

Modeling Tweet using Propositional Logic

Vishal Mehta
Mphasis Limited
Cybercity, Tower IV
Magarpatta City, Hadapsar,
Pune – 411013, India

ABSTRACT

Today the era is of data driven decision making. For any business it has become essential to listen what their customer's experience is? What they exactly want? How is their brand performing? Are they happy customers? How they can retain them and reduce the rate of churn and convert it into revenue? What's their user base? What is the age segment which gives them the most of the business? Depending upon the answers of all the above questions are we capable to take a decision which can increase the profits? Are all the answers helping businesses to arrive to some conclusion on the basis of which one can make valid decision? What is the truth value of this conclusion? In this paper we are trying to propose a framework which will help businesses to deduce inference from what the customers are talking about their brand and accordingly they can design a new strategy to save, retain and grow their business by taking effective decisions.

Keywords— Two-Phase-Splitter-module, Premise, Hypothetical Syllogism

1. INTRODUCTION

Customer experience gets derived from customer behavior. The rate of churn in the telecom industry is highest as compared to other and hence the companies are really looking curious to understand what their customer's experience is. Are they happy with the services provided? They want to hear their customer's talking about their brands and improve upon as with the changing time and needs. Customers have started sharing their experiences on social platforms like Twitter & Facebook and this gives the huge opportunity to business to listen what their customer's experiences are and can cope with.

According to the survey people are tweeting on an average rate of 9,153 tweets per second which is huge amount of human generated data which can be analyzed and understood and can be used for business purposes while keeping the privacy of the users intact. Lots of attempts are done to understand sentiments of the customer attached with the brand.

Customers tweet on twitter by posting, saying I am happy with the product or this product didn't meet my expectations or maybe they are talking about something else which have no context to the product at all. Now all these tweets can be categorized into three categories in precise. The classifier can classify the sentiment of the tweet as positive, negative or neutral on the basis of its past knowledge base and learning capabilities. But the accuracy of predicting the sentiments by the classifier is still in question.

Does doing sentiment analysis will answer all our questions? The answer is No. by only finding the sentiments of some set people and doing the cluster analysis on that will not answer all our questions. In this paper we are trying to take this idea ahead coupling with deducing the inference from tweets will make the analysis more effective and help crucial decision making.

2. FRAMEWORK

2.1 Design

In the design we propose

- Twitter : Datasource of the system
- Tweet Splitter : Splits the tweet on the basis of connectives & tweet being atomic or compound
- Tweet Modeler : Which converts the tweet into model
- Tweet Inference Engine : Model Truth validator

2.2 Implementation

Twitter is the social platform where customer share their experiences in the form of tweet. In this system framework the input for the data source for the system is twitter. When customer tweets about some product of the company on twitter and places the hashtag, the tweet comes to system through the Twitter streaming hose pipe API and gets collected in the one tweet per record in the CSV file format. The module stores 50 thousand records per CSV file. This CSV file is then forwarded to Two-Phase-Splitter module.

The Two-Phase-Splitter module takes one tweet (record) at a time and sends it to Phase-1-Tweet-Split module which is a sub module of Two-Phase-Splitter module Fig-1. Tweet which is received by Two-Phase-Splitter module is 140 characters in length. This 140 character tweet can have multiple statements embedded in it. Phase-1-Tweet-Split module checks if the tweet is Atomic (Single Sentence: One subject and Predicate) or it is a Compound tweet (Multiple Sentences: Multiple subjects and Predicates). Phase-1-Tweet-Split module will take tweet as an input and output multiple atomic tweets (tweets with one subject and predicate).

Once the tweets are split in their atomic form they are input in Phase-2-Tweet-Split module.

Phase-2-Tweet-Split module works in two steps:

- 1) Split the atomic tweet on the basis of propositional connectives: Conjunction, Disjunction, Negation, Conditional and Biconditional.
- 2) Assign all the atomic tweets to the decision variables P1, P2, P3....PN.

Once the tweets are split by Phase-2-Tweet-Split module they are input into the Tweet Modeler in the tweet modeling pipeline where the relation between all the variables is found and a model (formula) is modeled. Once we are able to derive the model out of the tweets this model is tested for its truth value and then send to Tweet-Inference-Engine.

The Tweet-Inference-Engine takes all the tweets and helps in finding out the conclusion out of the arguments made in the form of tweet. And this conclusion can be further tested for its truth value and if found sound can be used for decision making.

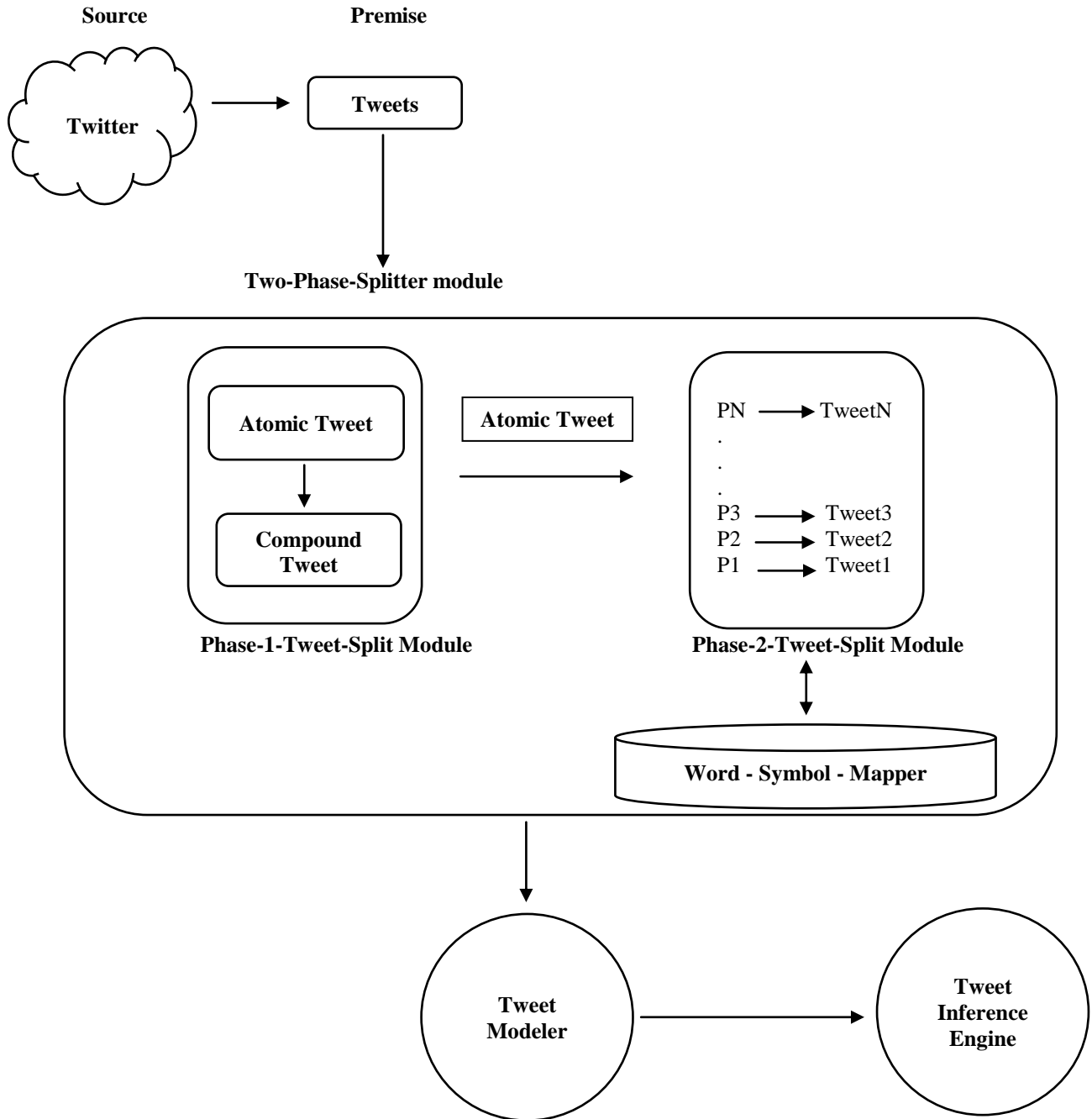


Fig.1

As shown in Fig-1 Phase-2-Tweet-Split module has a

Word – Symbol – Mapper it's a dictionary which works as a lookup table as shown below:

Dictionary

Word	Symbol
.... and	\wedge
Or	\vee
If then ...	\supset
iff	\equiv
not	\sim

In this module the tweets are referred as Premises or Propositions which can take the value either True or False.

All the atomic tweets are assigned a formula from

$P_1, P_2, P_3, \dots, P_N$ which can attain any truth value to prove their validity. The collection of all the formulas collectively as an argument or Hypothetical Syllogism which should be tested if it's a tautology or sometimes also faces contradiction if any of the premise is not a valid premise.

$$\sum P_1 \wedge P_2 \wedge P_3 \wedge P_4 \wedge P_5 \wedge P_6 \dots \wedge P(N - 1) \wedge P_N \Rightarrow C$$

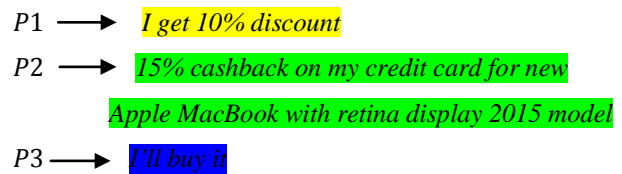
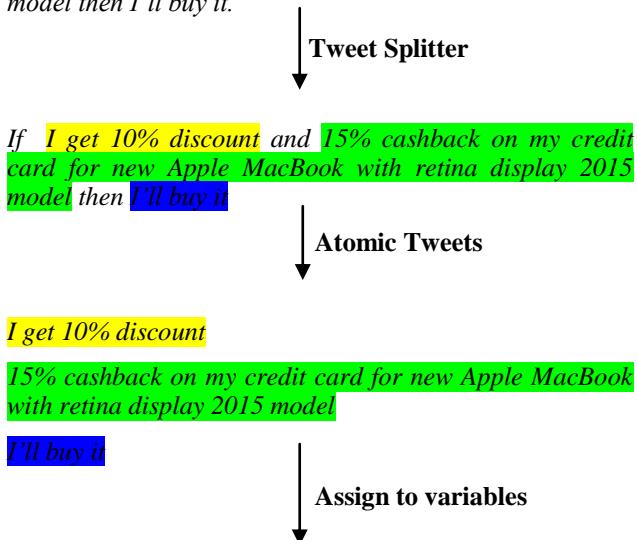
Now the Tweet Modeler helps in creating the models from the tweets done by the customer with some sentiment. As we can see below where a customer tweets his wish of buying a product:

"If I get 10% discount and 15% cashback on my credit card for new Apple MacBook with retina display 2015 model then I'll buy it."

In this tweet the person is already a customer of a telecom company and she wishes to buy a new Apple MacBook based on discounts offered as a conditional parameter.

Once the tweet modeler models the tweet it goes to next level where the Tweet Inference Engine will apply rules of inference to derive the conclusions on the basis of which a person can take the decision as we can see the system flow Fig-2.

"If I get 10% discount and 15% cashback on my credit card for new Apple MacBook with retina display 2015 model then I'll buy it."



If P_1 and P_2 then P_3

Model

$$P_1 \wedge P_2 \supset P_3$$

Inference

$$P_1 \wedge P_2$$

$$P_1$$

$$P_3$$

Fig-2

3. CONCLUSION

The above proposed system has variety of applications as in customer reviews where person wants to buy a new product but she reads through all the product reviews for the concerned product and finally gets confused after reading the reviews whether she should buy the product or not. This system will help in real time to guide the customer to buy the new product on the fly and make the right choice and mimic as an expert system by keeping an eye on the reviews posted by the buyers and help making sense out of reviews rather putting as dumb 5-star rated text on the ecommerce site.

The system also goes one step ahead in design and working from sentiment analysis done on the product based tweets and fills the gap in helping the businesses to arrive on the inference and taking effective decisions and keeping the customer behavior under their lens to predicting there needs in future.

4. REFERENCES

- [1] Leveraging Sentiment Analysis for Topic Detection IEEE Keke Cai IBM China Res. Lab., Beijing; Spangler, S.; Ying Chen; Li Zhang
- [2] Cambria, E. ; Nat. Univ. of Singapore, Singapore, Singapore ; Schuller, B. ; Yunqing Xia ; Havasi, C
- [3] Sentence-Based Sentiment Analysis for Expressive Text-to-Speech IEEE Trilla, T. ; Campus La Salle, Grup de Recerca en Tecnologies Media, Univ. Ramon Llull, Barcelona, Spain; Alias, F.
- [4] Automatic sentiment analysis of Twitter messages IEEE.