International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 3, May – Jun 2016

RESEARCH ARTICLE OPEN ACCESS

Social Media Ways Trade Market of Opinions Transaction Twitter Analysis (Time)

S.Nafeesa Begam
M.Sc., M.Phil (CS)
Assistant Professor
Aiman College of Arts & Science for Women, Trichy-21.
Tamil Nadu – India

ABSTRACT

Time-honored association rule mining and frequent Twitter message cannot meet the demands arising from some real applications. By considering the different types user of individual items as utilities, utility mining focuses on identifying the itemsets with high utilities. Recently, high utility Twitter mesages is one of the most important research issues in data mining. In this paper, present a utility mining using an algorithm named Generalized Linear Models anothername called Simple High Utility Algorithm (SHU-Algorithm) to efficiently prune down the number of candidates and to obtain the complete set of high utility itemsets. If we are given a large transactional Social Media consisting of different transactions this find all the high utility items efficiently using the proposed techniques. The proposed algorithm uses bottom up approach, where high utility items are extended one item at a time. This algorithm is more efficient compared to other state-of-the-art algorithms especially when transactions keep on adding to the Social Media from time to time. With the proposed SHU-Algorithm, incremental data Twitter mining can be done efficiently to provide the ability to use previous data structures in order to reduce unnecessary calculations when a Social Media is updated tweets, or when the minimum threshold is changed.

***Keywords*:-** Twiiter,High utility mining, data mining, minimum utility, simple high utility algorithm, utility mining, Generalized Linear Models.

I. INTRODUCTION

Data mining is the process of discovering or detecting something new from a large Social Media. The overall goal of data mining is to extract information from a data set and transform it into an understandable structure for future use. In today's world raw data is being collected by companies at an exploding rate. For example, Walmart processes over 20 million point-of-sale transactions every day. This information is stored in a centralized Social Media, but would be useless without some type of data mining software to analyse it. If Walmart analyzed their point-of-sale data with data mining techniques they would be able to determine sales trends, develop marketing campaigns, and more accurately predict customer loyalty. Market basket analysis [1] relates to data mining use in retail sales.

If a clothing store records purchase of customers, a data mining system could identify those customers who favour silk shirts over cotton ones, the frequently purchased cloth, the clothing that is purchased along with silk shirt etc. These are done using techniques called association rule mining[1], [13]frequent Twitter mining [1] multi level association rule mining[13], sequential Twitter mining[2].

A.Frequent Twitter Mining

Frequent Twitter mining [1] is the most popular one among these. It finds the number of items frequently purchased by customers, but it treats all items with same importance. So it may present frequent but unprofitable itemsets to users and may lose infrequent but valuable itemsets. The purchased quantities and unit profits are not considered with this mining technique. Association rule mining [1] is concerned with just whether an item is purchased or not. Therefore it also cannot satisfy the requirements of users who are interested in discovering the itemsets with high sales profit, since the profits are composed of unit profits and purchased quantities. Therefore these popular mining methods cannot fulfil the requirement of finding the most valuable itemsets that contribute to the major part of the total profits in a retail business. This gives the motivation to develop a mining model to discover the itemsets contributing to the majority of the profit.

B. *High Utility Mining*

Recently a few utility mining models [5] were defined to discover more important knowledge from a Social Media. Utility mining is concerned with both the purchased quantities of items and unit profit of each item. So by utility mining several important business decisions like what are the items to be put on display, how the items can be arranged in shelves, how the coupons can be designed etc can be made. Mining high utility items from Social Medias is a difficult task since download closure property [1] applicable in frequent Twitter mining does not hold in this i.e., a superset of a low utility itemset may be a high utility itemset. Hence finding all the high utility itemsets from a transactional Social Media without any miss is a big challenge in high utility mining.

C. *Tree Based Approach*

Existing studies [30] using tree based approach, first build a tree scanning the Social Media many times and using the minimum utility value. This tree is then mined to find high utility itemsets. But if the minimum utility value is changed, or if transactions are added from time to time, the tree has to be built starting all over again and then mined. To address these issues we propose a novel SHU- algorithm, with which we can use the previous data structures and avoid unnecessary repeated calculations, when the Social Media is updated or the mining threshold is changed. The rest of this paper is organized as follows: In section 2, we introduce the related work for high utility mining and in section 3, the proposed algorithm is explained in details.

II. RESEARCH METHODOLOGY

In the following sections, we will discuss research on frequent Twitter mining, weighted frequent Twitter mining, high utility Twitter mining, association rule mining, sequential Twitter mining and weighted association rule mining.

Extensive studies have been proposed for mining frequent Twitter s [1]. Among the issues of frequent Twitter mining, the most famous are association rule mining [21] and sequential Twitter mining [2], [22]. One of the well-known algorithms for mining association rules is Apriori [1], which is the pioneer for efficiently mining association rules from large Social Medias. Twitter growth-based association rule mining algorithms [14] such as FP-Growth [14] were afterward proposed. Let *I*={i1,i2,...im} be a set of items and *D* be the Social Media containing transactions {T1,T2,T3,...Tn}. The support count of a Twitter is the number of transactions in which the Twitter is present. With frequent Twitter mining itemsets having minimum support count will be the output.

In frequent itemset mining, the importance of each item is not considered. Thus, the topic called weighted association rule mining was brought to attention [4],. Cai et al. first proposed the concept of weighted items and weighted association rules [4]. But since weighted association rules do not have downward closure property, mining performance cannot be improved. To address this problem, Tao et al. proposed weighted downward closure property [28]. There are also many studies that have developed different weighting functions for weighted

Twitter mining. Weighted association rule mining considers the importance of items, but quantities of items in transactions are not taken into considerations. Thus, the issue of high utility itemset mining is raised and many studies [3] have addressed this problem. Liu et al. proposed an algorithm named Two Phase [19] which is mainly composed of two phases. In phase I, it employs an Apriori-based method to enumerate HTWUIs. Candidate itemsets with length k are generated from length k-1 itemsets, and their TWUs are computed by scanning the Social Media once in each pass. After the above steps, the complete set of HTWUIs can be collected in phase I. In phase II, HTWUIs that are high utility itemsets are identified.

A. *Performance of Algorithms*

The number of candidates generated is a critical issue for the performance of algorithms. By applying the proposed algorithm, the number of generated candidates can be highly reduced and high utility itemsets can be identified more efficiently. With the proposed SHU-Algorithm, incremental data mining can be done efficiently to provide the ability to use previous data structures in order to reduce unnecessary calculations when a Social Media is updated, or when the minimum threshold is changed.

III. PROPOSED ALGORITHM

The goal of utility mining is to discover all the itemsets whose utility values are beyond a user specified threshold in a transactional Social Media. We start with the definition of a set of terms that leads to the formal definition of utility mining problem. I = {i1, i2,…, im} is a set of items. D = {T1, T2,…, Tn} be a transactional Social Media where each transaction Ti ∈ D is a subset of I. Suppose we are given a Social Media D as shown in Table 1. This is a transactional Social Media containing a number of transactions. Each transaction is identified by a unique identifier called TID. Each row is a transaction. The columns represent the number of items in a particular transaction. Each transaction consists of the items purchased and the quantities of each item. For example the transaction T5 consists of 10 quantities of b, 8 quantities of d, and 9 quantities of e. There is another table called Utility table as shown in Table 2. Utility table stores utility or profit of each item per unit in rupees present in the Social Media.

A.*Algorithm*(SHU)

Algorithm Largest Number Twitter User To transcation T_M

Input: A list of numbers L ->user List T_U.

Output: The largest number in the list L->Language Twitter.

if L.size = 0 return null Tweet

largest ← L[0] Return of Relevancy

for each item in L->Time user Frequency, do

if item > largest, then

largest ← item

Return largest

Table 1: Social Media (showing different transactions)

Item / TID	a	b	c	d	e
T1	**2**	**2**	**0**	**0**	**0**
T2	**3**	**0**	**12**	**4**	**2**
T3	**0**	**0**	**15**	**0**	**3**
T4	**4**	**0**	**0**	**0**	**0**
T5	**0**	**10**	**0**	**8**	**9**
T6	**0**	**7**	**3**	**0**	**4**
T7	**1**	**0**	**2**	**0**	**1**
T8	**2**	**0**	**0**	**1**	**3**

Table 2: Utility Table (showing profit in rupees of each item per unit)

Item	a	b	c	d	e
Profit(Rs) (per unit)	2	15	3	8	7

Utility of an item '*i* ' in a transaction *Td* is denoted as U(*i*,*Td*); and it is the profit of that item i in transaction *Td*. For e.g., U (a, *T1*) =2* 2 = 4.Utility of an itemset X in *Td* is denoted as u(X; *Td*) and is defined as the profit of X in *Td*; i.e., profit from all the items in itemset X in *Td*. For eg., U(ab, *T1*)= U(a,*T1*) + U(b,*T1*) = 4 + 30 =34.

Utility of an itemset X in the whole Social Media *D* is denoted as U(X) and is defined as the sum of the utilities of X from all the transactions in which X is present. For eg, U (de) = U (de, *T2*) + U (de, *T5*) + U (de, *T8*) = 46 +127 + 29 =202. If minimum utility is set to 30, (de) is a high utility itemset.

B.Problem Statement

Given a transaction Social Media D and a user-specified minimum utility threshold "*minimum utility*", the problem of mining high utility itemsets from D is to find the complete set of the itemsets whose utilities are larger than or equal to *minimum utility*. If an itemset has utility less than minimum utility threshold, it is a low utility itemset.

C. ***Proposed Algorithm: Simple High Utility Algorithm (SHU-Algorithm)***

In the first scan of the Social Media we are collecting frequency and profit of different items and filling the entries in the frequency table, profit table and total frequency table as shown in Table 3, 4 and 5 respectively. Frequency table stores the total frequency of occurrences of each item. In the example Social Media there are 5 items, 'a', 'b', 'c', 'd' and 'e' present in different transactions. The item 'a' is present in transactions T1, T2, T4, T7 and T8 with the quantity of 'a' purchased as 2,3,4,1 and 2 respectively. So total frequency of occurrence of 'a' is 2+3+4+1+2 =12. Similarly occurrences of 'b', 'c', 'd' and 'e' are 19, 32, 13 and 22 respectively.

Table 3: Frequency Table (showing frequency of each item in the whole Social Media)

Item	a	b	c	d	e
Profit(Rs) (per unit)	12	19	32	13	22

Table 4 shows the profit table, which is also filled during first scan of the Social Media. It stores different items present in the Social Media with utility of each item in the Social Media. Recall from the earlier section, the definition of utility of an item/itemset X in the whole Social Media D, denoted as U(X) is the sum of the utilities of X from all the transactions in which X is present. In other words profit of an item in the whole Social Media is the product of total frequency (obtained from Table 3) and unit profit (obtained from Table 2). So utility of item 'a' is calculated as follows: from the frequency Table shown in Table3, frequency of 'a' in whole Social Media is 12; and from utility table in Table 2, utility of item 'a' per unit is 2. So profit of 'a' in the whole Social Media is the product of total frequency and unit profit; i.e., 12*2= 24.

Table 4: Profit Table (showing profit in rupees of each item in the whole Social Media)

Item	A	b	c	d	e
Profit(Rs) (per unit)	24	285	96	104	154

Similarly profit table shown in Table 4 is filled for each item as follows:

Profit of 'b'= 19*15 = 285;

Profit of 'c'= 32*3 = 96;

Profit of'd'= 13*8 = 104;

Profit of 'e'= 22*7 = 154;

After filling frequency table and profit table, we need to generate all supersets with cardinality 2 of each item present in the Social Media. Since the items present in the Social Media are 'a', 'b', 'c', 'd' and more profit frequent sets 'e', the possible supersets with cardinality 2 are {ab},{ac}{ad},{ae},{bc},{bd},{be},{cd},{ce}and{ de}. Now create a table 'total frequency-2 table' (here by 2 we mean frequency of itemsets containing 2 items) containing all itemsets with 2 items as the first row. The frequency values under these are initialised to zero i.e., (0:0). Now consider the first transaction T1 in original Social Media (shown in Table 1), in which the items present are only 'a' and 'b'. The possible superset with 2 items 'a' and 'b' are {ab}. The frequency of {ab} in T1 from Table 1 can be obtained as (2:2). So store under {ab}, the current frequency value of {ab} plus (2:2); i.e., (0:0) + (2:2) = (2:2).

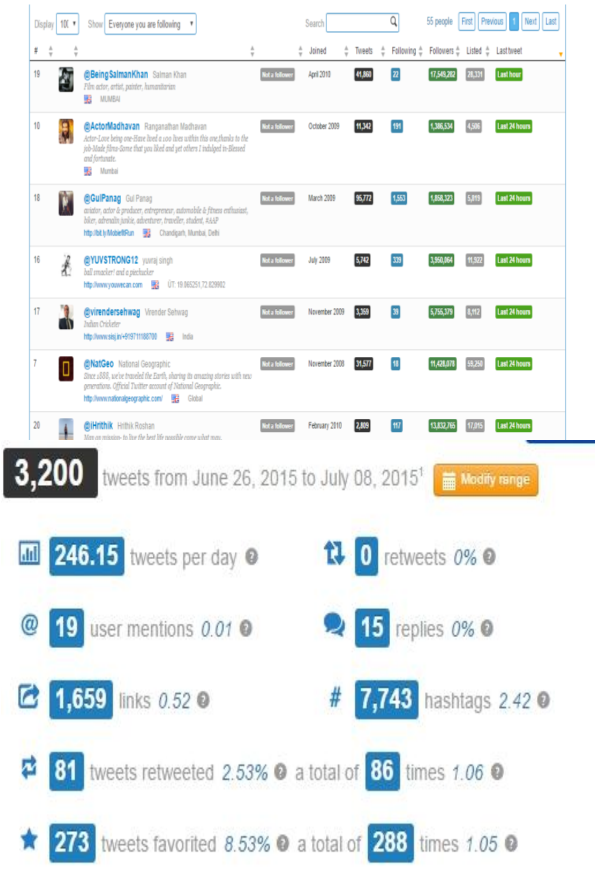


Fig. 1 Twitter Algorithms Analysis

Consider the next transaction T2 containing items ‘a’, ‘c’, ‘d’ and ‘e’. The supersets that can be formed with these items are {ac},{ad},{ae},{cd},{ce} and {de}. The respective frequency values of these are obtained from the corresponding row of T2 in Table1 as follows- (3:12), (3:4), (3:2), (12:4), (12:2) and (4:2). So these frequency values of each item are stored as their old value plus these new values.

Frequency of {ac} = (0:0) + (3:12) = (3:12)

Frequency of {ad} = (0:0) + (3:4) = (3:4)

Frequency of {ae} = (0:0) + (3:2) = (3:2)

Frequency of {cd} = (0:0) + (12:4) = (12:4)

Frequency of {ce} = (0:0) + (12:2) = (12:2)

Frequency of {de} = (0:0) + (4:2) = (4:2)

Consider the next transaction T3 from Table1. T3 consists of items ‘c’ and ‘e’ with frequencies 15 and 3. The super sets possible with ‘c’ and ‘e’ is only {ce}. So under {ce} store old value of {ce} plus (15:3); i.e., (12:2) + (15:3) = (27:5).

Consider the next transaction T4 from Table1. T4 consists of item ‘a’ only and so no super set with cardinality 2 is possible with item ‘a’ only.Consider the next transaction T5 from Table1. T5 consists of items ‘b’,‘d’ and ‘e’ with frequencies 10, 8 and 9. The super sets possible with ‘b’, ‘d’ and ‘e’ are {bd}, {be}and {de}.

Frequency of {bd} = (0:0) + (10:8) = (10:8);

Frequency of {be} = (0:0) + (10:9) = (10:9)

Frequency of {de}= (4:2) + (8:9) = (12:11)

After considering T6, T7 and T8 and filling corresponding frequency values under each superset, we get the entries as shown in Table 5. Thus during the first scan of the Social Media we fill the entries in tables 3, 4 and 5 respectively. By using information in these tables, we need to find all the potential high utility items. If the minimum utility is set to 200, all the item/itemsets with utility > 200 are considered as high utility itemsets, else low utility itemsets.

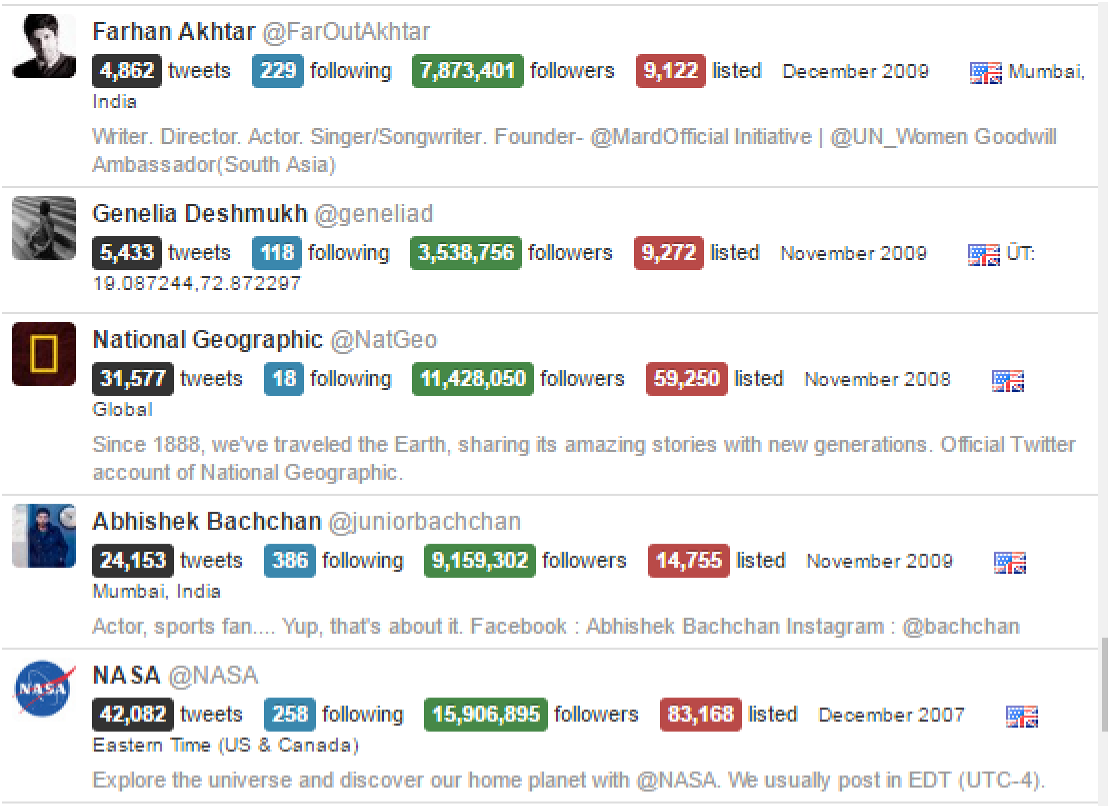


Fig. 2 Twitter User Transcation Details

Table 5: Total Frequency-2 Table (showing frequencies of itemsets with cardinality 2)

itemset	ab	ac	ad	ae	bc	bd	be	cd	ce	De
frequency	(2:2)	(4:14)	(5:5)	(6:6)	(7:3)	(10:8)	(17:13)	(12:4)	(32:10)	(13:14)

Now first we will find all the high utility itemsets containing 1 item, 2 items, etc. From Table 4 i.e., Profit Table showing profit in rupees of each item in the whole Social Media, we can find all the high utility items consisting of only one item. Out of the items ‘a’, ‘b’, ‘c’, ‘d’ and ‘e’ only ‘b’ is a high utility item, since its utility (U(b)=285) is greater than minimum utility i.e., 200.

The items ‘b’, ’c’, ‘d’ and ‘e’ are low utility items with utilities 24, 96, 104 and 154 respectively which are all less than 200. Now we need to consider itemset containing 2 items. Consider each of the supersets and their frequencies listed in Table5, they are ‘ab’, ‘ac’, ‘ad’, etc. Considering each of these one by one, f (ab) i.e., frequency of ‘ab’ in the whole Social Media is (2:2). The quantity of ‘a’ is 2, quantity of b is 2 and from Utility Table, unit profit of ‘a’ is 2 and unit profit of ‘b’ is 15. From this utility of ‘ab’ denoted as U(‘ab’)=2*2 + 2*15 = 34, which is <200; so ‘ab’ is a low utility itemset. Consider the next itemset ’ac’. From Table 5, f (‘ac’) = (4:14).

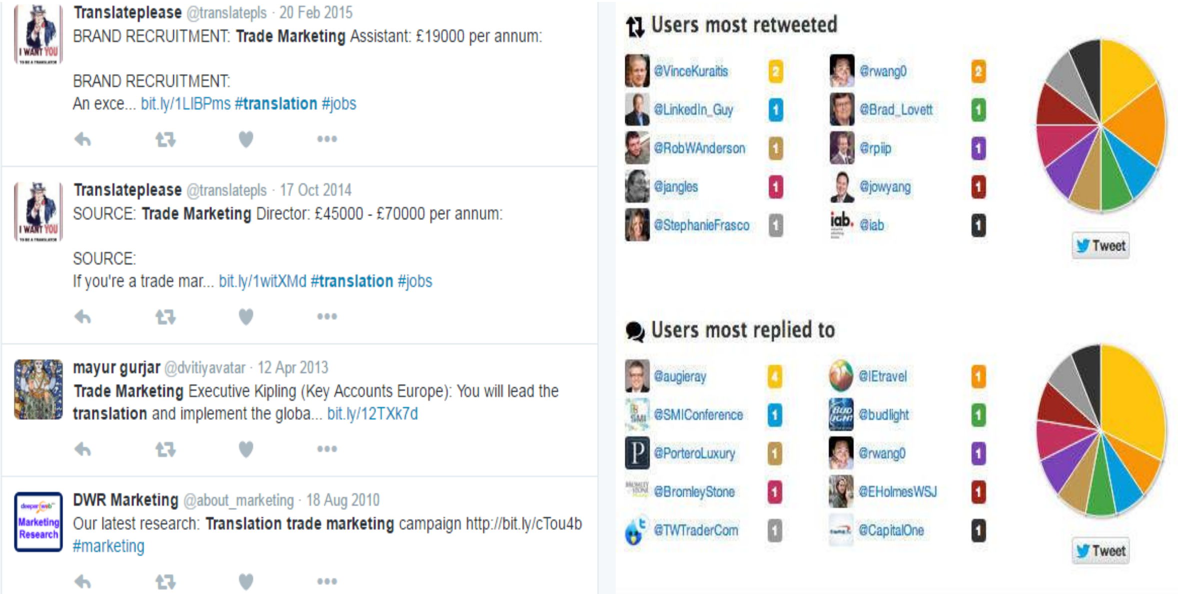


Fig. 3 Twitter Trade Market Transcation Analysis

So U (‘ac’) = 4*2+ 14*3 = 50, which is <200; so ‘ac’ is a low utility itemset.

f (‘ad’) = (5:5).So U(‘ad’)= 5*2+ 5*8 = 50, which is <200; so ‘ad’ is a low utility itemset.

f (‘ae’) = (6:6). U (‘ae’) = 6*2+ 6*7 = 54, which is <200; so ‘ae’ is a low utility itemset.

f (‘bc’) = (7:3). U (‘bc’) = 7*15+ 3*3 = 114, which is <200; so ‘bc’ is a low utility itemset.

f (‘bd’) = (10:8). U (‘bd’)= 10*15+ 8*8 = 214, which is >200; so ‘bd’ is a high utility itemset.

f (‘be’) = (17:13). U (‘be’) = 17*15+ 13*7 = 346, which is >200; so ‘be’ is a high utility itemset.

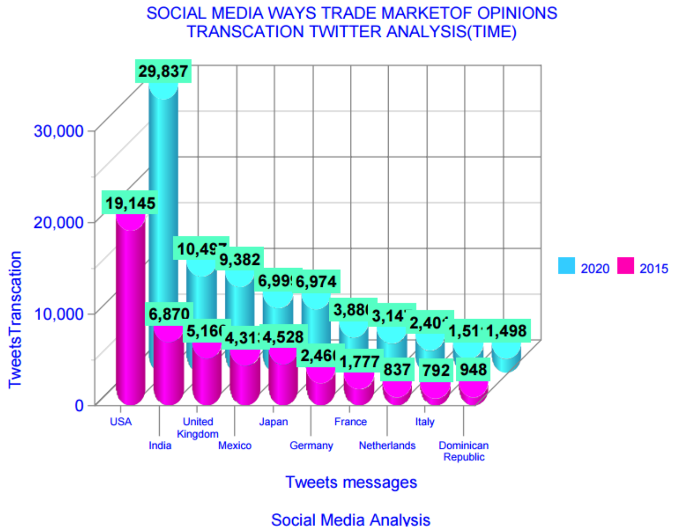
f (‘cd’) = (12:4). U (‘ac’) = 12*3+ 4*8 = 68, which is <200; so ‘cd’ is a low utility itemset.

f(‘ce’) = (32:10). U (‘ce’) = 32*3+ 10*7 = 166, which is <200; so ‘ce’ is a low utility itemset.

f(‘de’) = (13:14). U (‘de’) = 13*8+ 14*7 = 202, which is >200; so ‘de’ is a high utility itemset.

Till now we have scanned the Social Media only once and all the high utility items we have found till now are the actual high utility items itself, not potential high utility items. Now we need to find all the potential high utility itemsets (PHUI) with itemset containing 3 items and from these candidates we need to find the actual high utility itemsets. First we need to generate all the supersets containing 3 items each. They are ‘abc,’abd’, ‘abe’, ‘acd’, ‘ace’, ‘ade’ ‘bcd’ , ‘bce’, ‘bde’ and ’cde’. We need to consider each of these one by one and check if they are PHUI. For that we need frequency and unit profit, but frequencies for itemsets with cardinality 3 are not stored in Table 5(but unit profit can be calculated from Table 2). To find the frequencies we use a new approach that is explained below with examples.

D.*Discussion*



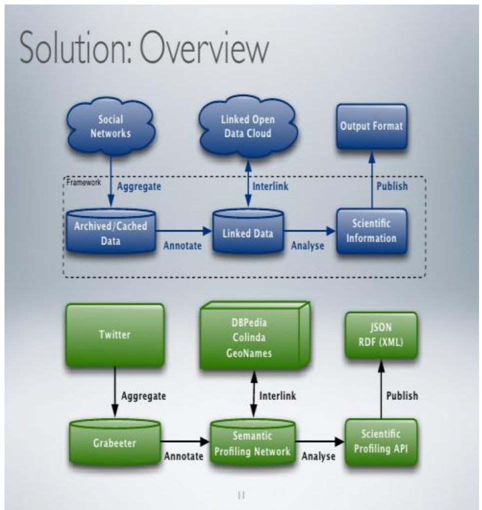


Fig. 4 Twitter Trade Market Transcation Analysis Result

From Table 5, f('ab')=(2:2) and f('bc')=(7,3). So f(a,b,c)<=(f('a') in f('ab') : minimum of{f('b') in f('ab') , f('b') in f('bc')}: f('c') in f('bc')). i.e., f(a,b,c)<=(2: minimum of{2,7}: 3), i.e., (2:2:3). From profit table unit profit of 'a', 'b' and 'c' are respectively 2, 15 and 3. So estimated utility of 'abc' is U('abc') = 2*2+2*15+3*3= 43<200, so'abc' not included in candidate set. Consider the next itemset 'abd'. Frequency f('abd') needs to be estimated. The overestimated frequencies can be as follows: From Table 5, f('ab')=(2:2) and f('bd')=(10,8). So f(a,b,d)<=(f('a')) in f('ab'): minimum of{f('b') in f('ab') , f('b') in f('bd')}: f('d') in f('bd')). i.e., f(a,b,d)<=(2: minimum of{2,10}: 8), i.e., (2:2:8). From profit table unit profit of 'a', 'b' and'd' are respectively 2, 15 and 8. So estimated utility of 'abd' is U('abd') = 2*2+15*2+8*8= 98<200, so'abd' not included in candidate set. Consider the next itemset 'abe'. Frequency f('abe') needs to be estimated. The overestimated frequencies can be as follows: From Table 5, f ('ab') = (2:2) and f ('be')=(17,13). So f(a,b,e)<=(f('a')) in f('ab'): minimum of{f('b') in f('ab') , f('b') in f('be')}: f('e') in f('be')). i.e., f (a,b,e)<=(2: minimum of{2,17}: 13), i.e., (2:2:13). From profit table unit profit of 'a', 'b' and 'e' are respectively 2, 15 and 7. So estimated utility of 'abe' is U('abe') = 2*2+15*2+7*13= 125<200, so'abe' not included in candidate set. Similarly for checking all the other supersets i.e., 'acd','ace', 'bcd', 'bce', 'bde', 'cde', 'ade' that if they can be included in the candidate set, we get the estimated frequencies as (4:12:4), (4:14:10), (7:3:4), (7:3:10), (10:8:14), (12:4:14) and (5:5:14)and estimated utilities as 76,120,146,184,312,166 and 148. All these utilities except that of 'bde' (whose overestimated utility is 312) are less than minimum utility threshold 200 and so these itemsets except 'bde' are not included in the candidate set. So after this we get candidate itemset formed with 3 items as {'bde'}. Similarly we can find candidate sets for itemsets containing 4 items, 5 items.

Top 10 Most Followed Twitter Accounts		
Twitter Account	**Followers**	**RTs/Tweet**
@KatyPerry	54,772,525	3,331
@JustinBieber	53,146,619	54,150
@BarackObama	44,386,962	1,358
@YouTube	43,889,182	226
@TaylorSwift13	42,127,368	5,554
@LadyGaga	41,676,842	5,004
@BritneySpears	38,214,513	1,400
@Rihanna	36,568,593	3,647
@Instagram	34,067,091	386
@Jtimberlake	33,306,591	932

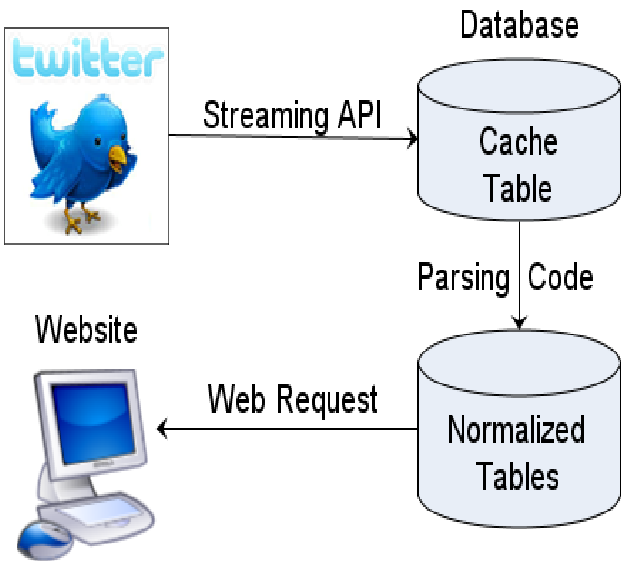


Fig. 5 Twitter Trade Market Transcation Analysis Result overview

After finding candidate sets of itemsets containing 3, 4,.. items, we need to find their actual utility from the Social Media and see if their actual utility is greater than minimum utility threshold. In this example, 'bde' is the only item in the candidate set. From the Social Media we can find the actual utility of 'bde' as 10*15+8*8+ 9*7=277, which is >200(please note here that actual utility is less than its overestimated utility 312). So 'bde' is a high utility itemset. So combining all the results we get the high utility itemsets as {'b':285, 'bd':214, 'be': 346, 'de':202, 'bde':277}.

IV. CONCLUSION

In this paper, we have proposed a SHU-Algorithm for finding all the high utility itemsets present in a transactional Social Media. This SHU algorithm can be then used for efficient mining of all the high utility itemsets from transaction Social Media. PHUIs can be efficiently generated using only one Social Media scan. Moreover, we have developed an approach which can be applied to a Social Media to which transactions keep on adding from time to time. Whenever a new transaction is added we need to update only a few corresponding values in the tables 3 and 5. Whenever minimum threshold is changed or a few transactions are added to a Social Media, we can still use the previous values with only a few changes i.e., incremental data mining can be done efficiently to provide the ability to use previous data structures in order to reduce unnecessary calculations when a Social Media is updated, Besides that we decrease overestimated utility of itemsets by using the minimum of the frequency values for that item and thus enhance the performance of utility mining. This proposed algorithm, will outperform the state of-the-art algorithms substantially especially when Social Medias contain small number of items or when transactions keep on adding to the Social Media from time to time.

REFERENCES

[1] M. Oussalah, B. Escallier, D. Daher , An automated system for grammatical analysis of Twitter messages. A learning task application ,Knowledge-Based Systems, Volume 101, 1 June 2016, Pages 31-47.

[2] Gi Woong Yun, David Morin, Sanghee Park, Claire Youngnyo Joa, Brett Labbe, Jongsoo Lim, Sooyoung Lee, Daewon Hyun,Social media and flu: Media Twitter accounts as agenda setters , International Journal of Medical Informatics, Volume 91, July 2016, Pages 67-73.

[3] Emilio Serrano, Carlos A. Iglesias, Validating viral marketing strategies in Twitter via agent-based social simulation,Expert Systems with Applications, Volume 50, 15 May 2016, Pages 140-150.

[4] Feixiong Luo, Guofeng Cao, Kevin Mulligan, Xiang Li,Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago Applied Geography, Volume 70, May 2016, Pages 11-25.

[5] Saifuddin Ahmed, Kokil Jaidka, Jaeho Cho The 2014 Indian elections on Twitter: A comparison of campaign strategies of political parties ,Telematics and Informatics, Volume 33, Issue 4, November 2016, Pages 1071-1087.

[6] Guy Lansley, Paul A. Longley,The geography of Twitter topics in London,Computers, Environment and Urban Systems, Volume 58, July 2016, Pages 85-96

[7] George Veletsianos, Royce Kimmon Scholars in an increasingly open and digital world: How do education professors and students use Twitter? ,The Internet and Higher Education, Volume 30, July 2016, Pages 1-10.

[8] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber, Carlos M. Dancausa, Frederic Stahl, João Bártolo GomesA rule dynamics approach to event detection in Twitter with its application to sports and politics, Expert Systems with Applications, Volume 55, 15 August 2016, Pages 351-360.

[9] Lucy L. Gu, Dorothy Skierkowski, Paul Florin, Karen Friend, Yinjiao Ye Facebook, Twitter, & Qr codes: An exploratory trial examining the feasibility of social media mechanisms for sample recruitment ,Computers in Human Behavior, Volume 60, July 2016, Pages 86-96

[10] Jihyun Kim, Hayeon Song ,Celebrity's self-disclosure on Twitter and parasocial relationships: A mediating role of social presence,Computers in Human Behavior, Volume 62, September 2016, Pages 570-577.

[11] Fábio C. Albuquerque, Marco A. Casanova, Hélio Lopes, Luciana R. Redlich, José Antonio F. de Macedo, Melissa Lemos, Marcelo Tilio M. de Carvalho, Chiara Renso, A methodology for traffic-related Twitter messages interpretation , Computers in Industry, Volume 78, May 2016, Pages 57-69.

[12] Paolo Nicola Barbieri, Francesca Fazio, Gabriele Gamberini ,Let young people join the legislative process. A Twitter based experiment on internships Computers in Human Behavior, Volume 60, July 2016, Pages 468-482.

[13] Kathleen M. Carley, Momin Malik, Peter M. Landwehr, Jürgen Pfeffer, Michael Kowalchuck ,Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia, Safety Science, In Press, Corrected Proof, Available online 13 April 2016.

[14] Megan A. Moreno, Alina Arseniev-Koehler, Dana Litt, Dimitri Christakis ,Evaluating College Students' Displayed Alcohol References on Facebook and Twitter ,Journal of Adolescent Health, Volume 58, Issue 5, May 2016, Pages 527-532

[15] Salvatore Gaglio, Giuseppe Lo Re, Marco Morana,A framework for real-time Twitter data analysis Computer Communications, Volume 73, Part B, 1 January 2016, Pages 236-242

[16] Siaw Ling Lo, Raymond Chiong, David Cornforth,Ranking of high-value social audiences on Twitter Decision Support Systems, Volume 85, May 2016, Pages 34-48

[17] Hassan Saif, Yulan He, Miriam Fernandez, Harith Alani,Contextual semantics for sentiment analysis of TwitterInformation Processing & Management, Volume 52, Issue 1, January 2016, Pages 5-19

[18] Francois R. Lamy, Raminta Daniulaityte, Amit Sheth, Ramzi W. Nahhas, Silvia S. Martins, Edward W. Boyer, Robert G. Carlson ,"Those edibles hit hard": Exploration of Twitter data on cannabis edibles in the U.S. Drug and Alcohol Dependence, In Press, Accepted Manuscript, Available online 26 April 2016

[19] Yun-Cheol Heo, Ji-Young Park, Ji-Young Kim, Han-Woo Park,The emerging viewertariat in South Korea: The Seoul mayoral TV debate on Twitter, Facebook, and blogs Telematics and Informatics,

Volume 33, Issue 2, May 2016, Pages 570-583

[20] Fabián Riquelme, Pablo González-Cantergiani,Measuring user influence on Twitter: A survey Information Processing & Management, In Press, Corrected Proof, Available online 23 April 2016

[21] Alina Arseniev-Koehler, Hedwig Lee, Tyler McCormick, Megan A. Moreno,#Proana: Pro-Eating Disorder Socialization on Twitter, Journal of Adolescent Health, In Press, Corrected Proof, Available online 12 April 2016

[22] Pete Burnap, Rachel Gibson, Luke Sloan, Rosalynd Southern, Matthew Williams,140 characters to victory?: Using Twitter to predict the UK 2015 General Election Electoral Studies, Volume 41, March 2016, Pages 230-233

[23] Ana Carolina E.S. Lima, Leandro Nunes de Castro, Juan M. Corchado, A polarity analysis framework for Twitter messages ,Applied Mathematics and Computation, Volume 270, 1 November 2015, Pages 756-767

[24] Heli Aramo-Immonen, Jari Jussila, Jukka Huhtamäki,Exploring co-learning behavior of conference participants with visual network analysis of Twitter data,Computers in Human Behavior, Volume 51, Part B, October 2015, Pages 1154-1162