How is my Event Rated? Rating an Event Using Social Media Data

Suman Silwal[1] and Dale W. Callahan[2]
[1,2]The University of Alabama at Birmingham (UAB), USA

***Abstract*— In recent years, events are continuously discussed on Social Media in the form of status updates, posts, discussions and comments by its participants, volunteers, and supporters. Social media content generated before, during, and after an event could add valuable insight into success and popularity of an event. It can also generate ideas for future improvement of the event. With the fast evolving nature of Social Media, current events' Social Media content is ignored, forgotten, and overlooked for new sets of future posts, discussions, and comments. In our research, we believe that any publically available Social Media data can be captured and analyzed to produce some meaningful information. In this research, we are building a rating system through a combination of multiple sentiment analysis models using SM data.**

***Keywords*– Social Media (SM), Social Networking Site (SNS), Rating System, Sentiment Analysis, Marathon, Twitter, Hashtag (#), Marathon Rating and Event Rating**

I. INTRODUCTION

Social Media (SM) data are "unstructured, informal, and fast-evolving" [1] in nature. In recent years, more and more people are sharing their thoughts, feelings, sentiments, etc. on SM about events, products and services [2]. As the growth of SM uses are happening around us, so is the interest on research and development to utilize these SM data [2], [3], [4].

In recent years, many different groups developed Natural Language Process (NPL) tools such as the Tweet NLP (http://www.ark.cs.cmu.edu/TweetNLP/) and the Stanford NLP (http://nlp.stanford.edu/) to understand sentiment of a SM post. The Tweet NLP uses tokenizer, clustering, and part-of-speech tagging approaches for Twitter data [5]. Even though it takes some effort to understand and get meaningful results using SM data, there are still a lot of unanswered questions regarding the findings of SM sentiment using existing sentiment tools and NLP [6], [7].

Social Networking Sites (SNS) have a lot of opinion spam [8] and fake opinions [9]. Due to the unstructured nature of SM data, finding quality user-generated content [10] from SM posts is always a challenge. Even with a data set that is filtered and domain specific, understanding and producing meaningful information by processing an individual SM post through a computer program provides added challenges.

A single SM post can consist of words, abbreviations, numbers, hashtags, images, mentions, links, special symbols, emoticons, etc.; furthermore, we believe that this information can provide insight into understanding the overall sentiment of a SM post.

In our previous research [2], [11], we built a foundation and road maps where we discussed ways not only to understand positive, negative and neutral sentiments of an event's SM posts but also to create a 1-5 rating system. So in this research, we are building multiple modeling techniques to capture different aspects of SM posts to achieve our overall goal of creating a rating system using SM data. We are building "systems of systems" [12] using a model of models in an interdisciplinary manner to capture and analyze SM data to produce some meaningful information. Our final outcome of this research is to build a numeric rating system of an event using SM data as well as compare and validate those output data.

Here are some of the core systems, events, and frameworks that are used to build our user rating models and processes:

- ***Twitter.com*** (Twitter) is used as the main SNS for our research. It provides its developer network limited access to its publically available data through its Application Programming Interface (API) [13].
- ***Marathon*** events are 26.2 miles foot races that are considered as our research event topic.
- ***MySql database system*** is a widely used open-source relational database management system (RDBMS) that provides different tools to access and manage the database.
- ***Java programming language*** is an open-source computer programming language, which is widely used for application development.
- ***Spring*** java framework provides a lot of different components to build a very powerful application, including Spring Social API. It is used to glue our application together.

In the future sections of this research paper, we will break our SM user rating building process into 3 different parts:

1) ***Data importing*** consists of importing and inserting SM data into a local database.

2) ***Sentiment dictionary*** building consists of creating processes to generate sentiment dictionaries.
3) ***Sentiment Modeling*** defines different possible modeling techniques needed to build a rating system.
   a. During testing, we will build a process to generate numeric rating SM data.
   b. In the results section, we will review results generated by our testing process.
   c. Within the validation section, we will compare human rating from results.
   d. In the discussion section, we will explain the relevance of our research model.

II. SM RESEARCH DATA COLLECTION PROCESS

Collecting valid sets of data plays an important role in the success of any research, including ours. Any open forum on the Internet can contain noise and misleading information [14]. In our research, we believe that it is important to find, filter and collect only necessary data to avoid overloading of unnecessary and excessive data.

Once a developer's access account is set up with the Twitter SNS (https://dev.twitter.com/), we request to set up consumer key, consumer secret, access token, and access token secret for authentication and authorization to its API.

During the data import process, once a valid handshake is made through Twitter's authentication APIs, our data import process gets access to Twitter's dataset. By default, the Spring Social (http://projects.spring.io/spring-social/) search API for Twitter can retrieve up to 50 of the most recent matching tweets per call. Also, Twitter allows only 180 requests/queries per 15 minutes to its API per run.

Our research is mainly based on yearly marathon events, which are heavily discussed close to the actual race day. To prevent accessing irrelevant Twitter data, we created a search look up table using search criteria (table 1) with active status. Our Twitter data collection process (figure 1) runs almost in real time; as a result, this process imports data into our SM warehouse database table.

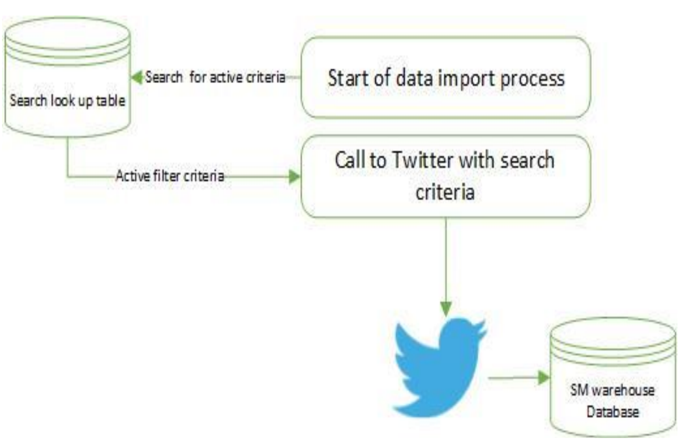

Figure 1: Twitter data import

Once we hand our search lookup criteria to our import process, most of the heavy lifting to retrieve proper Twitter data is done within Twitter search API.

Table 1: Some of Search Criteria

Search Lookup Criteria
#mercedesmarathon
@Run_Mercedes
#BostonMarathon2014
#ChicagoMarathon

At this time of our research, we are capturing Tweet Id, Tweets Text, Generated from user, Tweet created date, and Retweet count information from each Twitter per API call.

Imported tweets are stored in our local warehouse database table for future uses (Table 2). As we bring more marathons into our rating mix, this list of data is bound to grow.

Table 2: Some of the marathon's data collection counts

Event Names	Total Count
Boston Marathon	168357
Country Music Marathon	2997
Flying Pig	1884
Marine Corps Marathon	3349
OK City Marathon	1526
Richmond Marathon	305
St. Jude Marathon	1257

III. DICTIONARY BUILDING

Building a valid dictionary is a very important part of our research. A dictionary gives us an advantage in creating a structure around unstructured SM data. Each word on the dictionary table can have multiple attributes such as sentiment, trending count, weight, etc. to help us understand more about each word.

Furthermore, each word in our dictionary can also be clustered into positive (P), negative (N), neutral (NU) or not applicable (NA) sentiment category.

A. Initial sentiment dictionary building process

Initially, we imported predefined sentiment words from different websites into the sentiment dictionary table (Fig. 2). This provided us with a good set of data to start with predefined values.

Since our research is based on specific Twitter data and marathon running events, these initial sentiments were not sufficient. Therefore, additional words were added to the dictionary using the sentiment dictionary building process.

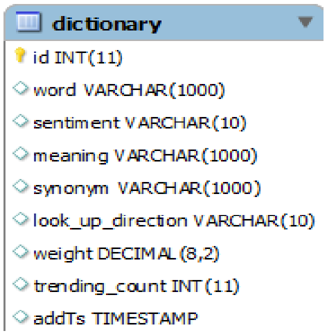

Figure 2: Dictionary table

B. SM data sentiment dictionary building process

In this sentiment dictionary building process (Fig. 3), at first, a call is made to the Twitter data warehouse table. Then, each of the resulting tweet posts are split into multiple word rows and stored into a temporary dictionary table. For this process, each word with a special character, symbol, link, numeric value, emoticon, etc. is ignored. A valid and unique word from a tweet post is inserted into a dictionary database table for future use.

At the time of writing this paper, each word on this dictionary is manually clustered into one of the default sentiment categories (Table 3).

Table 3: Default sentiment indicators

Sentiment Indicators	Descriptions
P	Positive
N	Negative
NU	Neutral
NA	Not Applicable

Even though it is a laborious process to create a word-based dictionary, we believe that this gives us control over how each word is perceived and evaluated without knowing the full context of a sentence. In this approach, each sentiment is defined purely on a word level. Table 4 lists some pros and cons of creating a domain specific lexicon.

Table 4: Pros and cons of creating a sentiment dictionary

Pros	Cons
Quickly build dictionary words	Have to look for words
Domain specific word	Getting unnecessary words into database table
Ability to expand attributes to understand a word	Manually enter dictionary
Clustering words to different categories	Misleading sentiment by just looking one word
Ability to create structure around words	
Reusability	
Grouping SM shorthanded words to real words	
Search ability	
Reusability	

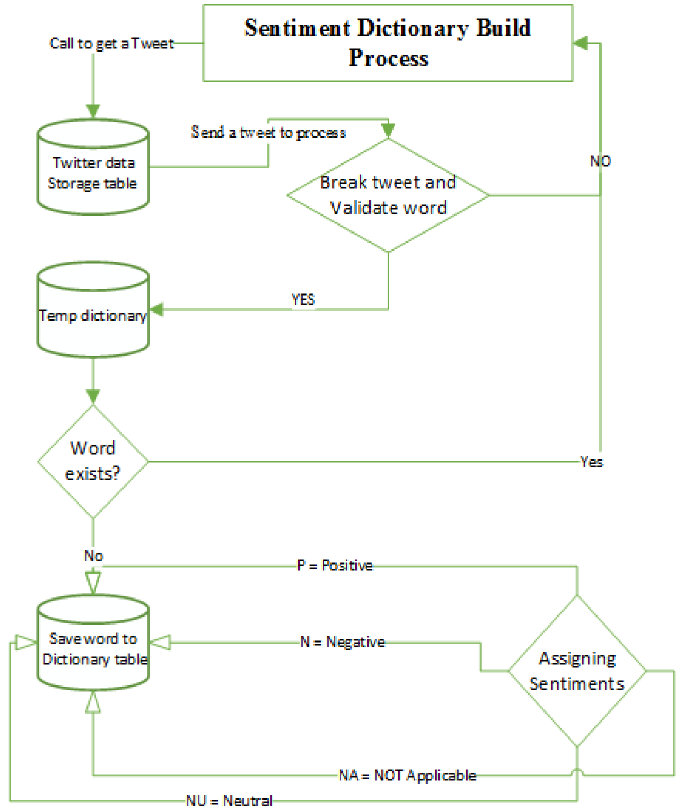

Figure 3: Dictionary data building process

IV. BUILDING A SM RATING MODEL USING SM DATA

Building a marathon event rating model using SM data in an interdisciplinary manner is core to our research. In our research, we believe that any piece of publically available SM data can be captured and analyzed to produce some meaningful information. An ultimate outcome of this research is to build a numeric rating model through a combination of multiple sentiment analysis models using SM data.

A SM API such as Twitter gives access to different types of data sets, including geo location, add timestamp, tweet id, text, etc. For our research, we are generally interested in text data of a SM post. In recent years, there has been a lot of interest in the study of SM data to find social interactions, emotion [15], election approval rating [4], etc. At this time of research, this area of processing SM data to create a numeric rating system is still a new field of interest.

SM users' rating model is built in the simple idea of an input-process-output model (Fig. 4), where input data is retrieved from a SNS data source. Those SNS data are processed to produce some meaningful information.

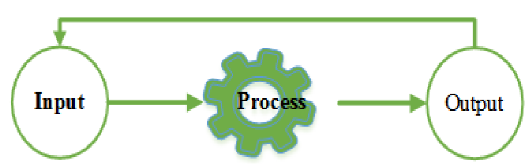

Figure 4: Input-process-out model

In this section, we will present multiple rating models to reflect a possible sentiment of a SM post. Eventually, each of these models is put together to create a unified rating model that contains a model of models. An example of a tweet post (Fig. 5) that consists of many different aspects of a SM post we will try to capture and understand in future sections of this paper.

Figure 5: MarathonRuns's tweet before Houston Marathon

A. *Word-by-word sentiment model*

The word-by-word sentiment model is the first of many sentiment models that we will build for this research. For this rating model, each word of a SM post is qualified to be reviewed for sentiment analysis. A single SM post can consist of many different types of words, abbreviations, numbers, hashtags, images, mentions, links, symbols, emoticons, etc. For this modeling, we ignore punctuation, reference to images, URLs, numbers, emotions, etc.

In this model, each SM post is sliced into multiple words. Each of these words is clustered into either the positive, negative, natural or not applicable sentiment category. Having these words clustered into 4 different sentiment categories give us a little sense of structure around the unstructured nature of a SM post.

The design of this model is dependent on the accuracy of each word's sentiment in obtaining the overall sentiment of an entire SM post. Eventually, each SM word's sentiment rating will produce an overall rating for an event associated with that SM post. Figure 6 shows what word-by-word rating looks like in a bigger picture.

Table 5: Default sentiment indicators numeric value

Sentiment Indicators	Descriptions	Numeric Rating
P	Positive	5
N	Negative	1
NU	Neutral	2.5
NA	Not Applicable	0

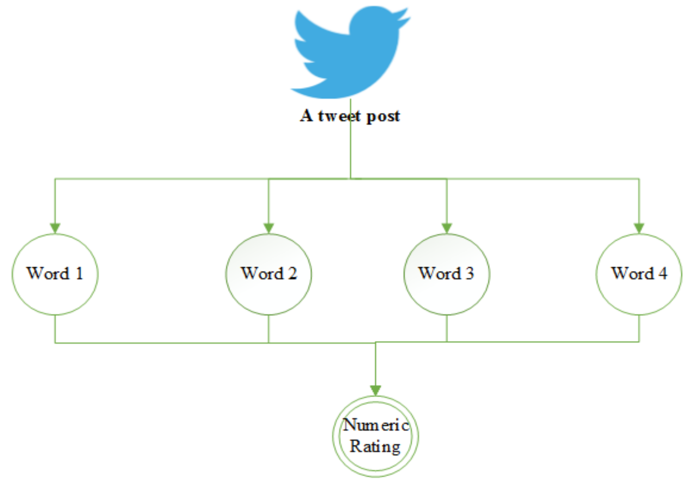

Figure 6: Word-by-word rating – a bigger picture

Table 6 shows how a SM post from Fig. 1 is broken into different words and sentiment categories.

Table 6: Word-by-word sentiment

Word	Sentiment
Good	Positive
Luck	Positive
To	Not Applicable
Racing	Natural
#HouMarathon	Natural
@HoustonMarathon	Not Applicable
!	Positive
Gr8t	Positive
Weather	Natural
www.weather.com/weather/weekend/l/	Not Applicable
Go	Positive
Get	Natural
Finisher	Positive
Medal	Positive
:)	Not Applicable

Each of these sentiment indicators from column 2 of Table 5 is associated with a numeric value. Table 5 shows a list of sentiment indicators and the assigned default numeric value for this model. For our research, those values defined on Table 5 are considered as default sentiment indicator values for all our current and future models. Since our numeric rating is based on a 1-5 rating system, we feel that this is a standard constant value for each sentiment indicator parameter. Also, it gives us a structure and consistent look of SM data.

For a single SM post, the sum of sentiment indicators is multiplied by each numeric rating value associated with it. The sum of these values is then divided by the sum of the total sentiment. The following formula (1) provides a numeric rating for a SM post:

Numeric rating using word-by-word sentiment

$$= \frac{\sum(\boldsymbol{P})\times\mathbf{5}+\sum(\boldsymbol{N})\times\mathbf{1}+\sum(\boldsymbol{NU})\times\mathbf{2.5}}{\sum \boldsymbol{P}+\sum \boldsymbol{N}+\sum \boldsymbol{NU}} \quad (1)$$

Based on this formula, the numeric rating for Fig. 5's SM post using the word-by-word rating model is 4.09. This result is very close to numeric rating compared to manual rating.

Due to the unstructured nature of SM data, the word-by-word sentiment model provides a great benefit in understanding SM posts through breaking each word into small units of its own. It can provide some insight into users' sentiments.

1) Testing

To test the word-by-word rating model, we built a testing process model. Initially, our process (figure 7) retrieves a single tweet post from our warehouse table and splits it into multiple words. Each of these valid words is searched for in the sentiment dictionary table to find an associated positive, negative, neutral, and not applicable sentiment category. Fig. 7 shows the core logic that we used for collecting different sentiments of each word.

A tally is kept for each tweet word's sentiment category assignment count and numeric value associated with them. At the end of processing each tweet post, we use the formula (1) defined by the word-by-word rating model to find the numeric rating of a tweet. These values are stored in the Twitter warehouse database table field for future calculation.

2) Results

Table 7 shows the final results of the word-by-word rating model after processing three different events' rating results. Based on an initial observation of our results, each of these events are getting positive ratings.

Table 7: Word-by-word rating model results

Event Name	Word-by-word rating
Boston Marathon	3.6161
Richmond Marathon	3.7403
Twin Cities Marathon	3.5991

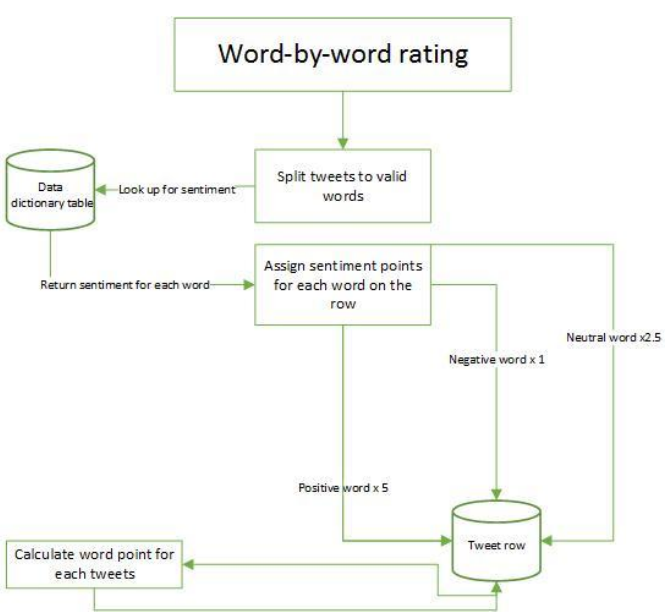

Figure 7: Word-by-word sentiment rating process

3) Validation

Validating our results from our rating system is an important part of our research. Since, we are trying to create a rating model using a computer process, there will always be misunderstanding between human speech vs. what our computer process translates it into.

By comparing the overall results of the human rating and the word-by-word rating, we acquired the following results (Table 8). By keeping human rating as our standard rating, we get less than 4% difference (2) between the word-by-word rating and the human rating.

$$\text{Difference \%} = \left(\frac{\textit{Absolute Value of }(\textit{Word by word rating} - \textit{Human rating})}{\textit{Word by word rating} + \textit{Human rating}}\right) \text{X } 100 \qquad (2)$$

Table 8: Comparing results

Event Name	Word-by-word rating	Human rating	Difference %
Boston Marathon	3.8258	3.5549	3.67%
Twin Cities Marathon	3.5991	3.5870	0.17%
Richmond Marathon	3.7403	3.7637	0.31%

4) Discussion

Due to the unstructured nature of SM data, the word-by-word sentiment model provides a great benefit of understanding SM posts through breaking each word into small units of its own. It also provides some insight into users' sentiments, but it does not provide all the answers. As cumulative data comparisons, the biggest difference between human rating and our model rating is 3.67%, which is within our confidence level of 10%. This is great news for our rating model.

After further reviewing line by line results from the Boston Marathon, we found that there are more than 72% of tweets with more than 10% difference in values. Some tweets were rated higher by human rating, while other tweets were rated higher by our rating process (Table 9). This may be due to our process only looking at one word at a time to make sense of the whole sentence, while human rating is looking at the whole context of a tweet.

Table 9: Example tweets

Tweet	Word-by-word rating	Human Rating
You don't get it, do you? It's not about winning--it's about participating. #BostonMarathon @JoeyDips	2.40	4.6
I could not imagine running 5 min miles for 26 miles #BostonMarathon	2.70	4.2

Ultimately, we believe that the word-by-word model is still a valid rating model as an initial model, but we feel that it is not sufficient enough to look at one word at a time to build an overall model of a SM post. So we decided to build a multi-words association sentiment modeling.

B. Multi-words association sentiment modeling

Within the multi-words association sentiment model, we look at sets of words to make sentiment analysis decisions. Unlike the word-by-word rating model, in this model, not every word is qualified for bi-direction look up. For those qualified words, we put a directional indicator on each word so that it can be viewed from a specific direction: forward, backward, or both sides of a word (Fig. 8).

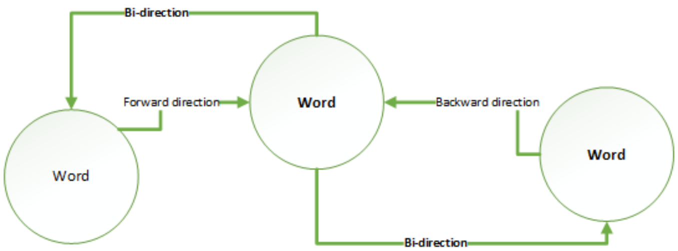

Figure 8: Multi-words association model

A multi-word association matrix (Table 10) is created to identify sentiment for multi-direction look up words. In this matrix, we assume that a negative word generates negative sentiment for an associated word regardless of whether it has a positive or neutral sentiment; also, a positive word with neutral sentiment will generate a positive sentiment.

Table 10: Multi-word sentiment matrix

	Positive (P)	Negative (N)	Neutral (NU)
Positive (P)	P	N	P
Negative (N)	N	N	N
Neutral (NU)	P	N	NU

Table 11 shows how each of these associated sentiment words generate positive or negative sentiment using multi-word association model.

Table 11: Multi-word association sentiment examples

Sentence	First lookup Word	Associated Word	Sentiment
Good Luck today!	Good	Luck	Positive
Good grief run fast.	Good	Grief	Negative
Love my Finisher Medal.	Finisher	Medal	Positive
Gr8t weather	Gr8t	Weather	Positive
I do not like this race	Not	Like	Negative

Similar to the word-by-word sentiment model, a default numeric value of 1-5 is assigned to each word and its associated word(s) with similar calculations. For those words that do not have an associated sentiment, it follows the word-by-word sentiment model to assign sentiment values. For our example SM post (Fig. 5), we received around a 4.16 rating using the multi-word sentiment model due to two positive words' association "Finisher" and "Gr8t" with other words.

1) Testing

To test our model, a multi-word association process is built to look up more than one word to obtain a more in-depth sentiment rating of a tweet post. To properly tag this new rating model, we manually added one more attribute to the dictionary table to indicate looking forward (F), backward (BK) or at both (B) directions of a word (Table 12).

Table 12: Word with sentiment and multi-direction look up

Word	Sentiment value	Look up direction
quick	P	F
lit	P	B
can't	N	F
looking	P	F
having	P	F
!	P	BK
cross	P	F

In this process, a complete sentence "I do not like this run" has a bi-direction word "not," which looks at both sides of the word. Even if "do" and "like" are two positive words with sentiment value of 5, in this process the word "not" looks at both directions, which creates negative outcomes for words "do" and "like."

This process works very similar to the word-by-word rating process that was described in the previous section but with the added difference of look up in the multi-words association sentiment matrix (Table 10). Every word in the dictionary table does not have bi-direction look up indicator. For those words without bi-direction indicator, they are treated as a word-by-word rating process. In general, this process works similar to the word-by-word rating model with multi-direction look up attributes.

2) Results

At a glance when we compared word-by-word rating and multi-word rating (Table 13), we were able to find improvement in most of the numeric rating.

Table 13: Multi word rating

Tweets	Word-by-word Rating	Multi word rating
Congrats to @dianamchard for finishing the #BostonMarathon !	4.38	5.00
Amazing that an American man won the #BostonMarathon. #StorybookEnding	4.00	5.00
.@TylerPennel just won the @tcmarathon. His. First. Marathon. That's wild. #tcmarathon	3.33	5.00
Meb Wins Boston: Amazing things happen. Never stop believing. #BostonMarathon http://t.co/eFCnF4KQCC via @Flotrack	4.29	4.64

After comparing cumulative results from word-by-word rating and multi-word rating process results (Table 14), we found very little difference on overall rating results.

Table 14: Comparing results from word-by-word and multi-word process results

Event Name	Word-by-word	Multi-word
Boston Marathon	3.6161	3.6083
Twin Cities Marathon	3.5991	3.5919

3) Validation

To validate our results from the multi-word association process, we took the sum of the results retrieved from the sample data and compared them against human rating (Table 15). We found that the Boston Marathon SM rating is still higher than human rating, while Twin Cities Marathon's rating shows a consistent look.

Table 15: Validation after multi-word association process run

Event Name	Multi-word rating	Human Rating
Boston Marathon	3.8242	3.5549
Twin Cities Marathon	3.5991	3.5870

To further validate Boston Marathon's rating, we took another sample set of the marathon data, which were rated by the multi-word rating process and were not part of our previous sample collections. Those sample data were sent for further validation. After taking the average from these new data sets, we were able to compare our results (Table 16). Now, we are able to see the overall Boston Marathon rating drop by .6185 from the prior overall rating as well as a 0.2861 difference between sample data used by multi-word process and human ratings.

Table 16: Validation after multi-word re-process run

Event Name	Multi-word rating	Human Rating
Boston Marathon	3.2057	3.4918

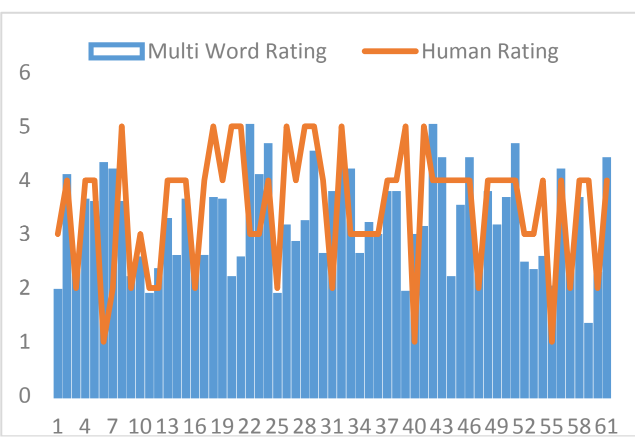

Figure 9: Sample of multi-word rating vs. human rating

4) Discussion

Looking at the sample set of data side by side (Fig. 9), we can see that a lot of data from human rating are very closely rated. Even though there was a big drop in numbers for overall rating using sample data, we feel that these were expected results for us due to multi-words rating.

We believe that the multi-word association process is a great addition to our overall look up of the SM rating process. At this time, we are using a limited amount of words for this multi-direction look up. As we add more words to find multi-directions, we feel that we will be able to get better results.

Even though multi-words give a lot more insight into a SM post and its sentiment, we are looking further into analyzing a SM post. In the next section, we will talk about the word weight factors sentiment model.

C. Word weight factors sentiment model

In the word weight factors sentiment model, we believe that any word may or may not have the same static numeric value even though it may be a synonym and have the same default sentiment category value (Table 5). So, in this model, each word is manually assigned a word weight numeric sentiment value; for example, "*good*" and "*great*" both are positive sentiment words, but the word "*great*" can be given a higher word weight than the word "*good*."

Table 17: Word weight chart

Word	Sentiment	Word weight
excited	P	5.00
good	P	4.50
great	P	5.00
please	P	3.00
join	P	3.50
us	NU	2.50
praying	P	4.00
injured	N	1.00

A SM post could have words with different word weight strength value. So, each of the numeric values associated with a word from a single SM post is summed together and divided by the sum of words to generate a SM rating. The following formula (3) shows how the word weight sentiment is calculated:

Numeric rating using word weight factor

$$= \frac{\sum(\boldsymbol{word}\times\boldsymbol{weight})}{\sum \boldsymbol{words}} \qquad (3)$$

Even though assigning weight to every word may be a difficult task, this model adds a different dimension to our SM rating models overall. With this model, we do not have to depend on static weight sentiment value. Table 18 gives a snap-shot of how a word weight range could look, where a positive word could have any numeric value from 3.5 to 5, while a negative word could have any numeric value from 1 to less than 2.5.

Table 18: Word weight range

Sentiment	Range
Positive (P)	3.5 - 5
Negative (N)	1 – less than 2.5
Neutral (NU)	2.5 to less than 3.5

1) *Test*

In this rating process, each word of a tweet post is viewed against its predefined word weight. Every word that gets processed has word weight associated with it (Table 17). Some words have greater strength than others. If predefined words are not found, we use a default sentiment process to find a word weight.

2) *Results*

Since each word could have potentially different weights, it is possible to get different results compared to previous rating models that we have used so far. In Fig. 10, we compared 12 sample results after running the word weight sentiment process. As we can tell from these sample results, some of the tweets' ratings improved, while most of these tweet rating results from the word weight model went down in numeric value.

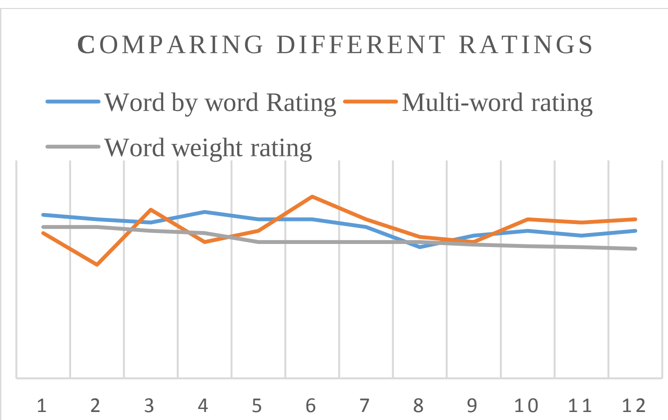

Figure 10: Comparing different rating results

Finally, Table 19 shows cumulative word weight rating. Due to a decrease in numeric rating on each tweet, there is also an overall downward rating for each event. Since our rating looks at the rating of individual word weight, it is important to have an accurate word weight rating for each word. To improve our results for word rating, we can look into further modifying the word weigh-in value of each word.

Table 19: After word weight

Marathon Name	Word-by-word	Multi-word	Word-weight rating
Boston Marathon	3.6161	3.6083	3.1980
Twin Cities Marathon	3.5991	3.5919	3.2129

3) *Validation*

After comparing results from the multi-word rating (Table 20), we found that Boston Marathon's rating is still higher for multi-word rating than human rating, but we noticed that the Twin Cities Marathon's numeric rating has gone down. We felt that the reason for a higher Boston Marathon rating may be due to our process not having enough sample data associated with word weight rating.

Table 20: Validation after word weight process run

Event Name	Word-weight rating	Human Rating	Difference %
Boston Marathon	3.1980	3.5549	5.2%
Twin Cities Marathon	3.2129	3.5870	5.5%

4) *Discussion*

After looking at the detailed results from human and word weight ratings, we can see that absolute difference is still less than the 10% range. We believe that our overall result matches well with expected results, but we are still finding a lot of discrepancy for line-by-line rating.

D. *Unified SM user rating model*

The SM user rating model is our final and core model for our research, which consists of a model of models. So far we have talked about three different ways to look at a single SM post. Even though each model has different ways to look at SM data, each model is built with a vision to create a single unified (Fig. 11) SM user rating model, where our previous three models work together to contribute to build an aggregate rating of an event.

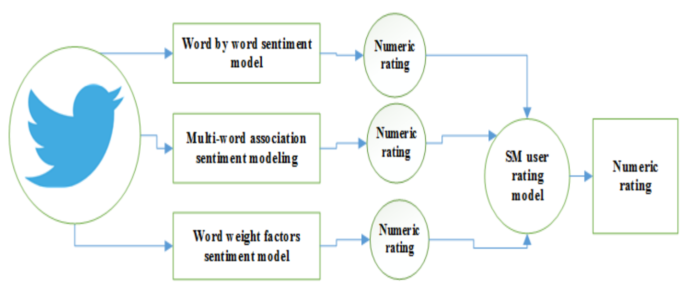

Figure 11: Unified model

Since most of the hard work is done by the previous models, in this model we take an average sum of each SM post's numeric value that was generated from the previous three models using the following formula (4):

$$\textit{SM user rating model} = \frac{\sum(\textbf{\textit{Word by word rating+multi word association rating +word weight factors rating}})}{\sum(\textbf{\textit{Number of sentiment model used}})} \qquad (4)$$

1) *Test*

SM user rating is our core and final processing model where we bring different modeling data together to create a unified model for an event using Twitter data. As of writing this research paper, this model consists of 3 different processing models to review a SM post where we sliced and diced a SM post into 3 different dimensions to find its rating. As described in the previous sections, each of these models have their own way of looking at a SM post. Since most of the work was done by the previous processes, in this process we take the average sums of word-by-word rating, multi-word rating, and word weight rating results to get the final rating of each tweet post. Finally, the average sum of all SM posts from each event creates a final SM user rating for an event.

2) *Results*

In these results (Table 20), we are taking the average sum of 3 different events and producing final results.

Table 20: SM rating model

Event Name	Rating Models used	SM user rating
Boston Marathon	3	3.4742
Twin Cities Marathon	3	3.4680

3) *Validation*

For our validation, we are able to see that our cumulative rating for both the Boston and Twin Cities Marathon are less than 2% (Table 20). This is much less than the 10% margin. So far, our modeling technique looks promising.

Table 20: SM rating model

Event Name	SM user rating	Human Rating	Difference in %
Boston Marathon	3.4742	3.5550	1.15%
Twin Cities Marathon	3.4680	3.5870	1.69%

4) *Discussion*

After further reviewing the Boston Marathon's line-by-line items with the overall rating, we are still seeing that more than 47% of data is above our standard 10% threshold between human rating and the combination of our process rating. Even though it is a drop from our previous rating, still these are very high percent values that do not match.

We feel that our rating models are getting better as we improve our processes and word sentiment categories as well as develop new ways to review SM data.

V. CONCLUSION

After creating three models to break SM posts into small units of words, multi-words, and word weight to understand sentiment of a SM post, we believe that our rating model result is getting closer to providing an understanding of the sentiment of a SM post by using computer-generated processes. In our approach, we looked at SM data beyond current trend and social experience.

Furthermore, we also believe that our research is built in a truly interdisciplinary manner to connect the multidiscipline of big data computation processing, social networking, sports, event, linguistic, etc. As these models and processes mature, this idea of event rating can run almost in real time manner to rate an event.

Due to many known and unknown variables, there will always be misunderstandings regarding true human sentiments vs. computer-analyzed sentiment of a SM post. During our testing and validating of a SM post, we realized that an individual SM post rating may vary between human rating and process rating, but we are able to create overall SM rating results within our standard threshold. We believe that this is due to our ways of looking at a SM post using multiple modeling techniques and dimensions.

In this research, we are successfully able to create a model of models to capture and analyze SM data to produce meaningful information. Initially, we were able to achieve our goal of producing a numeric SM user rating utilizing SM data by rating two different events (Table 20).

Despite our successes, we feel that there is still a lot of work remaining to complete our model of models. We believe that our concept is not invalid, but that it needs some more improvements.

VI. FUTURE WORKS

In the future, we will be looking at ways to bring in other part of a SM post, such as hashtags, emoticons, images, etc., to complete our user sentiment model. Also to further validate our processes, we are planning to evaluate more marathon events.

Since word sentiment category and word weight are an important part of our overall SM rating models and processes, we need to find ways to automate weighting words, categorizing sentiment, creating rating matrix, and finding more predefined words with associated sentiments.

VII. APPENDIX

A. *Appendix I*

For our validation process, we sent sample sets of multiple data to 10 different individuals with combined experience of more than 50 years of social media and more than 50 years of running. Even though these are not the same people who posted these SM posts, we felt that having humans to review helped to validate our results. Each individual was asked to read each sample Twitter post and rate them 1-5 according to sentiment towards a marathon.

REFERENCES

[1] X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised Sentiment Analysis with Emotional Signals," pp. 607–617.

[2] S. Silwal, "Using Social Media Data as Research Data," vol. 1, pp. 49–55, 2013.

[3] C. Wojtalewicz, "Social Media Use for Large Event Management," no. 1, pp. 24–29, 2012.

[4] D. Contractor and T. Faruquie, "Understanding election candidate approval ratings using social media data," *... 22nd Int. Conf. ...*, pp. 189–190, 2013.

[5] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, and N. Schneider, "Part-of-Speech Tagging for Twitter: Word Clusters and Other Advances," 2012.

[6] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter Sentiment Analysis : The Good the Bad and the OMG !," *Artif. Intell.*, pp. 538–541, 2011.

[7] T. Baldwin, "Social Media : Friend or Foe of Natural Language Processing ?," pp. 58–59, 2012.

[8] B. Liu, "Sentiment analysis and subjectivity," *Handb. Nat. Lang. Process.*, pp. 1–38, 2010.

[9] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, May 2012.

[10] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," *Proc. Int. Conf. Web search web data Min. - WSDM '08*, p. 183, 2008.

[11] S. Silwal and D. W. Callahan, "Building a Social Media rating model," *Ieee Southeastcon 2014*, pp. 1–3, Mar. 2014.

[12] A. Odusd and T. Sse, *Systems Engineering Guide for Systems of Systems*, no. August. 2008.

[13] J. Weng, E.-P. Lim, Q. He, and C. W.-K. Leung, "What Do People Want in Microblogs? Measuring Interestingness of Hashtags in Twitter," *2010 IEEE Int. Conf. Data Min.*, pp. 1121–1126, Dec. 2010.

[14] T. Ahlqvist, A. Bäck, M. Halonen, and S. Heinonen, "Social media roadmaps," *Helsinki Ed. Prima Oy*, 2008.

[15] F. Kivran-swaine and M. Naaman, "Network Properties and Social Sharing of Emotions in Social Awareness Streams," *Media*, no. July, pp. 379–382, 2009.