**Sentiment Analysis and Opinion Mining: A Survey**

Kaushik Hande

M.E Student, Dept. of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India

**ABSTRACT**: With the advent of Web 2.0, people became more eager to express and share their opinions on web regarding day-to-day activities and global issues as well. Evolution of social media has also contributed immensely to these activities, thereby providing us a transparent platform to share views across the world. Sentiment analysis (SA) is a computational study of opinions, sentiments, emotions, and attitude expressed in texts towards an entity. Sentiment Analysis identifies the polarity of extracted public opinions. This survey paper tackles a comprehensive overview of the last update in this field and presents a survey which covers Opining Mining, Sentiment Analysis, techniques, tools and classification.

KEYWORDS: Sentiment Analysis, Opinion Mining, Machine learning

1. **Introduction**

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features.

1. **Related work**

In [2] authors used average residual battery level of the entire network and it was calculated by adding two fields to the RREQ packet header of a on-demand routing algorithm i) average residual battery energy of the nodes on the path ii) number of hops that the RREQ packet has passed through. According to their equation retransmission time is proportional to residual battery energy. Those nodes having more battery energy than the average energy will be selected because its retransmission time will be less. Small hop count is selected at the stage when most of the nodes have same retransmission time. Individual battery power of a node is considered as a metric to prolong the network lifetime in [3]. Authors used an optimization function which considers nature of the packet, size of the packet and distance between the nodes, number of hops and transmission time are also considered for optimization. In [ 4] initial population for Genetic Algorithm has been computed from the multicast group which has a set of paths from source to destination and the calculated lifetime of each path. Lifetime of the path is used as a fitness function. Fitness function will select the highest chromosomes which is having highest lifetime. Cross over and mutation operators are used to enhance the selection. In [5] authors improved AODV protocol by implementing a balanced energy consumption idea into route discovery process. RREQ message will be forwarded when the nodes have sufficient amount of energy to transmit the message otherwise message will be dropped. This condition will be checked with threshold value which is dynamically changing. It allows a node with over used battery to refuse to route the traffic in order to prolong the network life. In [6] Authors had modified the route table of AODV adding power factor field. Only active nodes can take part in rout selection and remaining nodes can be idle. The lifetime of a node is calculated and transmitted along with Hello packets. In [7] authors considered the individual battery power of the node and number of hops, as the large number of hops will help in reducing the range of the transmission power. Route discovery has been done in the same way as being done in on-demand routing algorithms. After packet has been reached to the destination, destination will wait for time δt and collects all the packets. After time δt it calls the optimization function to select the path and send RREP. Optimization function uses the individual node’s battery energy; if node is having low energy level then optimization function will not use that node.

1. **Proposed algorithm**
   1. *Design Considerations:*

* Initial battery energy (IBE) is 50Jules for each node.
* Nodes are able to calculate its residual battery energy (RBE).
* Keeping track of previously used paths.
* Considered all possible paths at beginning.
* Receiving energy is not considered.
* The time when no path is available to transmit the packet is considered as the network lifetime.
  1. *Description of the Proposed Algorithm:*

Aim of the proposed algorithm is to maximize the network life by minimizing the total transmission energy using energy efficient routes to transmit the packet. The proposed algorithm is consists of three main steps.

Step 1: Calculating Transmission Energy:

The transmission energy (TEnode )of each node relative to its distance with another node is calculated by using eq.(1)[8].

TEnode α dn

TEnode = k dn  eq. (1)

where k is constant and n is path loss factor which is generally between (2-4) [8].

Step 2: Selection Criteria:

Node should have more residual battery energy (RBE) than the required transmission energy (TEnode ) to transmit the packet to the next node in the route. All the nodes in the route will be checked with this condition even if one node of a route is not satisfying this condition then that route will not be considered as a feasible solution. All the other routes having all the nodes with sufficient amount of energy are considered as the feasible solution. And those nodes having equal RBE than (TEnode ) are made to go into sleep mode. This selecting criterion helped to prolong the network life by avoiding the link breakage. We tried to avoid the repeated use of the path. But at one stage we have to compromise with energy efficiency when we have a route with less energy consumption but it is already being used and a rout with maximum consumption of energy which is not used. So till this point we avoided repeated use of the paths and tried to increase the network life. Transmission energy of a node to node in a rout is calculated according to the distance and the total transmission energy (TTER) for that rout is calculated using eq. (2).

TTER= eq. (2)

where m is the number of hops in the route, TE = TEnode is the transmission energy between the nodes. The route having minimum total transmission energy i.e. min (TTER)will be selected as energy efficient route.

Step 3: Calculating Residual Battery Energy (RBE):

After transmitting the packet, residual battery energy for each node of the route is calculated using eq. (3) with parameters initial battery energy (IBE) and TEnode.

RBE = IBE – TEnode

eq. (3)

Step 1: Generate all the possible routes.

Step 2: Calculate the TEnode for each node of each route using eq. (1).

Step 3: Check the below condition for each route till no route is available to transmit the packet.

if (RBE < = TEnode)

Make the node into sleep mode.

else

Select all the routes which have active nodes

end

Step 4: Calculate the total transmission energy for all the selected routes using eq. (2).

Step 5: Select the energy efficient route on the basis of minimum total transmission energy of the route.

Step 6: Calculate the RBE for each node of the selected route using eq. (3).

Step 7: go to step 3.

Step 8: End.

1. **Simulation Results**

The simulation studies involve the deterministic small network topology with 5 nodes as shown in Fig.1. The proposed energy efficient algorithm is implemented with MATLAB. We transmitted same size of data packets through source node 1 to destination node 5. Proposed algorithm is compared between two metrics Total Transmission Energy and Maximum Number of Hops on the basis of total number of packets transmitted, network lifetime and energy consumed by each node. We considered the simulation time as a network lifetime and network lifetime is a time when no route is available to transmit the packet. Simulation time is calculated through the CPUTIME function of MATLAB. Our results shows that the metric total transmission energy performs better than the maximum number of hops in terms of network lifetime, energy consumption and total number of packets transmitted through the network.

The network showed in Fig. 1 is able to transmit 22 packets if total transmission energy metric is used and 17 packets if used maximum number of hops metric. And the network lifetime is also more for total transmission energy. It clearly shows in Fig. 2 that the metric total transmission energy consumes less energy than maximum number of hops. As the network is MANET means nodes are mobile and they change their locations. After nodes have changed their location the new topology is shown in Fig .3 and energy consumption of each node is shown in Fig. 4. Our results shows that the metric total transmission energy performs better than the maximum number of hops in terms of network lifetime, energy consumption and total number of packets transmitted through the network.

 

Fig.1. Ad Hoc Network of 5 Nodes Fig. 2. Energy Consumption by Each Node

 

Fig. 3. Ad Hoc Network of 5 Nodes Fig 4. Energy Consumption by Each Node

1. **Conclusion and Future Work**

The simulation results showed that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric. The proposed algorithm provides energy efficient path for data transmission and maximizes the lifetime of entire network. As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. We have used very small network of 5 nodes, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

**References**

1. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79-86, 2002.
2. B. Pang and L. Lee,“Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval,vol. 2, no. 1-2, pp. 1-135, 2008
3. P. Turney, “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews,” Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2002.
4. Pang B. and L. Lee. “ Opinion Mining and Sentiment Analysis.” Foundation and Trends in Information Retrieval, 2(1-2):1-135 2008.
5. S. Das and M. Chen,“Yahoo! for Amazon: Extracting market sentiment from stock message boards,” Proceedings of the Asia Pacific Finance Association Annual Conference, 2001.
6. Rui Xia , Chengqing Zong, Shoushan Li, “ Ensemble of feature sets and classification algorithms for sentiment classification” Information Sciences 181 (2011) 11381152
7. R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi and T. Li, Dual sentiment analysis Considering two sides of one review, in IEEE Transaction Knowledge. Data Engg, vol. 27, no. 8, pp. 2120 - 2133, Aug. 2015.

**BIOGRAPHY**

**Kaushik Hande** is a student pursuing M.E. in the Computer Engineering Department, Pune Institute of Computer

Technology, Pune. His research interests are Data Mining , Data Analysis and Machine Learning.