Optimizing IT Operations with Natural Language Processing

Natural Language Processing techniques can help IT Operations teams gain enhanced understanding of their operations landscape and accelerate optimization of ticket management

Executive Summary

A significant portion of any IT operation is maintenance and support of applications and infrastructure. In such an engagement, every problem or request is initiated as a ticket that is worked upon by the operations team. In large operations, these tickets could run into very large volumes. This necessitates continuous optimization of the management of these tickets to keep the operational costs under control.

Traditional approaches to identify optimization opportunities require significant manual effort from Subject Matter Experts (SMEs), which turns out to be a big inhibitor in the continuous improvement journey. Natural Language Processing (NLP) techniques help overcome this challenge, resulting in cost optimization as well as service level improvement.

Challenges in identifying the right treatment strategies

Typically, end users describe their problem or request in free text as part of the ticket description. Although the range of issues are limited, given that the underlying application portfolio has a defined functionality, the users often describe similar, or even same, problems in many different ways.

To enhance efficiency, the operations teams use many different strategies such as

- Automating resolution of tickets.
- Documenting series of steps required to resolve the tickets, known as Standard Operating Procedures (SOPs).
- Assigning less experienced or right skilled personnel to the tickets, a strategy known as "shift-left".
- Eliminating the underlying problem and hence recurrence of such tickets, resulting in reduced ticket volumes in future.

Identifying and deploying the right treatment strategy requires SMEs to go through the ticket descriptions and develop an understanding of the ticket patterns. Such an approach is effort intensive and, when done manually, could be both inefficient and sub-optimal.

At times, SMEs define a dictionary of keywords or phrases corresponding to different categories of tickets. An automated program looks for presence of these keywords in the ticket description to map tickets to categories. However, this approach has its own challenges. One, such a dictionary of keywords is never comprehensive and requires constant updates. Two, mere presence of a keyword is not adequate to map a ticket to a category. Three, this approach allows mapping to a set of pre-defined ticket categories only.

Leveraging NLP to identify patterns in tickets

NLP can help address some of the challenges discussed above, by automatically assessing the free text descriptions across thousands of tickets. To accomplish this, a vector representation or embedding[1], of ticket descriptions is created using the word2vec[2] technique. The ticket description and other relevant characteristics of the tickets are used to automatically identify patterns and come up with homogeneous groupings or clusters of tickets. Leveraging these insights, SMEs can not only identify the right treatment strategies but also develop an enhanced understanding of the ticket landscape.

This approach eliminates the need for SMEs to go through individual ticket descriptions or define a comprehensive keyword dictionary. It can also handle the different ways in which different users describe the same ticket.

The deployed solution involves three key steps

- Automated processing & clustering[3] of tickets.
- SMEs to identify treatment opportunity, and execute on that such as preparing an SOP.
- Assigning every new, incoming ticket to the right cluster for appropriate processing.

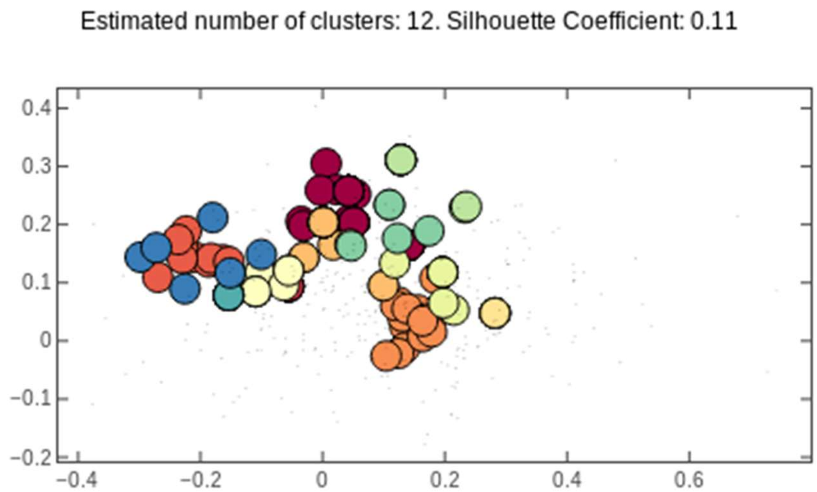
Automatic identification of semantically similar tickets

Table 1 illustrates how three similar tickets, articulated differently by the users, are grouped together under cluster number 2 without using any pre-defined keywords.

S. No.	**Ticket Description**	**Assigned Cluster**
1	Sup : SMP : Unable to launch ILT course --- FAILED	2
…	…	…
…	…	…
102	SUP:SMP:user cannot finish a training	2
…	…	…
3013	SUP:SMP:Problem completing a web-based learning activity	2
…	…	…

Table 1

Figure 1 shows a scatter plot of tickets, color coded by cluster, along with the optimal number of clusters suggested by the DBSCAN algorithm and a measure of cluster quality called silhouette coefficient[4]. The plot axes are not directly meaningful and are used only to show cluster separation.



Interaction diagram – clustering tickets in the repository:

In Figure 2 the sequence of actions needed to cluster or separate tickets into groups by similarity is shown.

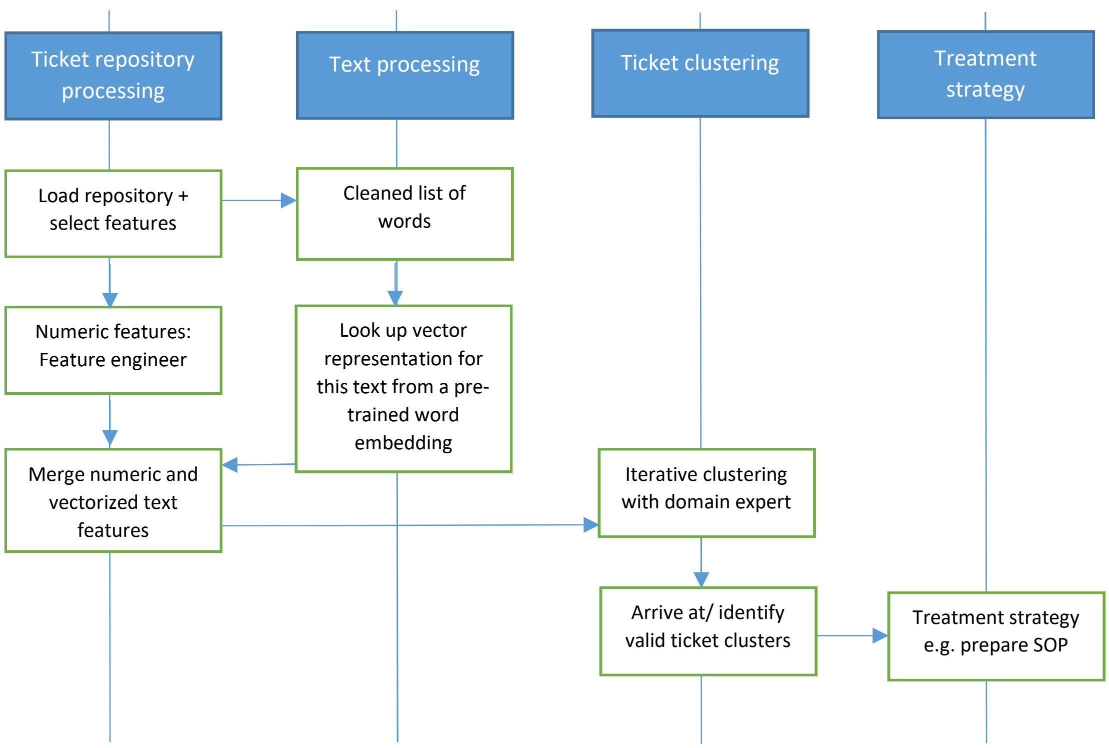


Figure 2

Interaction diagram – assigning new tickets to a cluster

Figure 3 shows the process to assign a new ticket to one of the previously calculated clusters. The output includes a similarity score calculated using cosine similarity[5].

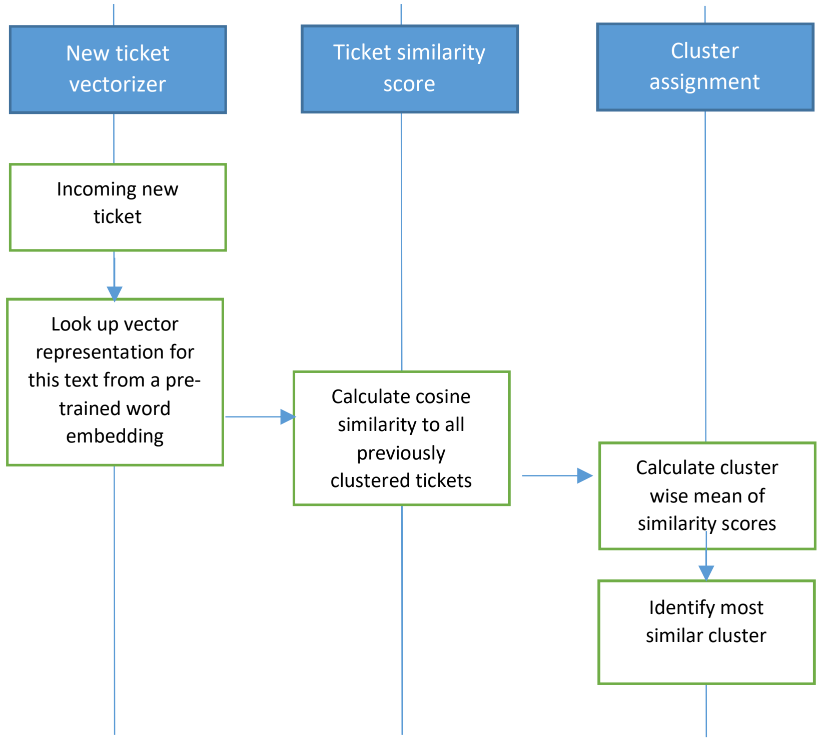


Figure 3

Solution architecture

The solution is built in Python using open source libraries like NLTK, BeautifulSoup, scikit-learn and genism (word2vec implementation) and is shown in Figure 4.

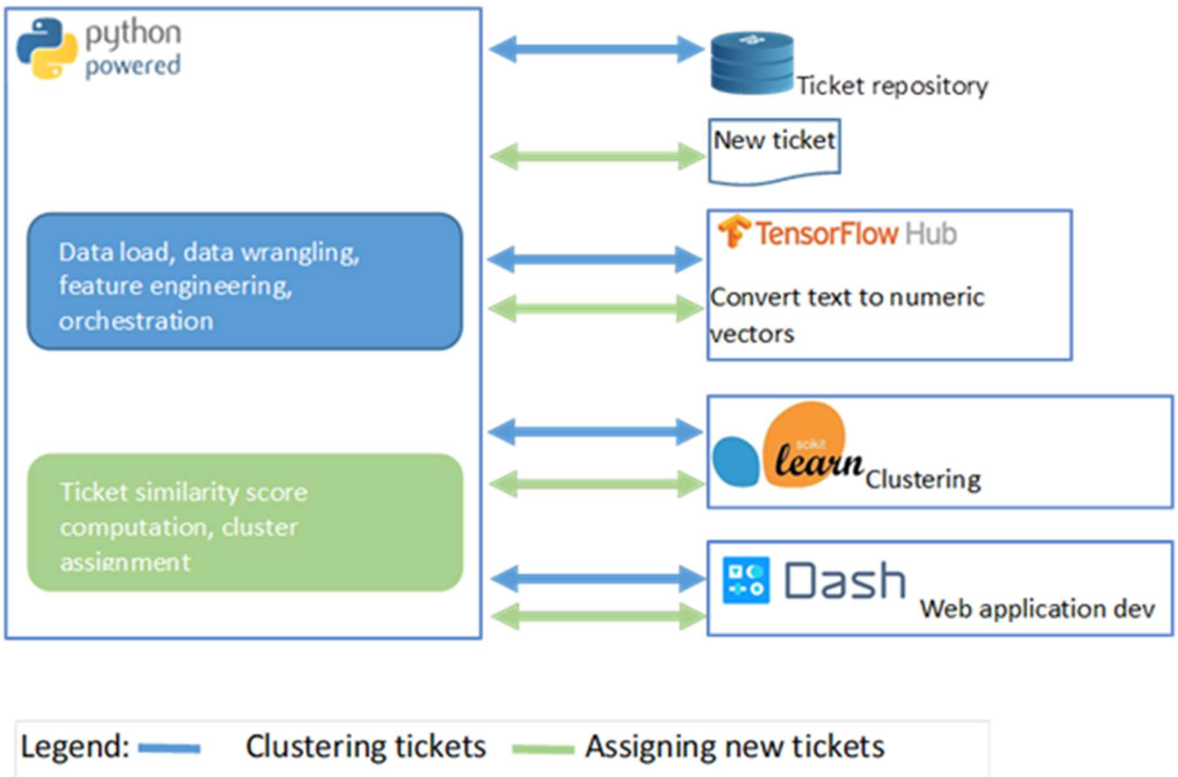


Figure 4

Opportunities Ahead

The ability to automatically process the free text ticket description, without the need for SMEs to laboriously go over each ticket or identify a set of keywords to categorize them, opens up many opportunities.

We have successfully piloted this solution in engagements across Industry segments. For an engagement with a Life Sciences client, some of the realized benefits included

- 84K USD annual savings through effort reduction and redeployment.
- First Level Resolution (FLR) improvement by 16 percentage points.
- Mean Time To Resolve (MTTR) reduction by 40%.

The solution also enabled identification of automation and problem management opportunities across engagements, which will eventually lead to non-linear cost savings.

We see other potential opportunities like insights to enable optimal team staffing and prioritizing knowledge transfer from the incumbent vendor during the transition for a new engagement. The solution can be extended for use in other scenarios that require processing of free form text such as identifying patterns in software defects, identifying duplicate incident alerts etc.

Footnotes

1. Word embedding: "Vector space models (VSMs) represent (embed) words in a continuous vector space where semantically similar words are mapped to nearby points ('are embedded nearby each other')." Source: https://www.tensorflow.org/tutorials/representation/word2vec
2. Word2vec: "Word2vec is a particularly computationally-efficient predictive model for learning word embeddings from raw text." Source: https://www.tensorflow.org/tutorials/representation/word2vec
3. Clustering: "Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters)." Source: https://en.wikipedia.org/wiki/Cluster_analysis
4. Silhouette coefficient: The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). Source: https://en.wikipedia.org/wiki/Silhouette_(clustering)
5. Cosine similarity: Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. Source: https://en.wikipedia.org/wiki/Cosine_similarity

About the Authors