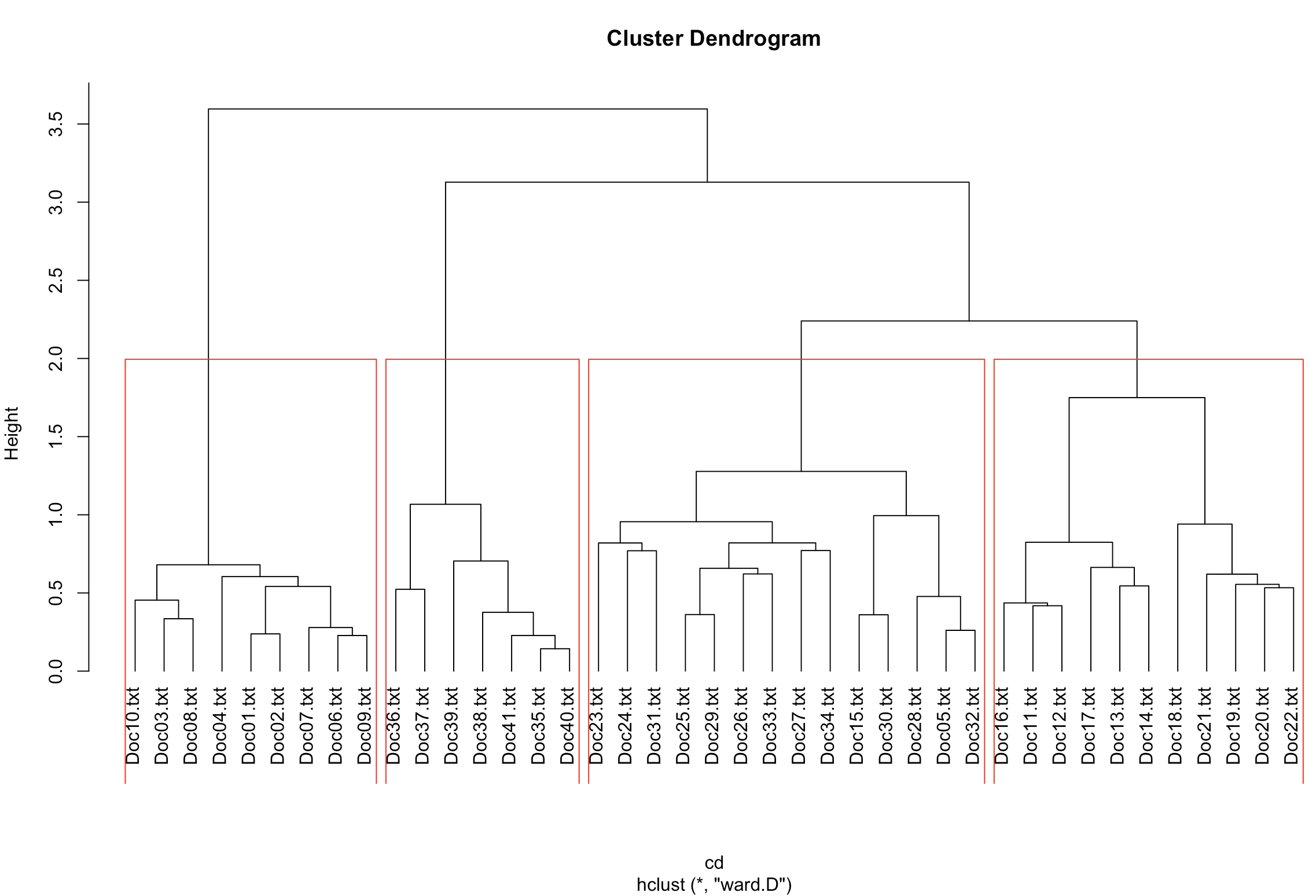
Coming from an e-commerce background, a customer-obsessed company, that takes the word of its customer into consideration for every decision taken, I see the huge varied implications that text mining and analytics potentially hold for e-retail. In a recent event wherein, the company announced a rise in the minimum wage for its employees, was followed by the quote from its CEO - "We listened to our critics, thought hard about what we wanted to do, and decided we want to lead." A quote that has rung a bell - "We listened to our critics".The company is certainly doing that, but how much of it? Is it when it is spoken by the customer first? Or is it when there is a group of customers speaking the same? Or is it when the press starts to talk about it? Or is it only when the stakeholders’ step in? It certainly cannot be that the company doesn't care about the customer's voice until the statement reaches mainstream media. What else could it be? Could it be that, even for a company of this magnitude, with all its human resources, is unable to infer the voice of the customer simply due to the overwhelming amounts of unstructured data that is being poured into its servers? Before we get there...

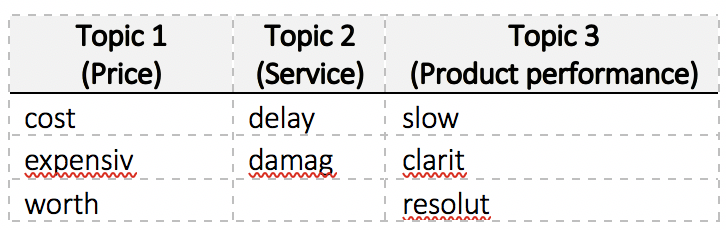
Sure, they have incredibly powerful product search engines, recommendation system, ad push systems, etc. It is how the e-retail sites improve the customers online shopping experience. You search for a shoe and their search engine throws up the search result with millions of products. From there you can narrow down to your choice by filtering through the brand, colour, style, material and what not! You name it, they got it! Once you add that splendidly good-looking brown leather shoe to your shopping cart and think of proceeding to the payment section is when the magic happens... boom... you are suggested to buy a matching pair of socks or a polish or if you are married - a shoe for your better half, perhaps! Of course, you knew you would have to get it eventually, but how did the system read your mind? You have the incredible minds of Machine Learning Gurus to thank for it (or curse for it if it is month end)! As simple as it may seem to be, it is not! Behind the web page where you type and hit ‘Enter’ there are potentially thousands of ML algorithms and neural network nodes taking decision based on outputs from complex statistical calculations, before putting a sock or polish on your screen for consideration. But, how did your text go to feed a statistical model? Have you not always related statistics to numbers with a mix of alphabets that never made any sense back in school? Well, you are not wrong completely! Text mining works by converting the unstructured text data into numerical distances which can then be statistically modelled using various unsupervised and supervised learning techniques that are available.

Getting back to my thought on how text analysis can be used to read customer feedback and take decisions based on that… Apart from the star ratings received by products online, e-retail companies also source a wealth of information from their customers in terms of customer feedback and review on the product website pages. Considering the amount of products beings sold online by some of these e-retail giants, it is not practical for someone to sit and read through each of the customer feedback. For scale, there were over 480 million products listed only in Amazon US back in 2015! First off, let us understand why it would be important for the company to understand all of its customers feedback, when they have 480 million products that are listed! Remember, to the company it is a 480 million listings, but for a customer who is looking to buy an item, say iPhone xs, it is just a single listing and considering the amount a single person is going to spend on that single purchase, one can be sure that the customer is going to read through every single customer feedback before buying that product.

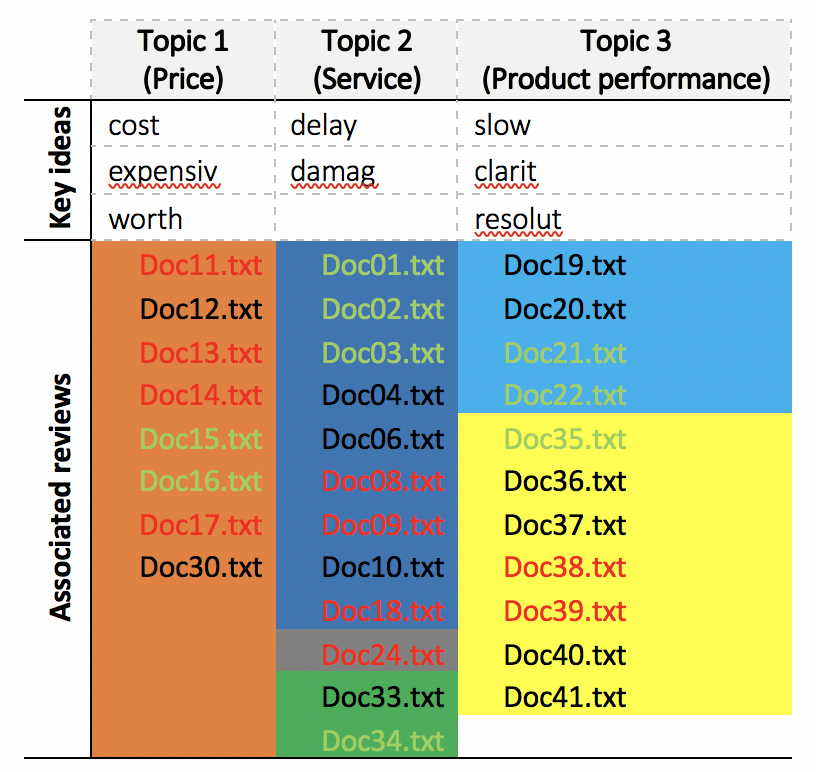
Now that we understand the importance for the company to understand its customer’s reviews, let us see how the company can go about decoding the vast amount of unstructured data. This is where the text analytics techniques such as clustering, topic modelling, network graphing, etc. are going to come in handy. First, the process can begin by simply querying all the customer reviews from their datawarehouse along with metadata on the customer ID, details of purchase such as purchased/not purchased/returned, product of purchase, ID of the seller (multiple vendors could be selling the same product and the customer will have the option to pick a seller of choice based on factors like delivery time, discounts, offers, etc.) who delivered the product to the customer, average rating of the product/seller, etc. With these now available in a dedicated database, it is time to get busy with text analytics! First off, all the customer reviews will have to be converted into single or multiple corpus (seller, product, customer, etc.). This will have to be followed by cleaning of data, with the cleaning conditions periodically supervised and updated. The cleaned data can then be converted into a DTM or TDM, on which we can perform one of the many available clustering techniques. Although the clusters are not going to give us any information on the nature of a cluster themselves, the following methods will start adding in colours to the clusters. Below we see hierarchical clustering applied on a set of 41 customer reviews.



Next step would be to proceed to topic modelling, which would give us the topics widely split under a desired number of categories such as the follows –



We would also have each review classified under a particular topic based on the probabilities of the keywords. Further, network graphing could help identify reviews that are closely associated and enable us to draw classification of reviews. Finally, sentiment analysis of each review could add in colours like green for positive and red for negative reviews.



Now this makes it easier and saves a whole lot of times in segregating the comments and focus on the ones that are in red and might need immediate attention!

End of the day, even if Amazon did use our text analysis model and get to this comment, too bad they are not going to be able to get some wild wolves to act in the movie unless you want a havoc at the studios! :P

