Social Media Hashtag: Affective Assessment Based On Kansei Approach

Anitawati Mohd Lokman, Ahmad Azran Awang, Shamsiah Abd Kadir and Saidatul Rabah Hamidi
Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

Abstract: Kansei engineering was initially designed to accommodate the field of product design and business processes. However, it is now observed compatible to be applied for the purpose of understanding human emotions in terms of communication, psychological and information warfare. Malaysia’s political scene evolves alongside the prolific rise of social media where it becomes a battlefield for political parties to win the hearts and minds of the people through hashtags. Consequently, this study aims to comprehend the sensitivity of Malaysians towards various political hashtags and also whether a relationship exists between Kansei and the hashtags. In addition, this study employs social media mining—a clustering hashtags approach to categorising political hashtags used in this study. Then, a Kansei approach is implemented to selected hashtags. From the results obtained, it is evident that Malaysians do have political awareness, sensitivity and are able to respond to the hashtags that represent various political issues. Moreover, Kansei is able to verify the results, hence proving the existence of associations between Kansei and political hashtags. Ultimately, the results are significant in providing a prelude to Malaysian’s response towards political issues, especially, through the medium of hashtags.

Key words: Hashtag, social media mining, clustering, information warfare, Kansei, response

INTRODUCTION

The advent of Twitter enables people to write about anything they want as their status messages and share these messages with their friends and followers (Milestone et al., 2008). Important events like elections, disasters, concerts and football games can have immediate and direct impact on the density of status updates. As a result, Twitter provides a great medium for situational awareness using the messages written by millions of real world users.

Antenucci et al. (2011) enlightened that hashtags are a way for a user to designate a subject of a tweet in a way that is easy to search for, hence, it could deliberate the metadata. In this situation, they scrutinised that hashtags could follow that when multiple hashtags occur in the same tweet, they represent a similar approximation of content, the more often two hashtags co-occur, the more similar their meanings are.

Nowadays, social media play significant roles in supporting the democratisation of knowledge and information. It has since, transformed the common people from information seekers and users into information producers (Sani and Zengeni, 2010). Social media are distinct compared to traditional media which include newspapers, television and film. Relatively, social media are budget-friendly besides can be easily available to people who wish to circulate or access information. On the other hand, specific skills and resources are required when publishing information via traditional media. However, both of these media channels are equally capable to reach small or large audiences. Thus, as made evident by the Arab Spring, the Internet is capable of causing destabilising effects on intellectual life in authoritarian regimes and can possibly pose a threat towards the downfall of the regime itself (Habermas, 2006).

This rapid growth of social media has influenced the lives of Malaysians in various aspects and one of the most important aspects is politics. The presence of social media has changed how the political stage is set in the country as how it is today (Muniandy and Muniandy, 2013). A standout feature of the social media is currently the use of the hashtag function in relations to certain topics, issues and events therefore this study hopes to undertake on the perceptions of Malaysians towards trending political hashtags on social media by using a Kansei analysis.

Literature review

Twitter and hashtags in political mobilisation: The category of political hashtag is a hashtag that often refers to a politically controversial topic (Romero et al., 2011). This can include a political figure political commentator, a political party or movement, a group on Twitter devoted...
to discuss a political cause, a location in the world that is the subject of controversial political discussion or a topic or issue that is the subject of controversial political discussion globally.

The Arab Spring Revolutions of 2011 have fostered a budding dialogue about the role of social media and networking as a tool for political mobilisation towards regime change and pro-democracy movements. Some political pundits, academics and journalists have embraced social media as an undeniable force for good, claiming that, “democracy is just a Tweet away” (Morozov, 2012). The internet has showed indication that it can be an effective weapon of the weak and marginalised against their authoritarian leaders. During the revolution, the most popular hashtags in the Arab region between January and March 2011 were #Egypt with 1.4 million mentions, #Jan25 with 1.2 million mentions and #Libya with 990,000 mentions. There are an estimated 131,204 Twitter users in Egypt between January 1st and March 30th 2011, generating an average of 24,000 Tweets a day during that time period. The Arab social media report tracked the volume of daily Tweets in Egypt throughout the time period of January 1st through February 28th with the peaks in Twitter use revolving around major events in the uprising, including the protests in Tunisia on January 14th and Hosni Mubarak stepping down on February 11th.

On a more recent analysis, the estimated number of active Twitter users in the Arab region at the end of June 2012 was 2,099,706. The estimated number of Tweets generated in the Arab region in March 2012 by “active users” was 172,511,590 Tweets. The estimated number of daily Tweets is 5,750,386 Tweets per day or 3993 Tweets a min or roughly 67 Tweets every sec. The most popular trending hashtags across the Arab region in March 2012 were #Bahrain with 2.8 million mentions in the Tweets generated during that period, followed by (Arabic for Syria) with 1.5 million mentions (Arabic for Bahrain) with 1.48 million mentions, #Syria with 1.3 million mentions) and #Egypt with 900,000 mentions and #Kuwait with 860,000 mentions (Mourtada and Salem, 2012).

Social media mining: Hashtag clustering: Social media is the “group of internet-based applications that build on the ideological and technological foundation of Web 2.0 and which allow the creation and exchange of user generated content” (Kaplan and Haenlein, 2010). Generally, social media could be considered as a world of social atoms (for example, individuals), entities (for example, content, sites, networks, etc.) and interactions between individuals and entities (Zafarani et al., 2014). In their study, Zafarani et al. (2014) collected information about individuals and entities, measure their interactions as well as discover patterns to understand human behaviour which could indicate the effective of social media mining. However, social media mining has to face challenges such as the big data paradox, obtaining sufficient samples, the noise removal fallacy and evaluation dilemma.

Social media mining represents the virtual world of social media in a computable way, measures it as well as design models which could help others to understand its interaction (Zafarani et al., 2014), for example, Twitter—a short-text social networking on the internet that grows continuously since its first introduction. With a large amount of data or micropost, there is a need for effective categorisation and search of data. Twitter-as one of the largest microblogging sites-allows users to utilise hashtags to categorise their posts. This pattern of categorising could also be distinguished as clustering in terms of unsupervised learning in the social media mining. Unsupervised learning is the unsupervised division of instances into groups of similar objects (Zafarani et al., 2014).

Antenucci et al. (2011) presented two techniques to aid in the retrieval of information from Twitter. In order to categorise the topic groups on Twitter, they indicated a technique to cluster hashtags by using a combination of co-occurrence frequency, graph clustering and textual similarity. They also indicated a technique to classify Tweets in terms of those topic groups based on their word contents and examined the relationship between the clustering step and classification step to evaluate the performance of each step.

As a result, they found that there are no rules of hashtags usage—they are used in a diverse ways and there are many unique, semantically similar hashtags but not structurally similar. Nonetheless (Antenucci et al., 2011) initially believed that they would be able to predict what certain user is going to tweet about next based on their previously used hashtags. Therefore, this study employed Antenucci et al. (2011) technique in clustering of hashtags. The study scrutinised that hashtag clustering could also be used independently to classify for example by inferring a user’s topic of interests (such as politics) from the clusters to which their most frequently used hashtags belong, users could be shown other Tweets user profiles or advertisements that correspond to those interests.

Internet democratisation in Malaysia: The impact of the internet in Malaysia has frequently been discussed, mainly because it represents a distinctive case of an administration that implemented a policy of
non-censorship of the internet as an effort to boost economic competitiveness of the country in addition to attract Foreign investment into the sector of information communications (Abbott et al., 2013). Yet, civil permissions and freedom of expression are strictly restricted. The result from this is the upsurge of an energetic and parallel, online public sphere alongside a generally controlled real-world civil society. The connection of these two contrasting public spheres provides an ideal environment to observe the significant impacts these two elements can make on a democratic institution or on regimes in which prohibit democratic freedom despite regular elections.

Over the last twenty years, there have been significant changes in regards to the communication and multimedia setting in Malaysia (Rahim and Pawanteh, 2011). The citizens now have broader selection and alternatives at their disposal to gain access to the informative materials that they desire from mass access information through broadcasting services. In addition, there are other alternatives of apparently independent news blog online such as Malaysiakini and Malaysian Insider that provide critical news and discussions on latest issues. Besides that, the trend of blogging is also on the rise and there are bloggers with political inclination that provide alternatives views and comments.

The massive impacts that online and social media play in the political discourse in Malaysia is reflected by the total internet penetration figures in Malaysia which has increased by 4,121,100 from 2008-2012. Facebook users in Malaysia have reached 13 million while Twitter users reached 2 million from the overall population of 29 million. Besides that, Malaysian youths are engaging heavily with social media. Studies illustrate that 64% of Facebook users in Malaysian are those aged between 18-34 while 62% of the total unique visitors of the internet are those aged between 15-34. Overall, social media assisted by key hashtags played a major role in mobilising rallies such as the Bersih rallies and in spreading anti-government materials by young people.

**Kansei and Kansei Engineering (KE):** Kansei is intended to express people’s feelings towards their surroundings or situations. As Kansei is derived purely from the Japanese culture, it is very challenging to define it into other languages. Kansei in general can be easily refers to sensitivity, sensibility, feeling and emotion (Yoshikawa, 2000; Ishihara et al., 2005 Nagamachi, 2003).

Kansei engineering is an approach initially applied in daily electronic appliances, computer systems, automobile manufacturing, cosmetics, apparel product, community design and so forth (Childs et al., 2003). While relatively Japanese based organisations are leading the way in product developments based on Kansei engineering, researchers, business practitioners and academic scholars from different parts of the world have shown interest towards the field (Nagamachi and Lokman, 2015). In addition, the cultural and indigenous traits that Kansei engineering has potentially allow it to be applied within other applications.

**MATERIALS AND METHODS**

**Identification of specimen:** This study illustrates the formation of political hashtags from social media websites that are relevant with the Malaysian citizens by using clustering of hashtag’s techniques, adopted from (Anteneh et al., 2011). This study data set is a collection of 9000 Tweets, from 858 unique users spanning from March -April 2014. The period chosen is a period identified to be the peak of 2014 Kajang by Election (pre and post-election) which the nomination day was held on March 11th 2014. Firstly, this study has performed several pre-processing steps to prepare the data for clustering, such as remove all Tweets containing non-ASCII characters in order to focus on malay language content. Then, this study removes the Tweets that contain no hashtags and also remove all non-alphanumeric characters. The final processing steps is to remove certain hashtags that are either auto-generated by external systems or are overly-general hashtags for example, those whose usage is too broad to consider their co-occurrence with other tags to be meaningful. From this pre-processed data, this study generates a list of the most frequently used hashtags, along with the co-occurrence frequency of all hashtags with each other.

In order to get the ideal specimen, this study also adopted the spectral clustering and MEITIS approach. For comparison purposes, this study performs spectral clustering and normalised spectral clustering. Even though these methods cluster over the 100 most used hashtags, the study has to maintain the co-occurring lists with each of these hashtags. As a result, even the hashtags which are not in the top 100 are then assigned to a cluster. Then, the result also been categorised into most and least significance for the study. Based on the time frame spanning from March 2014-April 2014 pre-election and post-election for 2014 Kajang by-Election, only 10 most significance political hashtags are analysed.

Tsui and Rappoport (2012) suggested these main categories which included character and words length, the hashtag orthography, hashtag location and emotional and
cognitive dimension. These attributes are extracted based on the hashtag alone, e.g., #TolakGST is 9 characters long, 2 words long, both words are listed lexical items, GST is a policy and the word “Tolak” represents a negative sentiment word. Approximately 50% of the hashtags selected were compounds of more than one word such as #KajangMove and #LawanTetapLawan. A word compound can make a hashtag or meme either clearer or too complex. Thus, this is more or less appealing to be adopted by users. Three bins were used for this feature: 1-3 words or more. Acronyms includes #GST and #BR1M were considered as one word.

For this study case, political hashtag has been selected as specimen. As political hashtags can be in abundance, the scope had to be narrowed down. Thus, the 2014 Kajang by-election was selected as a case study and timeline was determined which was 1 month before and after the by-election. In searching for the relevant hashtag, several search keywords that relates with the hashtags were used. These words were mainly referred from the official social media accounts of political parties and figures.

**Identified Kansei words:** Familiar, honest, boring, aggressive, confusing, presentable, impressive, reasonable, outstanding, original, straight-forward, interesting, emotional, appealing, strong, important, bad, effective, exciting, logical, excellent, positive, simple, deep, friendly, doubtful, fancy, complicated, symbolic, functional, radical, memorable, harmonious, passionate, ordinary, brilliant, offensive, generous, trustworthy, inspiring.

Additionally, the keywords were then used in the search function of both social media platforms to view opinions and hashtags from various users across the platforms. Facebook was not mainly used as a platform because the appearance of hashtags is not as extensive and evident as compared to Twitter and Instagram. As previously mentioned, the limited number of characters that Twitter allows for its posts tend to encourage users to apply hashtags functions more. This is a function not often seen being practiced on Facebook. Table 1. The social analytic tool, social mention illustrates the trends of hashtags while on Instagram allows viewing of Instagram images that contains the related political hashtags used. On a specific note, #MH370 was selected because despite not being directly related with a political issue, the issue was politicised but certain individuals and groups as it seemed to happen in the midst of the Kajang by-elections. Table 2 shows the 10 specimens that were selected for this study based on the tools and methods previously explained.

| Table 1: List of official social media accounts referred |
|---------------------------------|-----------------|-----------------|
| **Twitter**                      | **Instagram**   | **Keywords referred** |
| @NajibRazak                     | @najib_razak    | PKK Kajang       |
| @anwaribrahim                   | @anwaribrahim_my | Kajang by-election |
| @khairolKJ                      | @khairolKJ      | Kajang by-election |
| @gu_izzah                       | @gu_izzah       | Kajang Y25       |
| @bursainsarional                | @bursainsarional |                |

<table>
<thead>
<tr>
<th>Table 2: Identified specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hashtags</strong></td>
</tr>
<tr>
<td>#KajangMove</td>
</tr>
<tr>
<td>#BLACKOUTS05</td>
</tr>
<tr>
<td>#BR1M</td>
</tr>
<tr>
<td>#TolakGST</td>
</tr>
<tr>
<td>#AzizahKajang</td>
</tr>
<tr>
<td>#AlimahAllah</td>
</tr>
<tr>
<td>#LawanTetapLawan</td>
</tr>
<tr>
<td>#MH370</td>
</tr>
<tr>
<td>#Demokrasi</td>
</tr>
<tr>
<td>#Reformasi</td>
</tr>
</tbody>
</table>

**The instruments:** Kansei words are used in defining the perceptions in a selected product or research domain. The Kansei words used for this research were selected, according to Lokman and Kamaruddin (2010) study on Kansei affinity cluster that listed out Kansei words based on several different clusters. The technique was adopted from the KJ method of clustering descriptions into their respective categories.

The selected words should be related and able to describe the domain for the research. As this research domain is “political hashtags” and thus reflects on beliefs and emotions of the people. Therefore, adjectives such as “honest”, “radical”, “harmonious”, “offensive” etc. are best describing political perception and emotions of people. Besides that, it is proper to use common words that can be understood easily by all races in Malaysia. The adjectives must also represent both positive and negative adjectives as to have a balance response of perception from respondents. Finally, the selected words must be checked again to identify words that are possibly similar in meaning and may cause redundancy. Next, all forty sample words were displayed in random orders to reduce biasness by ambiguity. Besides, it will help indicate the level of understanding by the respondents. As for respondent selection, equal number of respondents was selected from each three major races in the country.

**RESULTS AND DISCUSSION**

**Clustering:** This study focused on clustering the top 100 most used hashtags. Also, this study also performed several trials on the 200 most frequently used hashtags without running the step but most of the hashtags ended up in one large cluster while the rest of the clusters consisted of single hashtags. Therefore, this study
Fig. 1: Structure of specimen and Kansei responses

believes that this was caused by noisy connections, which in inadvertently linked most of the hashtags closely together. This issue is removed and filtered.

In order to evaluate the clustering techniques, this study randomly selected 100 of the 200 clusters of political hashtags and score them manually on a scale of 1-5 with 1 representing a cluster that makes no sense and 5 being a perfect cluster. Factors that go into this rating include size of the cluster, number of matching hashtags that relate to the topics and relation of topics if more than one topic arises. As a result, only 10 hashtags remained after filtering, based on the significance of the study which focuses on political hashtags towards 2014 Kajang by-election.

**Sensitivity to specimen and emotion:** This study discusses the relationships between specimen and emotions that are formed from the results obtain by evaluation of all ethnic groups, named ethnic 1-3. The discussion is based on the calculated component vector which visualises the relationship between specimen and emotions over the first two components. The plotted vector enables the display of all specimen axes against the set of emotion axes in 2 dimensional spaces. Figure 1 shows an example of evidence that most specimens are very well distributed to all emotion vectors and have good influence to emotion. Besides that, the figure also shows that all specimens are well distributed along the emotion axes and the distinction of specimen could be observed clearly. This shows strong evidence that the evaluation subjects from ethnic 1 were able to respond well to the specimen and that the specimen prompted significant emotions to the ethnic group. Similar observation found in ethnic 3. However, different pattern is seen in ethnic 2. Most specimens are fairly distributed to all emotion vectors and have weak influence to emotion. It provides evidence that the evaluation subjects from the ethnic 3 were not so much able to respond well to the specimen and that the specimen evoked less significant emotions to the ethnic group. It is evident that the relationship between specimen and emotion could be clearly observed in ethnic 1 and 2. However, the same cannot be said to ethnic 2. It seems that the ethnic group did not respond well to the specimen. This is perhaps due to the lack of knowledge or interest towards the selected hashtags.

**The Kansei responses:** This section discusses the Kansei responses that formed from the evaluation result by the three ethnic groups. Figure 2 illustrates the strengths and influence of the two components to the structure of
emotion for an ethnic group. The emotions that produced large negative first component loading (x-axis) are “Offensive”, “Aggressive” and “Bad”. On the other hand, emotions that produced large positive loadings are “Friendly”, “Positive”, “Reasonable” and many more which represented by the dense area on the most right of the semantic space. It appears that this axis shows the dimension of “Acceptance”.

In the second component loading (y-axis), emotions that have large positive loading are “Emotional” and “Deep”. Emotions that have negative loadings are “Boring” and “Ordinary”. It appears that the axis shows the dimension of “Emotive”. The result indicates that Acceptance and Emotive are the determinants for a successful hashtag for ethnic 1.

It is then observed similarly that “Optimistic” and “Satisfaction” are determinants for a successful hashtag for ethnic 2. In contrast, “Acceptance” and “Satisfaction” are the determinants for a successful hashtag for ethnic 3. The above results put into view that evaluation results by the three ethnic groups produce somewhat similar determinants. Therefore, hashtag initiators can use this as a clue to create hashtag that can be strong and effective enough to stand out from the rest, hence, capture attention and influence the mind of the people in accepting the particular hashtag. Table 3 illustrates the result obtained which provides a clue on how certain characteristic of hashtags would result different kind of Kansel to the people.

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Kansel</th>
</tr>
</thead>
<tbody>
<tr>
<td>KajangMove</td>
<td>Original, symbolic, radical,</td>
</tr>
<tr>
<td>BLACKOUT505</td>
<td>Boring, offensive, dubtful,</td>
</tr>
<tr>
<td>LawanTetapLawan</td>
<td>complicated, aggressive,</td>
</tr>
<tr>
<td>TolakGST</td>
<td>confusing</td>
</tr>
<tr>
<td>BR1M</td>
<td>Logical, functional, appealing, generous</td>
</tr>
<tr>
<td>KalimahAllahMH370</td>
<td>Important, emotional, deep</td>
</tr>
</tbody>
</table>

It can be observed from the above results that the hashtags, “LawanTetapLawan”, “KajangMove”, “BLACKOUT505” and “TolakGST” produced the most feedbacks and mostly tilted towards negative perception. On the other hand, “BR1M” produced feedbacks that were well accepted by the people. “kalimahAllah” and “MH370” produced a more emotional and sensitive feedback.

The result, however is findings based on the limited scope of the research instrumentation. Therefore, the findings are not to be used as generalised findings to all hashtags and for the whole Malaysia’s population.
CONCLUSION

Political hashtags studied were structured by four factors—pleasant, heartfelt, ordinary and familiar. Therefore, these four factors are likely to have strong influences in an individual’s state of mind that need to be considered in designing powerful yet strong enough to win the hearts and minds of the people in relation to political hashtags. This study basically has produced Kansei analysis of political hashtags which hashtags namely KajangMove, Blackout505, LawanTetapLawan and TolakGST produced the most feedbacks and they were mostly tilted towards negative perceptions—radical, boring, offensive, doubtful, complicated, aggressive and confusing. BRIM has produced feedbacks that were well accepted by the people which they perceived the hashtag as logical, functional, appealing and generous. However, KahlimahAllah and MH370 produced moving and susceptible feedback which the respondents believe they were important, deep and emotional.

Generally, this research managed to provide an idea on the perception of Malaysians towards various political hashtags that decorate Malaysia’s political scene. Thus, it is up to the respective political parties through their strategic communications divisions to develop a well-structured political hashtags in order for their agenda-setting purposes to be effective and well accepted by the citizens, instead of simply making up a single word with the hope that it can be an instant hit for political campaigns. Besides that in terms of academic it is recommended that this study should be expanded into deeper perspective possibly until a solid model or framework on hashtag development can be developed as this may prove to be a useful finding not only in the academia but also in politics and commercial use.

ACKNOWLEDGEMENTS

The researchers would like to thank RIG KAE, UiTM as well as Malaysia Association of Kansei Engineering (MAKE) for all their supports. The study is partially supported by Research Management Institute, Universiti Teknologi MARA, Malaysia under the Research Entity Initiative Grant Scheme (Project Code: 6004-RMI/DANA S3/REI(10/2017)). The Researchers wish to thank Muhammad Danial Idris for his initial investigations.

REFERENCES


