 comments are accurate enough. Meanwhile, the figure 3 display top ten songs. The number one song is a song called "Confused Girl" which has 5569 likes, 3085 shares and 100% positive feedback from the fans. Even number 5th song has more likes, but the positive the feedback just only 84 %. The song number 2 and number 5 are the song related to one of the biggest holiday in Cambodia, that is why there are many positive feedbacks from fans. Beside these songs, the songs are self-compose songs by the famous singers

in Cambodia.

#	🎵 Title	Likes	Share	Positive	Neutral	Negative	Comments	📊 Average
1	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	568	365	100	0	0	👍	0.1000000000000000
2	🎵 អរគុណ	512	254	100	0	0	👍	0.1000000000000000
3	🎵 អរគុណប្រទេសខ្មែរ-ស្រុកខ្មែរ	381	124	100	0	0	👍	0.1000000000000000
4	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	44	24	100	0	0	👍	0.1000000000000000
5	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	124	131	84	4	12	👍	0.1000000000000000
6	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	444	456	96	9	9	👍	0.1000000000000000
7	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	1182	1042	83	12	4	👍	0.1000000000000000
8	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	630	123	96	0	10	👍	0.1000000000000000
9	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	1822	179	78	11	9	👍	0.1000000000000000
10	🎵 ប្រទេសខ្មែរ-ស្រុកខ្មែរ	212	190	52	10	1	👍	0.1000000000000000

Fig. 3. Top Ten Songs

6. Conclusion and Further Research

Cambodian Popular Music was completely destroyed during Khmer Rouge. During that genocide, over two million people were killed include 90% percent of nation's popular singers and musicians. After the country recovered from the wars during the mid 1990s, electronic broadcast and consumer media were specifically accessible to Cambodians. Currently, the Cambodian popular music dynamic and flows are observably directed by the multifaceted socioeconomic, political and creative forces. Due to these competing forces, Cambodian popular music status is in risky (Mamula, n.d.). To advertise the music, the companies uploaded the music into their official Facebook and YouTube accounts. To check the popularity and feedback of the songs, the companies manually check them once in a while according to the interview with^[55]. The study proposed a ranking chart system which based on the number of likes and share, and the sentimental analysis of fans comments. To conduct the research, the study used implementation driven research to recommend a ranking chart websites. The development of the website followed system development life cycle using XP agile development methodology. The development has four main processes - data collection, translate the comments, sentiment analysis, and ranking. To collect the Facebook data, the study used Facebook Graph Explore API, Facebook4J and Spring Thymleaf Frameworks. Meanwhile, the translation and sentiment analysis were developed by using Google Cloud Natural Language API. The result has shown 80% of accuracy of

translation and sentiment analysis. Some fans commented in Khmer using English spelling which lead to error for translation and sentiment analysis. The songs that got the positive feedbacks are mostly the original songs.

With the methodology of the research, the music production can expand competitive advantage and effectively assess the business competitive environment. It is believed that competitive intelligence can help the music companies to realize the feedback from the fans include weakness and strengths, enhance the music compositions. Moreover, it is hoped that the research would be a contribution to the evolution of Cambodian popular music evolution to grow more just like back in Golden age before the genocide. Plus, the companies also could find the trends of the music that currently people like to listen. Furthermore, the advertising operations can hold all the connection details they have with fans, and conducting smart publicity strategies that enable them to maintain the high level of buzz, not just only during release times but all over the entire year. By using the fans' comments as the feedback about the artwork and music, it would encourage the electronic word of mouth which is called "e-WOM". This can be discovered online without using the huge amounts of money on investing the traditional advertising and marketing campaigns. Also, the negative opinions should not be ignored because such opinions could be utilized for detecting what are the flaw with music and service. The comments on each post can be used effectively to distinguish fans' preferences, to get the related dissatisfaction to a song deficiency, and to also to correct the intention of errors. The sentiment analysis lets the companies be more efficient and functional in dealing with fans because sentimental analysis can be useful in such the areas as mentioned above such as analysing the trends in music, handling fans' feedback and targeting the advertising operations. The comments on each post can be used effectively to distinguish fans' preferences, to get the related dissatisfaction to a song deficiency, and to also to correct the intention of errors. The sentiment analysis lets the companies be more efficient and functional in dealing with fans because sentimental analysis can be useful in such the areas as

mentioned above such as analysing the trends in music, handling fans' feedback and targeting the advertising operations.

For the further study, the study suggests to include the sentiment analysis from other social networks particularly YouTube for better ranking algorithm. Also, the further study should include the videos from other entertainment sectors such as movies, and TV shows. Moreover, it would be better to include the users' statistics on each posts. With this study approach, sentiment analysis will be introduced to Cambodia for other industries such as tourism, education, and customer relationship management. Through the accuracy of result from the study, public and private organization are able to use this technology for purpose to find the knowledge from the massive data from social network.

References

- [1] "Entertainment industry facing 'collapse', Business, Phnom Penh Post" Entertainment industry facing "collapse", Business, Phnom Penh Post, 26-May-2009.
- [2] "Not the same old song, Post Weekend, Phnom Penh Post" Not the same old song, Post Weekend, Phnom Penh Post, 03-Oct-2015.
- [3] S. Mamula, "Starting from Nowhere?: Popular Music in Cambodia after the Khmer Rouge" Starting from Nowhere?: Popular Music in Cambodia after the Khmer Rouge, *Asian Music*, vol. 39, pp. 26-41, 2008.
- [4] "What Kind of Impact Does Our Music Really Make on Society?" What Kind of Impact Does Our Music Really Make on Society?, 09-Apr-2017.
- [5] D. E. Giles, "Increasing returns to information in the US popular music industry" Increasing returns to information in the US popular music industry.
- [6] D. E. Giles, "Survival of the hippest: life at the top of the hot 100" Survival of the hippest: life at the top of the hot 100, *Appl. Econ.*, vol. 39, pp. 1877-1887, Aug. 2007.
- [7] Y. Yang, C. Chen, and Y. Chen, "The Measurement of the Search Charts of Music" The Measurement of the Search Charts of Music, in *International Conference on Networks Security, Wireless Communications and Trusted Computing*, 2009, pp. 346-349.
- [8] A. S. Eric and T. Clive, "The Dynamics of Chart Success in the U.K. Pre-Recorded Popular Music Industry" The Dynamics of Chart Success in the U.K. Pre-Recorded Popular Music Industry, *J. Cult. Econ.*, vol. 24, no. 2, pp. 113-134, May 2000.
- [9] N. Koenigstein and Y. Shavitt, "SONG RANKING BASED ON PIRACY IN PEER-TO-PEER NETWORKS" SONG RANKING BASED ON PIRACY IN PEER-TO-PEER NETWORKS.
- [10] "Billboard biz | Billboard" Billboard biz | Billboard, 16-Jul-2017. [Online]. Available: <http://www.billboard.com/biz>. [Accessed: 16-Jul-2017].
- [11] S. Han, "Korea Launches First Official Music Charts Gaon" Korea Launches First Official Music Charts Gaon, 23-Feb-2010.
- [12] Q. Keith, "8 Ways to Improve Customer Relationships With Social Media: Social Media Examiner" 8 Ways to Improve Customer Relationships With Social Media: Social Media Examiner, 01-Jul-2014.
- [13] "How can social media impact my business make your marketing personal" How can social media impact my business make your marketing personal, Mobile commerce stats for small business. 25-Mar-2017.
- [14] T. Hennig-Thurau, E. C. Malthouse, C. Friege, S. Gensler, L. Lobschat, A. Rangaswamy, and B. Skiera, "The Impact of New Media on Customer Relationships" The Impact of New Media on Customer Relationships, *J. Serv. Res.*, vol. 13, Aug. 2010.
- [15] L. Kwok and B. Yu, "Spreading Social Media Messages on Facebook: An Analysis of Restaurant Business-to-Consumer Communications" Spreading Social Media Messages on Facebook: An Analysis of Restaurant Business-to-Consumer Communications, *Cornell Hosp. Q.*, vol. 54, Feb. 2013.
- [16] T. Grizane and I. Jurgelane, "Social Media Impact on Business Evaluation" Social Media Impact on

- Business Evaluation, *Procedia Comput. Sci.*, vol. 104, 2017.
- [17] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums" *Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums*, *ACM Trans. Inf. Syst.*, vol. 26, Jun. 2008.
- [18] S. Ahn, S. Oh, and J. Byun, "A Big Data Study on Viewers' Response and Success Factors in the D2C Era Focused on tvN's Web-real Variety 'SinSeoYuGi' and Naver TV Cast Programming" *A Big Data Study on Viewers' Response and Success Factors in the D2C Era Focused on tvN's Web-real Variety "SinSeoYuGi" and Naver TV Cast Programming*, *Int. J. Adv. Cult. Technol.*, vol. 4, pp. 7-18.
- [19] R. C. Britto, "Branding and communication on Twitter" *Branding and communication on Twitter*.
- [20] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter Power: Tweets As Electronic Word of Mouth" *Twitter Power: Tweets As Electronic Word of Mouth*, *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, Nov. 2009.
- [21] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *Twitter as a Corpus for Sentiment Analysis and Opinion Mining.*, in *LREC*, 2010.
- [22] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in Twitter events" *Sentiment in Twitter events*, *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, Feb. 2011.
- [23] G. M. Thomaz, A. A. Biz, E. M. Bettoni, L. Mendes-Filho, and D. Buhalis, "Content mining framework in social media: A FIFA world cup 2014 case analysis" *Content mining framework in social media: A FIFA world cup 2014 case analysis*, *Inf. & Manag.*
- [24] A. Go and R. B. L Huang, "Twitter sentiment analysis"
- [25] A. Go and L. H. R Bhayani, "Twitter sentiment classification using distant supervision"
- [26] A. Sarlan and S. B. C Nadam, "Twitter sentiment analysis" *Twitter sentiment analysis*.
- [27] H. Wang and A. K. D Can F.Barçın, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle" *A system for real-time twitter sentiment analysis of 2012 us presidential election cycle*.
- [28] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning" *Sentiment analysis in Facebook and its application to e-learning*, *Comput. Hum. Behav.*, vol. 31, Feb. 2014.
- [29] C. Troussas and K. E. M Virvou, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning" *Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning*, in *IISA 2013*, 2013.
- [30] O. Kucuktunc and I. W. BB Cambazoglu, "A large-scale sentiment analysis for Yahoo! answers" *A large-scale sentiment analysis for Yahoo! answers*.
- [31] O. Kucuktunc, B. B. Cambazoglu, I. Weber, and H. Ferhatosmanoglu, "A large-scale sentiment analysis for Yahoo! answers" *A large-scale sentiment analysis for Yahoo! answers*, in *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*, 2012, p. 633.
- [32] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding Domain Sentiment Lexicon Through Double Propagation" *Expanding Domain Sentiment Lexicon Through Double Propagation*, in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 2009, pp. 1199-1204.
- [33] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques" *Thumbs up?: sentiment classification using machine learning techniques*, in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02*, 2002, pp. 79-86.
- [34] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis" *Opinion Mining and Sentiment Analysis*, *Found. Trends Inf. Retr.*, vol. 2, 2008.

- [35] M. Hu and B. Liu, "Mining and summarizing customer reviews" Mining and summarizing customer reviews, in Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04, 2004, p. 168.
- [36] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, "Why people hate your app: making sense of user feedback in a mobile app store" Why people hate your app: making sense of user feedback in a mobile app store, in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13, 2013, p. 1276.
- [37] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market" Twitter mood predicts the stock market, *J. Comput. Sci.*, vol. 2, Mar. 2011.
- [38] P. Hennig, P. Berger, C. Lehmann, A. Mascher, and C. Meinel, "Accelerate the detection of trends by using sentiment analysis within the blogosphere" Accelerate the detection of trends by using sentiment analysis within the blogosphere, in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 2014, pp. 503-508.
- [39] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment in Short Strength Detection Informal Text: Making the case for logarithmic binning" Sentiment in Short Strength Detection Informal Text: Making the case for logarithmic binning, *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, no. 12, pp. 2417-2425, Dec. 2010.
- [40] T. Pedersen, "A Decision Tree of Bigrams is an Accurate Predictor of Word Sense" A Decision Tree of Bigrams is an Accurate Predictor of Word Sense, in Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies, 2001, pp. 1-8.
- [41] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews" Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, in Proceedings of the 12th International Conference on World Wide Web, 2003, pp. 519-528.
- [42] P. D. Turney and M. L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association" Measuring Praise and Criticism: Inference of Semantic Orientation from Association, *ACM Trans. Inf. Syst.*, vol. 21, no. 4, Oct. 2003.
- [43] D. J. Xu, S. S. Liao, and Q. Li, "Combining empirical experimentation and modeling techniques: A design research approach for personalized mobile advertising applications" Combining empirical experimentation and modeling techniques: A design research approach for personalized mobile advertising applications, *Decis. Support Syst.*, vol. 44, Feb. 2008.
- [44] A. Wong and B. Fatura, "Amazon Digital Music: Sentiment Analysis and Text Mining" Amazon Digital Music: Sentiment Analysis and Text Mining.
- [45] D. Pop, "Machine Learning and Cloud Computing: Survey of Distributed and SaaS Solutions" Machine Learning and Cloud Computing: Survey of Distributed and SaaS Solutions, Mar. 2016.
- [46] "Google Cloud Machine Learning Engine: The smart person's guide - TechRepublic" Google Cloud Machine Learning Engine: The smart person's guide - TechRepublic, 10-Aug-2017.
- [47] J. Kaminski, Y. Jiang, F. Piller, and C. Hopp, "Do User Entrepreneurs Speak Different?: Applying Natural Language Processing to Crowdfunding Videos" Do User Entrepreneurs Speak Different?: Applying Natural Language Processing to Crowdfunding Videos, in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17, 2017, pp. 2683-2689.
- [48] "Retrieving publications from Library" Retrieving publications from Library.
- [49] "Graph API Overview" Graph API Overview, 27-Jul-2017.
- [50] "CLOUD TRANSLATION API" CLOUD

TRANSLATION API, 24-Jul-2017.

- [51] "CLOUD NATURAL LANGUAGE API" CLOUD NATURAL LANGUAGE API, 24-Jul-2017.
- [52] S. Joseph, "Cambodia's 2017 Social Media & Digital Statistics" Cambodia's 2017 Social Media & Digital Statistics, 09-Apr-2017.
- [53] R. Nicholas, "Music Marketing Tips and Ideas - 50 Ways to Promote Your Music" Music Marketing

Tips and Ideas - 50 Ways to Promote Your Music, 27-Jul-2016.

- [54] "Sinn Sisamouth and the golden age of Khmer pop | Savong School Cambodia" Sinn Sisamouth and the golden age of Khmer pop | Savong School Cambodia, 02-Sep-2017.
- [55] V. Pol, "The Popularity of the song" The Popularity of the song, 20-Aug-2017.



Saksonita Khoeurn (saksonita@gmail.com)

2012~ Teacher, Hun Sen Saang High School, Ministry of Education Youth and Sports
2014 Bachelor of Computer Science, Norton University
2014~2015 IT Instructor, Korea Software HRD Center
2017 Master in ICT Convergence, Handong Global University, Pohang, South Korea

관심분야 : Web Development, BigData, Data Mining, Machine Learning, Cloud Computing



Yun Seon Kim (sean0831@handong.edu)

2001 한동대학교 경영학사, 전산공학사 학사
2004 University of Pittsburgh Information Science 석사
2010 Wayne State University Industrial and Systems Engineering 박사
2008 YS security and computers 대표
2010 Robust system inc. 부사장
2013 연변 과학기술대학교 상경학부 교수
2015~ 현재 한동대학교 재직 중

관심분야 : Knowledge management, intelligent system, data analysis, global development