Detecting Redundancy in Reported CGM Device Issues

By: Iris Dai, Chuangji Li, Srihita Nangunuri, Nandini Neralagi
Project Advisor: Peter Freeman

Background & Introduction

Continuous glucose monitoring (CGM) is a new technique that has revolutionized the tracking of people's blood sugar levels. Dexcom, a prominent glucose monitoring device company, is helping to lead this revolution. To track reported issues with its devices, Dexcom uses the Jira software platform, which records the issues in the form of *tickets*. However, many of these tickets are redundant, with the same issues being recorded multiple times. Our solution to this problem is to group related tickets together using statistical models and natural language processing methods.

Issue_Key	Status	Created	Os	Description	Summary	COMMENTS
RSP-1099	Closed	2022-07-20 13:0		email quotegood morning ruth work remote yesterday leave phone message desk phone review request vendor package organization	albany med health request completion vendor onboarde form	await ruth send p
RSP-1221	Closed	2022-09-30 12:2	iOS	ensure pt attempt wifi cellular pt institutional wifi issue happen connection firewall device dnd screen time block connection unsure cause	g6 app phone get incompatible message server error app launch despite compatible device	
RSP-1297	Closed	2022-11-08 6:44		christine wozniak ask feedback hi timothy hcp question patient tir report < 1 low daily detail show alert show time low 0.5 hcp state pt report hypo	cams clarity clinic report pt experience hypo event daily overlay	hi ~cxw20918
				hello review report clarity page device patient show agent notice test account dexcom app show dexcom g6 mobile app unidentified clinic report issue clinical		

Figure 1: Examples of three CGM Jira Tickets with fields Issue Key, Status, Date Created, OS type, Description, Summary, and Comments

Data Pre-Processing

Our dataset comprises a collection of 10,011 reported CGM device issues. For each, there are 16 associated fields, including the following:

`Status`	tracks the progress of the ticket through the workflow ("open", "closed", "awaiting action")
`OS`	denotes the operating system concerned
`Description`	the body of the ticket description
`Summary`	the title of the ticket
`Comments`	detailed discussions or notes added to the ticket

We conducted a detailed cleaning process to improve the usability of text (**`Description`**, **`Summary`** and **`Comments`**). Initially, we removed HTML tags and addressed formatting issues. Subsequently, we normalized text through lemmatization (a process that simplifies words to their base or root form) and removed stopwords, punctuation, and excess spaces. Further, with the removal of replicated issues using **`Issue_Key`** as an identifier, we reduced the dataset to 1,673 unique tickets.

Methods

- Using the **`Summary`** column primarily, we grouped the text based on the frequency of keywords. To do this, we used N-grams, a contiguous sequence of words of length N, to find the most frequent series of words.
- We converted the text inputs (words) into vectors of numbers using a technique called Word2Vec. These vectors contain information about the relationships between the N-grams that is not observable, or that is latent.
- The K-means algorithm then assigned the elements in the vectors to different clusters based on the distance between elements.
- The clusters were then visualized with t-SNE, with a perplexity value of 50 and 100. The perplexity value controls the effective number of neighbors that each point considers when reducing dimensionality, and we found that 50 and 100 were optimal for our clustering, based on data size.
- Clustering was finally evaluated using the Silhouette Score, which ranges from -1 to 1, with greater than 0.5 being optimal.

Analysis and Results

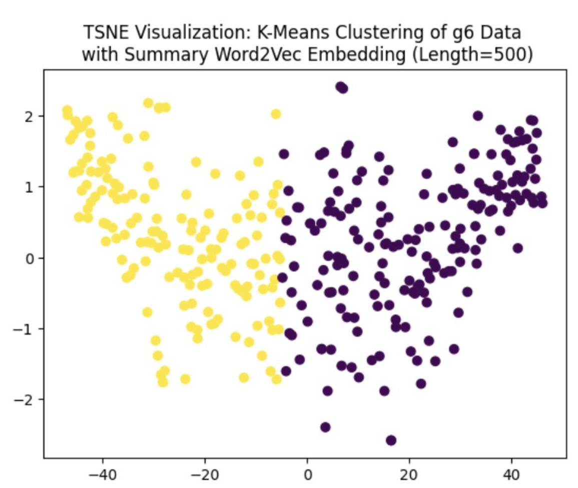

Figure 2: t-SNE visualization (perplexity=50) of G6 data. The Silhouette Score is 0.518, indicating good clustering.

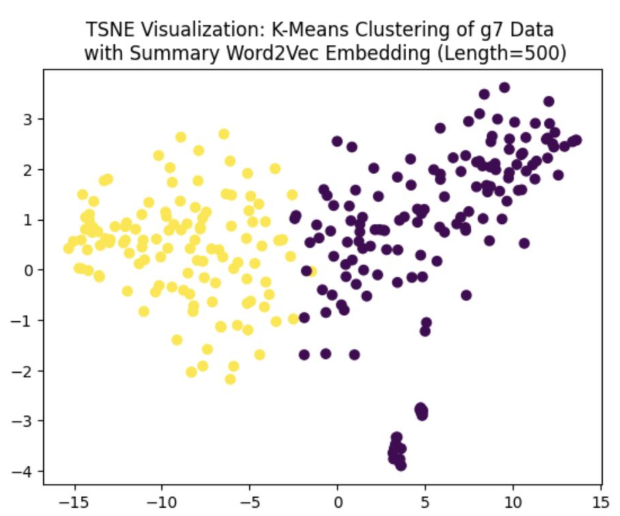

Figure 3: t-SNE visualization (perplexity=50) of G7 data. The Silhouette Score is 0.374, indicating fair clustering.

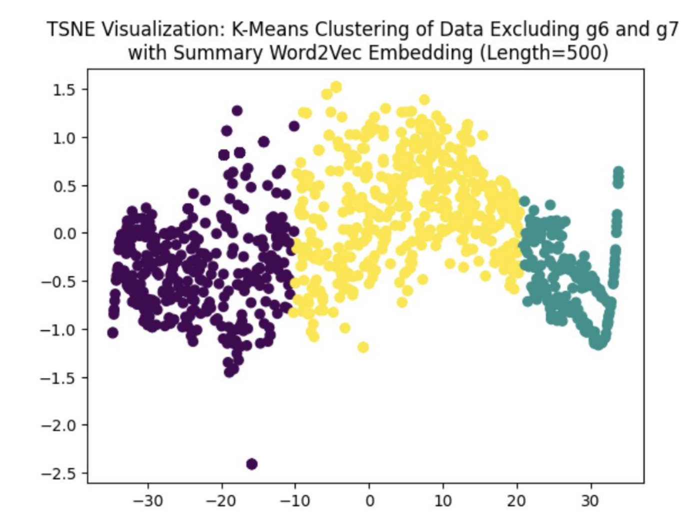

Figure 4: t-SNE visualization (perplexity=100) of remaining data. The Silhouette Score is 0.512, indicating good clustering.

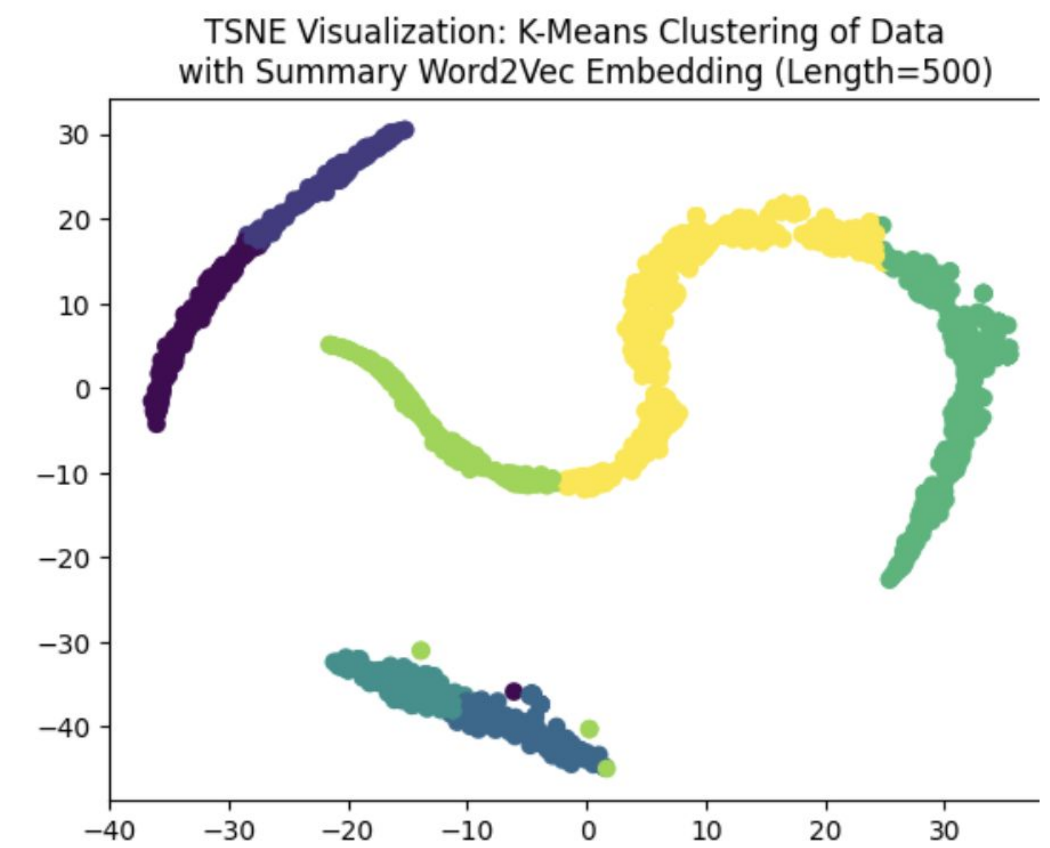

Figure 5: t-SNE visualization (perplexity=50) of all tickets Silhouette Score: 0.475. Top left clusters represents G6 data, bottom-most clusters represents G7 data, and three-colored center clusters represents other data.

Top 5 Most Frequent N-Grams By Cluster

N-Gram Cluster 0	Cluster 0 Frequency	N-Gram Cluster 1	Cluster 1 Frequency	N-Gram Cluster 2	Cluster 2 Frequency	N-Gram Cluster 3	Cluster 3 Frequency	N-Gram Cluster 4	Cluster 4 Frequency	N-Gram Cluster 5	Cluster 5 Frequency	N-Gram Cluster 6	Cluster 6 Frequency
g6 app	37	g6 app	77	g7 android	43	g7 app	41	clarity clinic	138	clarity clinic	60	clarity clinic	178
g6 ios	23	mobile app	25	g7 ios	38	g7 ios	28	follow app	41	merge request	19	clinic request	70
g6 android	20	g6 mobile	24	bug risk	24	g7 android	27	clinic request	37	clarity clinic merge	18	clarity clinic request	65
g6 mobile	12	g6 android	23	bug risk rating	24	signal loss	13	clarity clinic request	35	clarity clinic merge request	18	cams clarity	50
jp g6	11	g6 mobile app	22	bug risk rating g7app	24	android ous	10	error message	26	clinic merge	18	cams clarity clinic	44

*Figure 6: Top 5 N-grams for each cluster, from N=2 to N=10 (bi-grams to deca-grams). Clustering was aggregated from each data split. **G6 Data**: Cluster 0, Cluster 1; **G7 Data**: Cluster 2, Cluster 3; **Remaining Data**: Cluster 4, Cluster 5, Cluster 6*

Cluster Interpretation: Cluster 0 covers G6 system issues, like device compatibility, software glitches, and UI concerns, while Cluster 1 seems to focus primarily on G6 app-specific problems like compatibility errors and signal loss. Cluster 2 focuses on the G7 app's bug risks across devices while Cluster 3 is more concerned about on G7's signal loss and connectivity issues. Cluster 4 focuses on patient account and Clarity app issues. Cluster 5 covers merge requests and system functionalities, while Cluster 6 centers on administrative tasks and user account management within Clarity.

Conclusions

- Splitting the data into multiple datasets according to device type creates stronger and more compact clusters
- A total of seven clusters yields an aggregated Silhouette score of 0.475, where a Silhouette score of 0.5 or higher is considered optimal
- Further work could focus on improving the metric score by including more data or using more complex clustering methods
  - Different clustering algorithms such as Gaussian Mixture Models or Balance Iterative Reducing and Clustering using Hierarchies (BIRCH) might account for features of the data that was missed by the methods used here