Customer requirement analysis using User Generated Content

Problem Statement

Understanding and identifying the customer need is very much important from the perspective of marketing strategy, product development and marketing research. By knowing the consumer needs, sales for the product can be increased. So customer need is nothing but to describe the benefits that a consumer expects from a product or service obtained from customer from his own words. In this context, user generated content (UGC) provides a good way to identify consumer requirements for managerial impact. So for understanding the consumer requirement, companies generally do surveys, experimental interviews or direct input from customer but these methods are expensive, time taking and involve delays as well. Apart from this, these methods are non-efficient and not that effective when the data volume of user generated content is too large as it may contain information is not relevant and gets repeated. User generated content can make this fast and efficient as it gets updated continuously, has low cost and less delays. At the same time, some problems also exist with UGC like content is repetitive and not relevant, huge volume of data exists and data is not structured and mostly text based. Therefore for such content, some machine learning algorithms may be helpful in selecting the content for review analysis. So the problem statement here is to examine or analyze a corpus of data which is user generated and identify the consumer requirements.

Background

In previous times, researchers from marketing and engineering have defined many methods to identify customer requirements which are direct input from customers like interviews, surveys, focus groups, or ethnography as input. Then trained professional process and analyze the data and identify the consumer requirements. In today's time, there are machine learning based solutions which are much efficient than traditional methods. For example - researchers in marketing have developed variety of methods to mine data in unstructured form to answer managerial questions. Apart from this deep learning models like CNN (Convolutional neural network) has been used as state-of-art solution for problems like relation extractions, object recognition and sentiment analysis. So here the approach is to use the machine learning algorithm to analyze the user generated content to determine the consumer requirements.

Methodology

Below mentioned are the steps which are used to extract the customer requirements from the user generated content (UGC) and it’s architecture in presented in the Figure 1.

Step 1: Pre-processing of User Generated Content

The data for user generated content can be taken from public sources or some company databases. After this, the content is split into the form of sentences, the stop-words, numbers and punctuations are removed from the content and then concatenate frequently occurring words.

Step 2: Training of word embeddings

After pre-processing the user generated sentences, using skip-gram model word embeddings are trained.

Step 3: Identification of informative data

The CNN (Convolutional neural network) is trained on the user generated content and then CNN filters are applied for filtering out the non-informative content in such a way that the remaining content or corpus is more towards informative content.

Step 4: Sampling of diverse content

Sentence embeddings and sample sentences are clustered from different clusters so that set of sentences are selected which represents diverse consumer requirements.

Step 5: Extraction of customer needs manually

Professional analysis can be done by reviewing the diverse informative content to identify the customer needs.

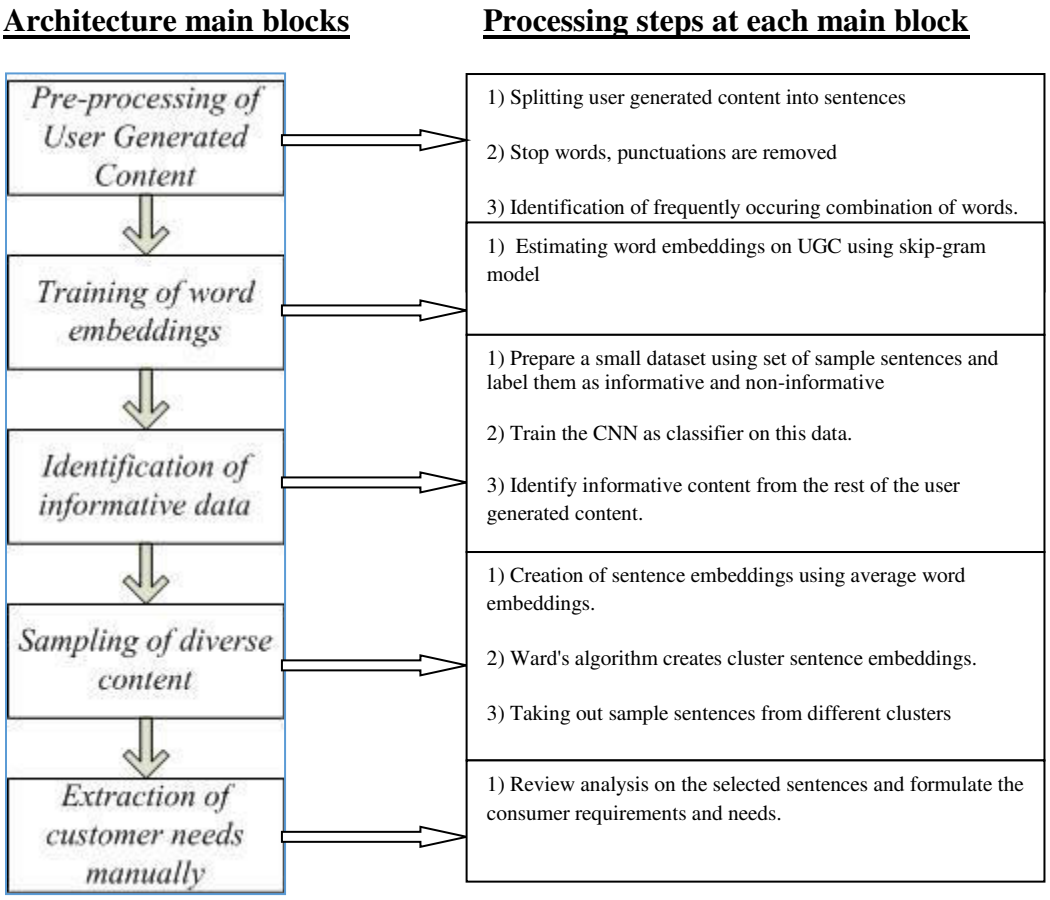

Figure 1. Architecture of customer requirement analysis using user generated content.

Experimental Design

Dataset
Dataset in the form of experimental interview transcripts, sample of some product or service review sentences can be taken from a professional marketing consulting company.

Evaluation Measures
Comparison of efficiency, delays, expensive incurred from machine learning method and traditional methods.

Software and Hardware Requirements

Python will serve the purpose along with deep learning libraries like tensorflow, and Keras. Tools like Anaconda can be used. For training purpose, NVDIA GPUs are helpful in terms of hardware.