August 2021

DETERMINING LIKE PEERS AND DOMINANT FEATURES USING MACHINE LEARNING

Xinjun Zhang

Doosan Jung

Qihong Shao

Kevin Broich

Yue Liu

See next page for additional authors

Inventor(s)
Xinjun Zhang, Doosan Jung, Qihong Shao, Kevin Broich, Yue Liu, and Gurvinder Singh

DETERMINING LIKE PEERS AND DOMINANT FEATURES USING MACHINE LEARNING

AUTHORS:
Xinjun Zhang
Doosan Jung
Qihong Shao
Kevin Broich
Yue Liu
Gurvinder Singh

ABSTRACT

Peer comparison is one of the most desirable features for network customers. One of the most frequently asked questions is, "How does my company's network performance compare with that of my peers?" To provide effective peer comparison results there are two fundamental questions that must be resolved – the first question concerns finding the most similar peers and the second question addresses understanding why the peers are similar. To address these types of challenges, techniques are presented herein that leverage machine learning (ML) models to resolve the two fundamental questions that were described above. Aspects of the presented techniques encompass an end-to-end system, which for convenience may be referred to herein as "DeepSense," which resolves the entire lifecycle mystery of peer comparison. Additionally, aspects of the presented techniques employ a singular value decomposition (SVD) algorithm to define similarity among customers in a way that is able to overcome the limitations that are caused by latent information. Further, aspects of the presented techniques leverage non-negative matrix factorization (NMF) to capture the dominant features which can influence the similarity among peers. Still further, aspects of the presented techniques support a user-friendly customer interface in real working production systems.

DETAILED DESCRIPTION

Peer comparison is one of the most desirable features for network customers. One of the most frequently asked questions is, "How does my company's network performance compare with that of my peers?" The answer to that question has a significant influence on a customer’s business decisions (e.g., whether to upgrade their services, buy more

network devices, renew a license, etc.). A good peer comparison framework may be leveraged by numerous areas within a customer (including, for example, marketing, sales, customer service, and all of the related departments) and bring significant benefits to the customer.

To provide effective peer comparison results, there are two fundamental questions that must be resolved. As illustrated in Figure 1, below, the first question concerns finding the most similar peers and the second question addresses understanding why the peers are similar. After resolving these two critical questions the most similar networks may be used in various applications (e.g., a marketing team can generate different reporting, a sales team can use the comparison result to upsell and motivate, a service department can find more relevant problem resolutions, etc.).

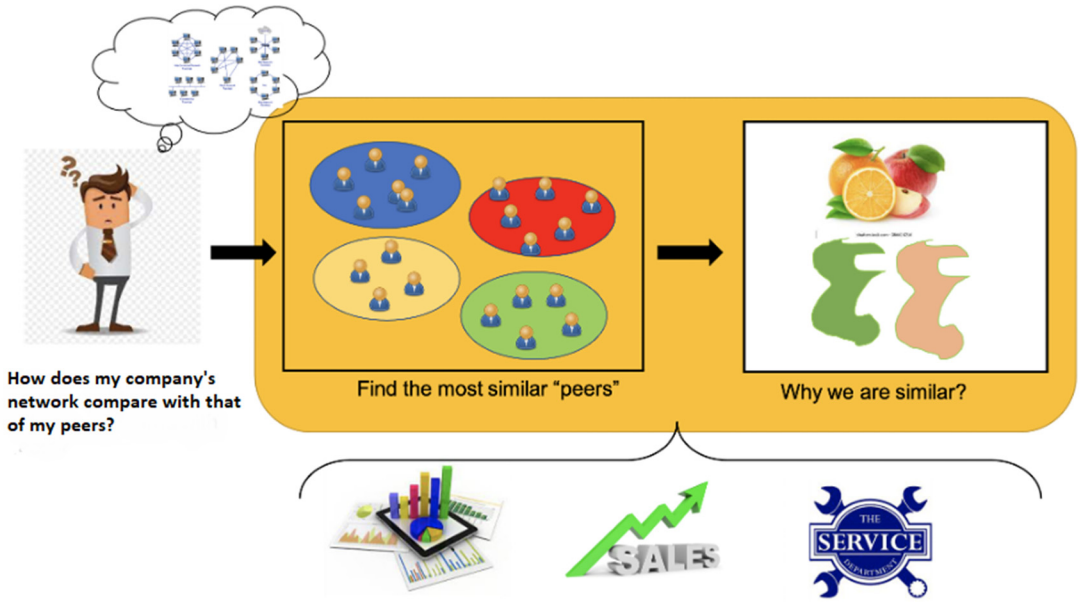

Figure 1: Peer Comparison Fundamental Questions

A customer's network profile may be composed of comprehensive information covering a number of areas, including, possibly among other things:

- Company information. For example, the company's name, the company's business category (such as bank, hospital, etc.), the number of end customers (e.g., the number of users for a bank, the number of patients in a hospital, etc.), etc.

- Network information. For example, the network topology, the number of devices (such as routers, switches, etc.), location, active licenses, etc.
- Device information. For example, product family, device type, etc.
- Current network status/issues. For example, security concerns, active tickets, software bugs, known issues, device end of life, license expiration, etc.

With such a complicated customer network profile, it is not trivial to answer the two fundamental questions that were noted above (i.e., finding the most similar peers and understanding why the peers are similar). Various of the associated challenges are briefly described below.

For the first question (i.e., how may the most similar peers be found?), considering all of the complicated information identified above in the customer profile, existing systems have difficulty finding the most similar peers. Various of the factors that contribute to that difficulty are briefly described below.

First, simple heuristics will not work. Existing systems try to use simple heuristics (e.g., grouping customers based on their business category, or by product family (router, switch, etc.), or by size, etc.). The results will not be accurate. For example, Bank A and Bank B may both be financial institutions, but how they arrange their networks could be totally different, based on, among other things, their specific budgets, etc. Treating a worldwide bank as a similar peer to a small local bank does not make sense.

Second, manually selection is not scalable. Individual people may try to manually find the most similar customers, but such an approach is not scalable. For example, it takes a considerable amount of time and effort to manually check customer profiles and to try and figure out a list of most similar networks.

Third, hidden or latent data is hard to capture. Some information in a customer profile (such as, for example, company information, device information, current network status, etc.) is visible while other information (such as, for example, network topology, architecture, etc.) is not easy to capture directly. Furthermore, certain information (such as, for example, a customer's preference) is hidden and thus latent data, which is impossible to capture. An interesting analogy involves movie review information. For a given movie and different customer reviews, the results are essentially based on a number of hidden

variables (e.g., customer preference, etc.) not simply by the customer's age, gender, education, etc.

For the second question (i.e., how may the similarity among customers be interpreted?) existing systems do not provide a transparent interpretation, they only provide a relative score or ranking of the peers based on heuristics (e.g., based on whether a customer is in a similar business category, or has similar devices, etc.).

Thus, an interpretable model is highly desirable as it enables a customer to understand, for example, what the key aspects are and how they compare with their peers.

To address the types of challenges that were described above, techniques are presented herein that leverage machine learning (ML) models to resolve the two fundamental questions that were described above. Aspects of the presented techniques encompass an end-to-end system, which for convenience may be referred to as "DeepSense," which resolves the entire lifecycle mystery of peer comparison. Additionally, aspects of the presented techniques employ a singular value decomposition (SVD) algorithm to define similarity among customers in a way that is able to overcome the limitations that are caused by latent information. Further, aspects of the presented techniques leverage non-negative matrix factorization (NMF) to capture the dominant features which can influence the similarity among peers. And finally, aspects of the presented techniques support a user-friendly customer interface in real working production systems.

Figure 2, below, depicts elements of the DeepSense framework and illustrates aspects of the workflow within such a framework.

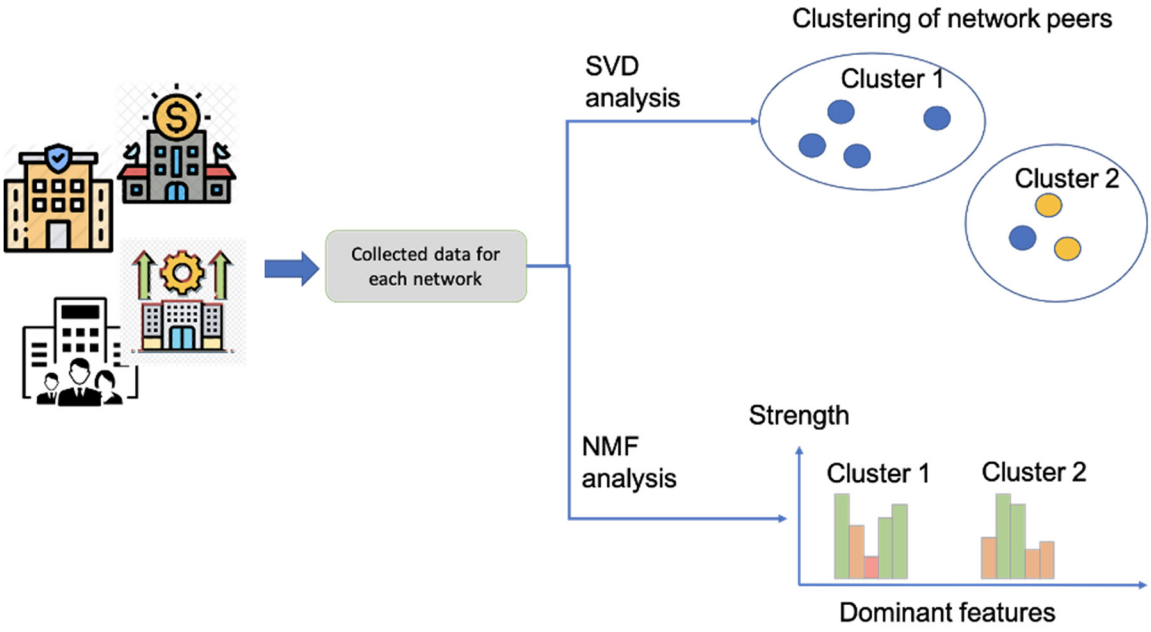

Figure 2: Illustrative DeepSense Framework

As depicted in Figure 2, above, as part of the DeepSense framework customer network data is collected. That data may include, as noted previously in connection with a customer's network profile:

- Company information. For example, the company's name, the company's business category (such as bank, hospital, etc.), the number of end customers (e.g., the number of users for a bank, the number of patients in a hospital, etc.), etc.
- Network information. For example, the network topology, the number of devices (such as routers, switches, etc.), location, active licenses, etc.
- Device information. For example, product family, device type, etc.
- Current network status/issues. For example, security concerns, active tickets, software bugs, known issues, device end of life, license expiration, etc.

The DeepSense framework comprises a "finding similar networks" module that employs SVD analysis, as depicted in Figure 2, above, to develop a list of most similar peers considering latent information. This functionality will be described and illustrated in the narrative below in connection with a first step of the DeepSense framework.

The DeepSense framework comprises a “discover dominate feature” module that employs NMF analysis, as depicted in Figure 2, above, to extract the dominant features and understand why they are similar. This functionality will be described and illustrated in the narrative below in connection with a second step of the DeepSense framework

Within the DeepSense framework a first step of processing data encompasses finding the most similar customers considering latent information. Among other things, aspects of this step leverage an SVD algorithm, as will be discussed in further detail below.

A challenge in defining similarity among customers is identifying all possible informative data (e.g., features). The more data that is available, the more accurate will be the result. However, it can be difficult to obtain large amounts of data in some instances. For example, customer networks are very complicated and quite varied, in terms of business categories, network architecture, total number of devices, the number of each type of device, the running software version and hardware version, different configurations and reported defects, security risks, and so on. Among all of this information, some is easy to collect and the measurement is straightforward (for example, the total number of devices, the number of each type of device, etc.), while other features are hard to collect and impossible to measure quantitatively (for example, customer preferences, etc.).

This necessitates the collection and use of only those visible features and treating the rest of the data as latent (i.e., invisible). Such a strategy can reduce the level of difficulty, but it can also raise the risk of an inaccurate measurement of similarity. Fortunately, this limitation can be addressed by the SVD and NMF algorithms.

As illustrated in Figure 3, below, the basic idea of an SVD algorithm comprises factorizing a matrix 'M' (of dimensions m*n) into the form 'USV', where 'U' is an m*m matrix, 'S' is an m*n rectangular diagonal matrix with non-negative real numbers on the diagonal, and 'V' is an n*n matrix.

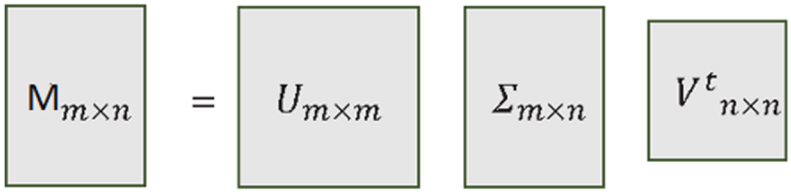

Figure 3: Singular Value Decomposition

Applying the example of user movie ratings to the model that was described above, the matrix 'M' would be the rating score matrix of customers over movies. The resulting matrix U basically describes the characteristics of each customer, the matrix V represents the characteristics of each movie, and the diagonal entries of matrix 'S' are known as the singular values of matrix 'M'.

Within the DeepSense framework, the matrix 'M' is defined where each row is the unit of network devices, which can be, for example, an entire company, or a subsidiary or branch of a company, or just a local office building. The columns of matrix 'M' are the measurements, which describe different aspects of a network (such as, for example, security issues, compliance, network management best practices, license expirations, hardware version, software versions, the total number of devices, the number of each type of device, reported bugs and defects, etc.). By applying an SVD algorithm to matrix 'M' it is possible to derive the profile matrix for the network unit (e.g., company, branch, or single office building) as well as the profile matrix for each type of measurement metric.

The first step of processing within the DeepSense framework may further encompass performing a peer comparison among devices with different network roles. The similarity of each pair of a network unit (e.g., company, branch, local office, etc.) may be calculated using profile matrix 'U'. The dimension for each unit is 'm', but here, only the first two dimensions are used (i.e., the top two principal components which preserves most of original information). Using the first two dimensions, it is possible to easily visualize the clustering of each network unit (as illustrated in Figure 4, below).

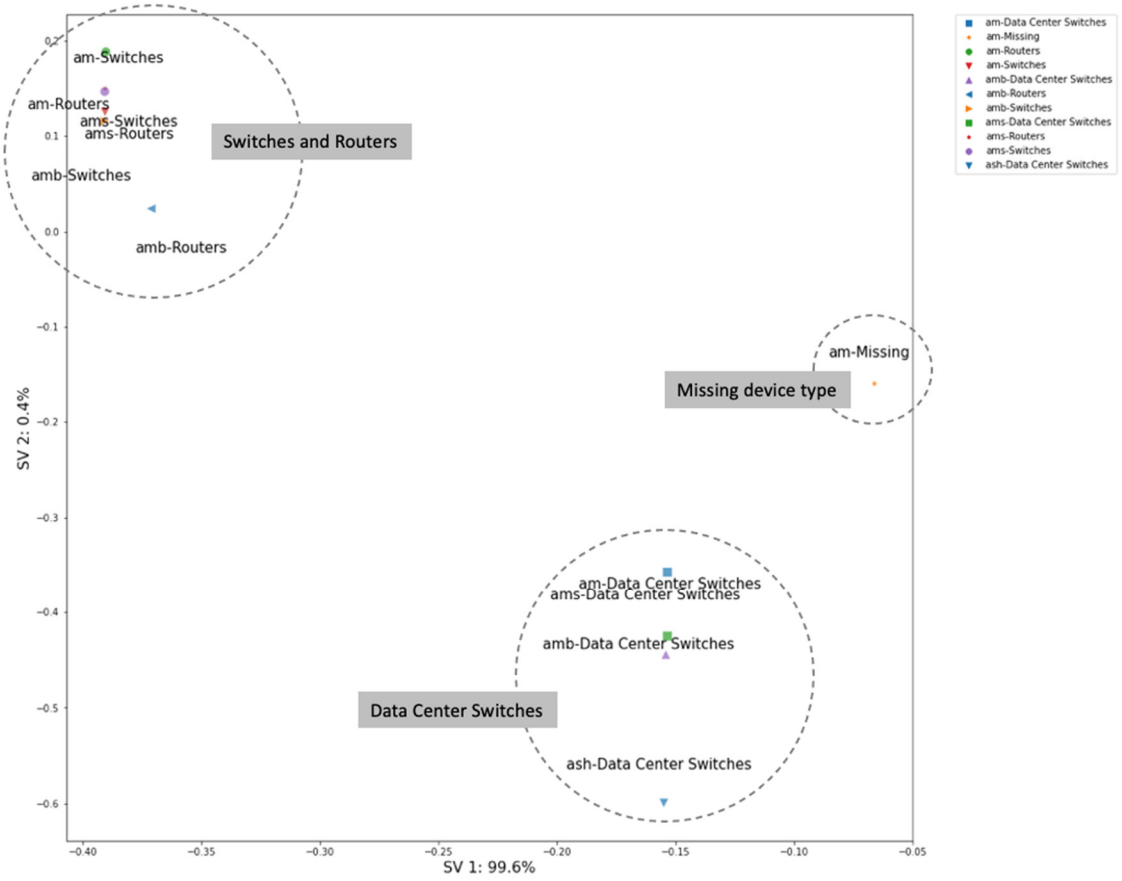

Figure 4: Clustering Using First Two Dimensions of Matrix 'U'

Figure 4, above, illustrates the clustering result of network devices. In one particular experiment the data labels were not exposed to the SVD algorithm. In this example, there are a group of "Data Center Switches", a group of "Routers" and "Switches," and one data point without device types. The prefixes such as "amb", "ash," etc. are identifiers for different geographic locations.

Under aspects of the techniques presented herein, the algorithm generates a clear grouping of all "Data Center Switches" for all local sites. Additionally, all of the "Routers" and "Switches" are grouped together and the two data points without device types are located very far away from the two major groups. Meanwhile, within the two clusters all of the data points are located in very close proximity to each other. Thus, it is demonstrated that the algorithm, according to aspects of the techniques presented herein, can group similar items together and separate different items apart.

The first step within the DeepSense framework may further encompass performing a peer comparison among different industries.

In one particular experiment, aspects of the techniques presented herein were applied to perform peer comparison among different industries. Intrinsically, different industries have different business operation styles, network traffic patterns, etc. so by assumption the objective is to identify similarity and dissimilarity among different industries. As shown in Figure 5, below, the SVD algorithm identified "Financial services" and "Government" as similar peers, "Professional services", "Health Care", and "Manufacturing" as similar peers, and "Retail," "Service Provider," and "Education-Public/Private" as similar peers. It is important to note that it is possible to work with domain experts to verify all of the model results in detail and to collect their feedback to help to improve the models.

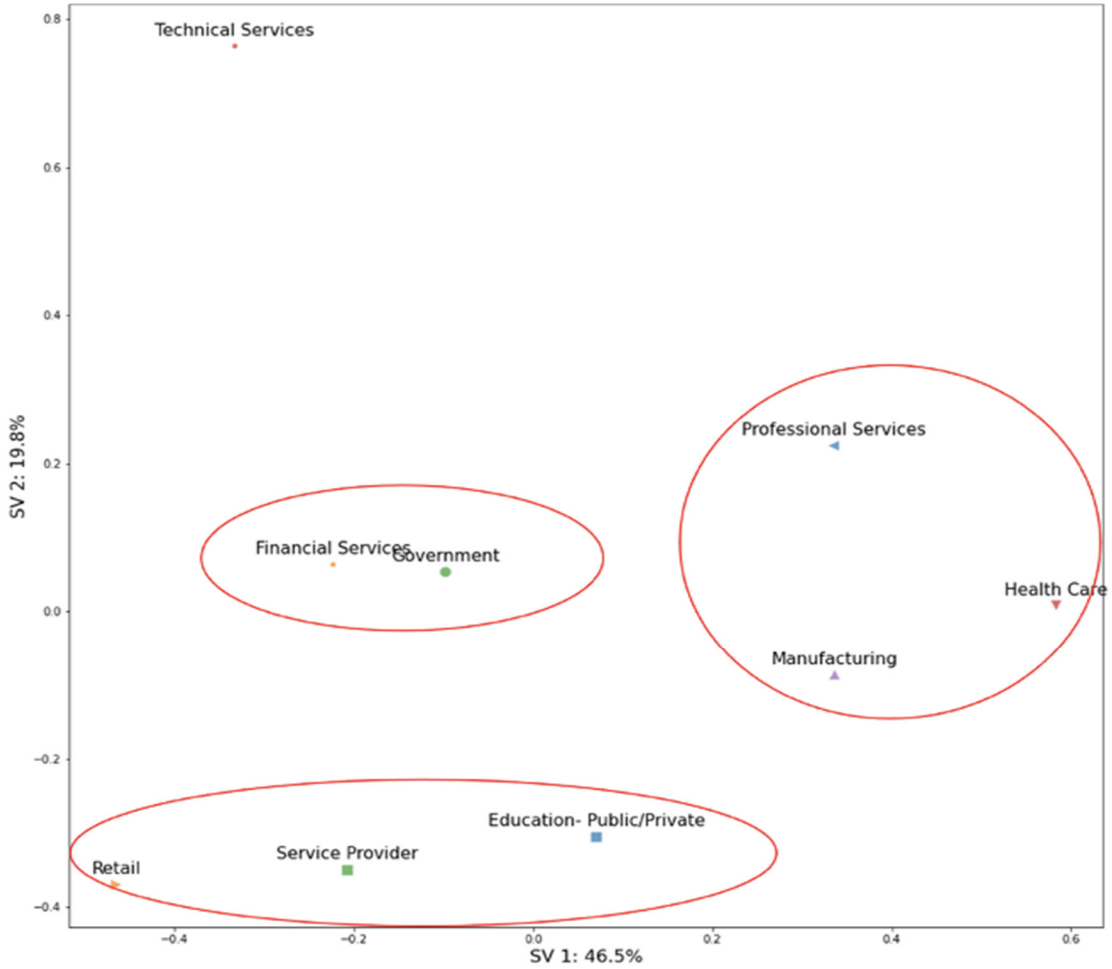

Figure 5: Illustrative Peer Comparison Among Industries

The first step of processing within the DeepSense framework may further encompass an application programming interface (API) for predicting top similar peers for a given customer.

A desired output from a peer comparison may include determining the most similar and dissimilar customers for a given customer. Here, the application is not restricted to comparing peers among different companies, but also providing a peer comparison among different branches and local offices within one customer.

Aspects of the techniques presented herein can provide the top 'N' or the bottom 'N' number of customers that are most similar or dissimilar. Figure 6, below, depicts an example of the similarity rank among all of the peers of the given "Financial services" customer. The indicated score is between [-1, 1] where 1 represents the most similar and -1 represents the most dissimilar.

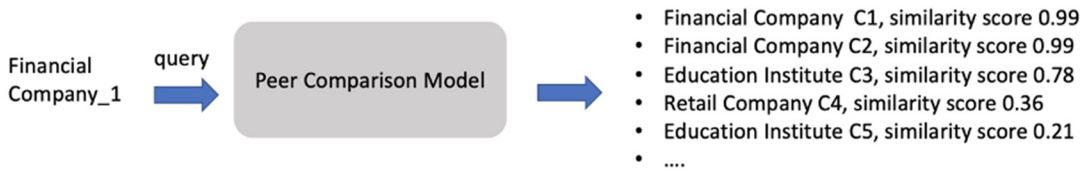

Figure 6: Rank of Similar Companies Among Customer's Peers

Within the DeepSense framework, a second processing step encompasses explaining customers' peers using dominant features. Among other things, aspects of this step leverage NMF, as will be discussed below.

In an NMF setting, algorithms are considered for solving the following problem – given a non-negative matrix 'V', as depicted in Figure 7, below, find two non-negative matrices 'W' and 'H' such that when 'W' and 'H' are multiplied together the result approximately reconstructs 'V'.

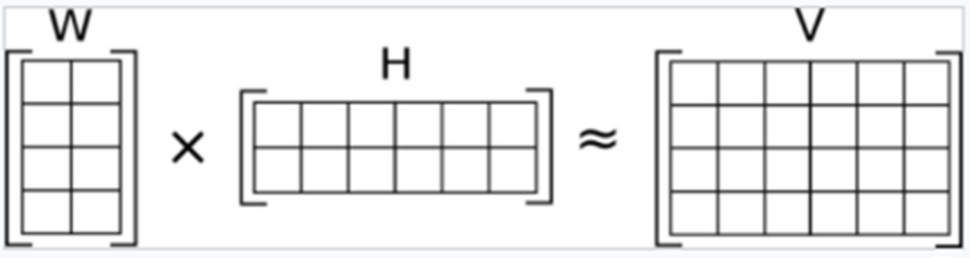

Figure 7: Exemplary Non-negative Matrix Factorization

The non-negative constraint often makes the resulting matrices easier to interpret because the whole (e.g., 'V') consists of parts (e.g., 'W' and 'H') and only "additive combinations" of the parts are allowed.

Under aspects of the techniques presented herein NMF can provide the parts-based representation of a network or customer. Also, for a given network or customer the importance of the parts can be obtained. Similar networks or customers will show similar importance (e.g., high or low) in similar parts. If a network or customer assigns high importance values on certain parts, these parts are considered to be the "dominant features," as they are the strongest characteristics or the main elements of the given network or customer. The dominant features determine the similarities within the group and make each group distinct from other groups.

The second processing step within the DeepSense framework may further encompass finding the dominant features in a peer group. The strongest characteristics or the main elements (i.e., the dominant features) that distinguish one group from other groups will be discovered during such an analysis. One particular experiment demonstrated that:

- The model is good enough to cluster the same customer's network into the same cluster when there are multiple measurements, each from a different timestamp.
- The factorized matrix provides a clue as to the reason why the customers were grouped together, based on the metadata that was used.

It is important to note that the customer names are annotated after the model groups the customers together. The experiment was set to treat the customer’s name as the ground truth. Additionally, the customer name information was not provided to the model when it factorized the original matrix.

Figure 8, below, shows the high similarity between multiple measurements of the same customer's network when a customer's metadata (such as its industry, sales territories, the number of devices per device type, etc.) was used. The names of the customers and the input feature names are anonymized to follow certain formats.

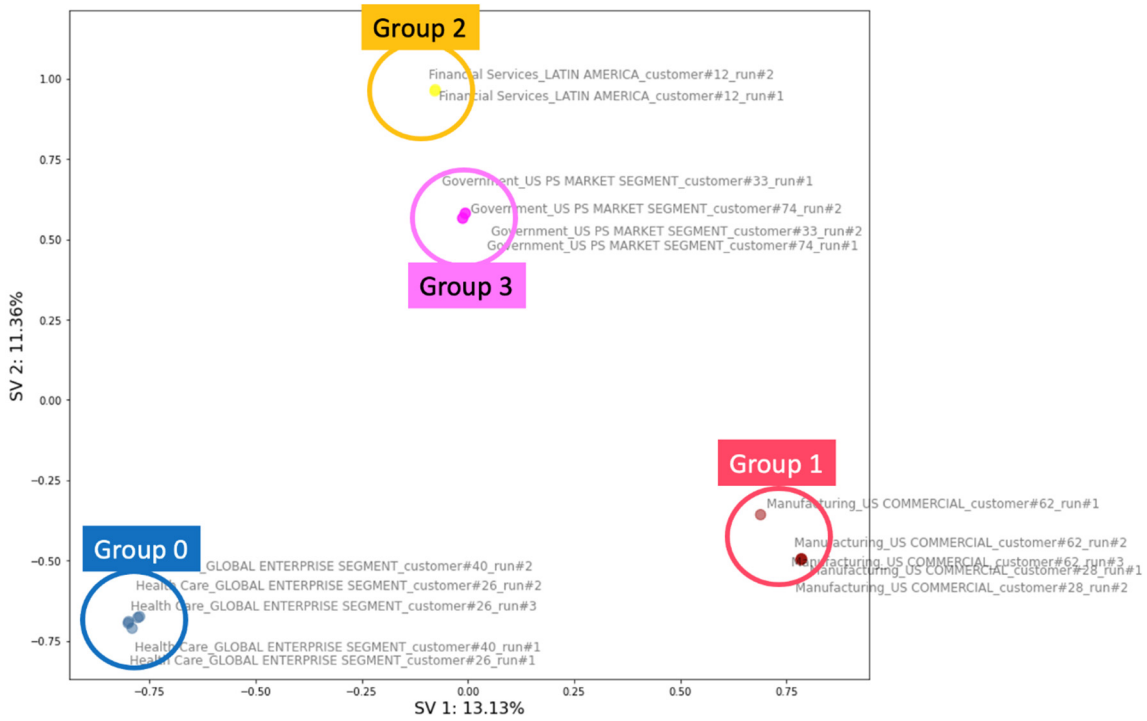

Figure 8: Exemplary Customer Group Clustering Results

With the given input data such as industry, sales territories, etc. the model, according to aspects of the techniques presented herein, can group the customers into four clusters (which are color coded blue, red, yellow, and pink as shown in Figure 9, below).

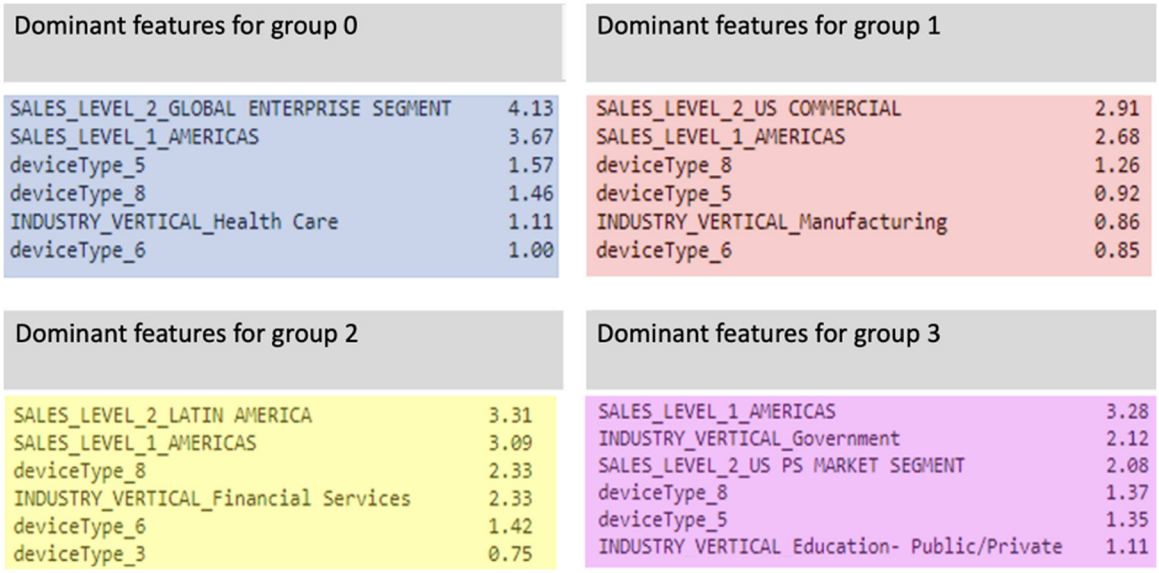

Figure 9: Illustrative Cluster Group Dominant Features

As illustrated in Figure 9, above, multiple measurements of the same customer are grouped together. For example, the "Health care GLOBAL ENTERPRISE SEGEMENT_26" (which is the anonymized customer name, where the number 26 represents the identifier of the customer) are clustered together in group 0, the "Manufacturing US COMMERCIAL_28" are clustered together in group 1, and the "Financial services LATIN_AMERICA_12" are clustered together in group 2. The model does not take the customer's name as the input, but it is able to correctly group the multiple measurements of the same customer into the same group.

Furthermore, the factorized matrix provides a clue about the reason why the model believes that a certain group of customers should be clustered together. Table 1, below, shows the dominant features that highlight each cluster. For example, in group 2 an emphasis is on the sales territory being Latin America and in group 3 an emphasis is on the industry being Government. In other words, sales territory being Latin America is one of the significant characteristics of group 2 while industry being Government is one of the significant characteristics of group 3.

Group	Dominant Features From H	Industry Composition in each Group	Sales Territory Composition in each Group

Group 0	Global Enterprise Segment sales territory, health care industry	Health care: 37% Manufacturing: 25% …	Global Enterprise Segment: 100%
Group 1	US Commercial sales territory, manufacturing industry	Manufacturing: 49% Professional Services: 31% …	US Commercial: 100%
Group 2	Latin America sales territory, financial services industry	Financial Services: 38% Service Provider: 10%	Latin America: 96% US Commercial: 2%
Group 3	US Government, Education	Government: 59% Education Public/Private: 36%	US Public Sector: 77% Canada: 21%

Table 1: Dominant Feature Per Group and Composition

When the customers within each group are examined, they match with the clustered customers. For example, health care customers appear in group 0, Latin America "mostly financials customers" appear in group 2, and U.S. government customers appear in group 3.

Aspects of the techniques presented herein may be applied to many different use cases. Several of those use cases will be described below, including a working system in a partner cloud and marketing reporting.

A first use case example encompasses a partner cloud with peer comparison models. Using similar network modeling to that which was described above, it may be determined that the partner cloud has a peer comparison feature as shown in Figure 10, below. Application of aspects of the techniques presented herein identifies the most similar peers and through a user-friendly user interface (UI) shows a side-by-side view of all of the details from their overview, health, stability, age, etc.

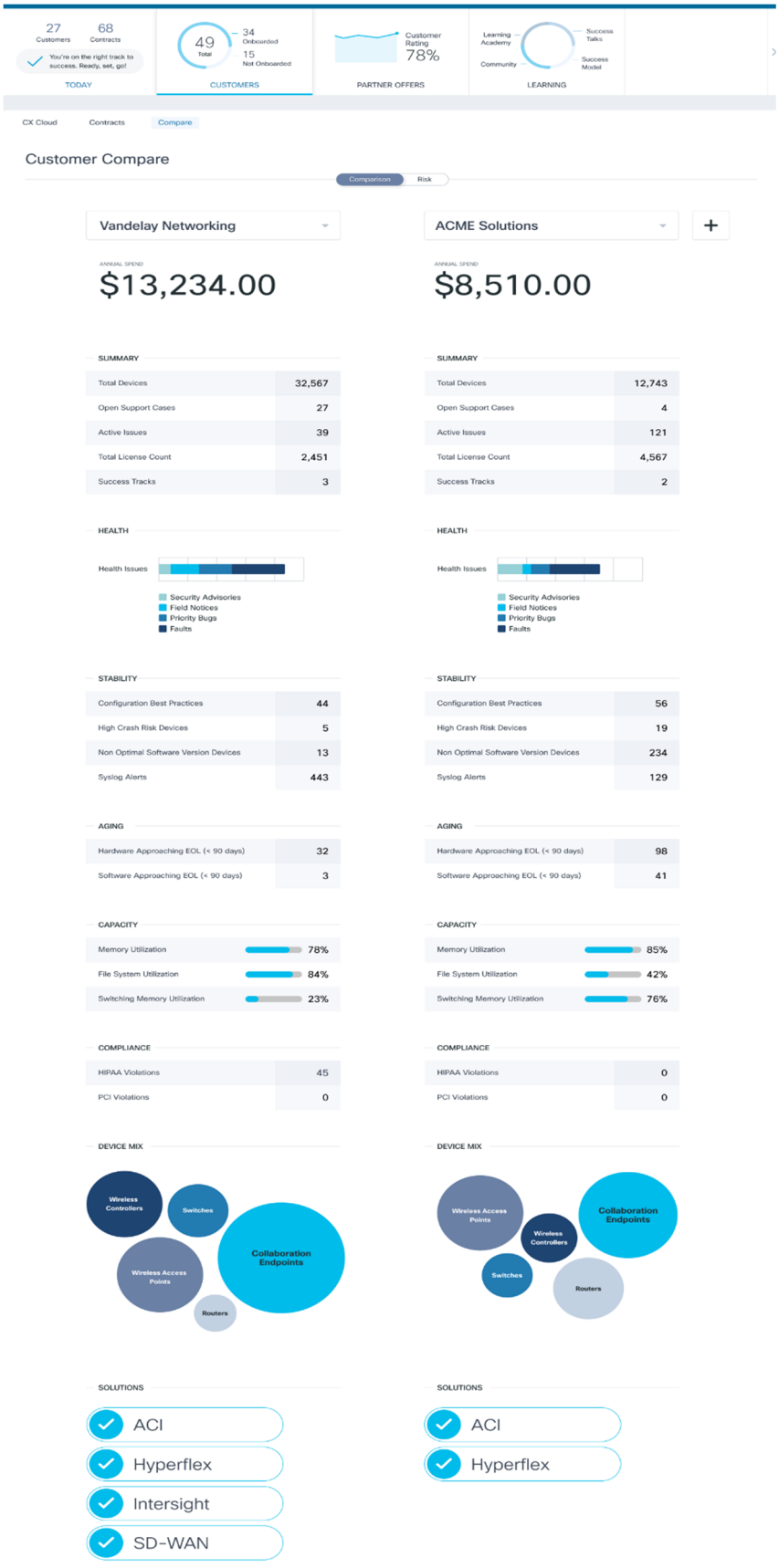
27
Customers
68
Contracts
You're on the right track to success. Ready, set, go!
TODAY
49
Total
34
Onboarded
15
Not Onboarded
CUSTOMERS
Customer Rating
78%
PARTNER OFFERS
Learning Academy
Community
Success Talks
Success Model
LEARNING
CX Cloud
Contracts
Compare
Customer Compare
Comparison
Risk
Vandelay Networking
ACME Solutions
ANNUAL SPEND
$13,234.00
ANNUAL SPEND
$8,510.00
SUMMARY
Total Devices 32,567
Open Support Cases 27
Active Issues 39
Total License Count 2,451
Success Tracks 3
SUMMARY
Total Devices 12,743
Open Support Cases 4
Active Issues 121
Total License Count 4,567
Success Tracks 2
HEALTH
Health Issues
Security Advisories
Field Notices
Priority Bugs
Faults
HEALTH
Health Issues
Security Advisories
Field Notices
Priority Bugs
Faults
STABILITY
Configuration Best Practices 44
High Crash Risk Devices 5
Non Optimal Software Version Devices 13
Syslog Alerts 443
STABILITY
Configuration Best Practices 56
High Crash Risk Devices 19
Non Optimal Software Version Devices 234
Syslog Alerts 129
AGING
Hardware Approaching EOL (< 90 days) 32
Software Approaching EOL (< 90 days) 3
AGING
Hardware Approaching EOL (< 90 days) 98
Software Approaching EOL (< 90 days) 41
CAPACITY
Memory Utilization 78%
File System Utilization 84%
Switching Memory Utilization 23%
CAPACITY
Memory Utilization 85%
File System Utilization 42%
Switching Memory Utilization 76%
COMPLIANCE
HIPAA Violations 45
PCI Violations 0
COMPLIANCE
HIPAA Violations 0
PCI Violations 0
DEVICE MIX
Wireless Controllers
Switches
Collaboration Endpoints
Wireless Access Points
Routers
DEVICE MIX
Wireless Access Points
Collaboration Endpoints
Wireless Controllers
Switches
Routers
SOLUTIONS
ACI
Hyperflex
Intersight
SD-WAN
SOLUTIONS
ACI
Hyperflex

Figure 10: Side-by-Side Peer Comparison

Additionally, a customer may have a multidimensional view as shown in Figure 11, below.

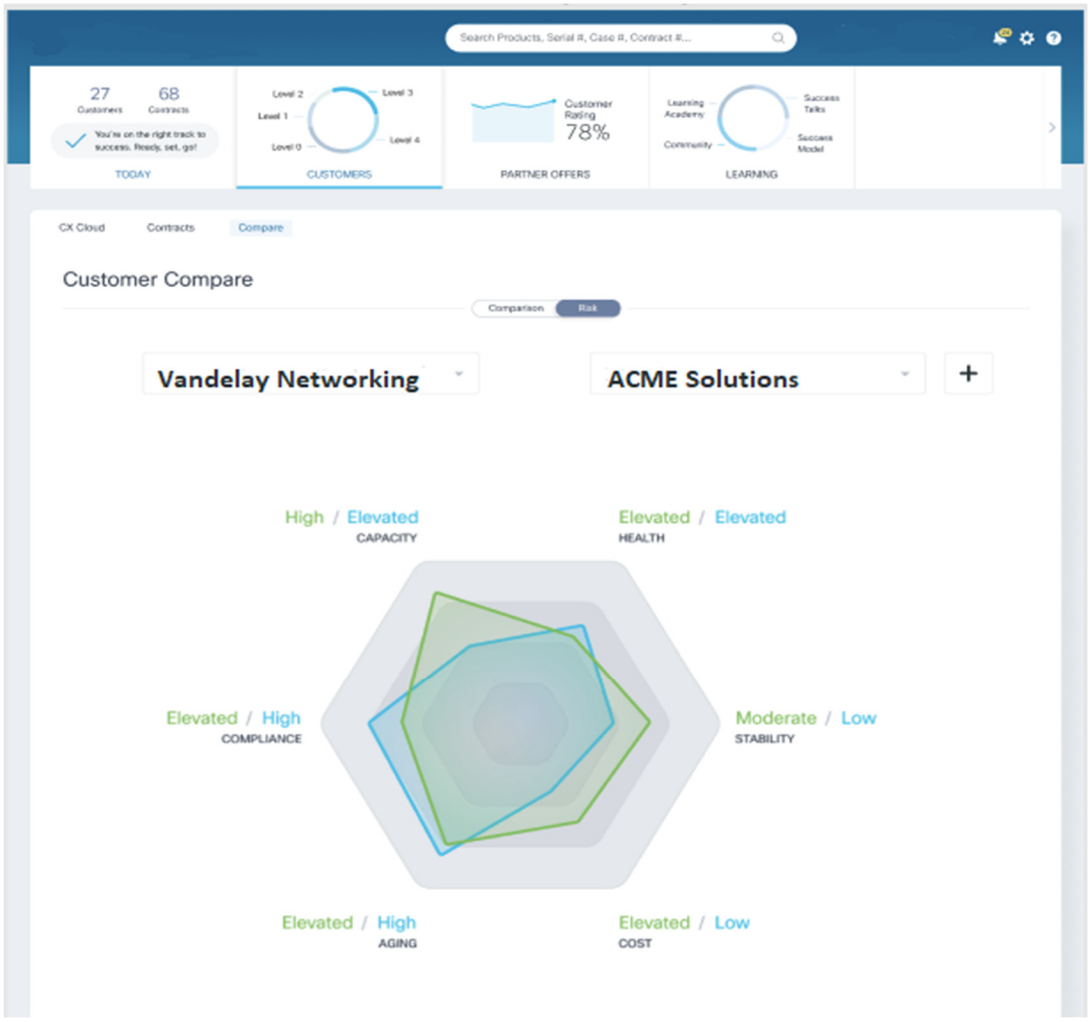

Figure 11: Multidimensional Peer Comparison

A second use case example encompasses a business comparison used in a marketing team. Figure 12, below, is a sample risk matrix evaluation report used by a marketing team as developed through the application of aspects of the techniques presented herein. The top distribution graph shows the business risk of the customer network and a red bar indicates the current situation for a given customer. The five graphs on the bottom

are "zoom in" views for different perspectives (such as, for example, security, aging, licensing, etc.). Similarly, the red bars highlight the current situation in each perspective.

Figure 12: Business Matrix with Peer Comparison Models

Such a report is very helpful for a marketing team to make upsell suggestions to a customer. For example, such a team may suggest that a customer purchase better devices to reduce their security advisory as the comparison clearly indicates that they are far behind their peers.

A third use case example encompasses a display of dominant features within a UI. By applying aspects of the techniques presented herein it is possible to show the dominate features in a customer network. For example, the customer can understand why Peer 1 and Peer 2 are similar to them (as they have similar features) as shown in Figure 13, below.

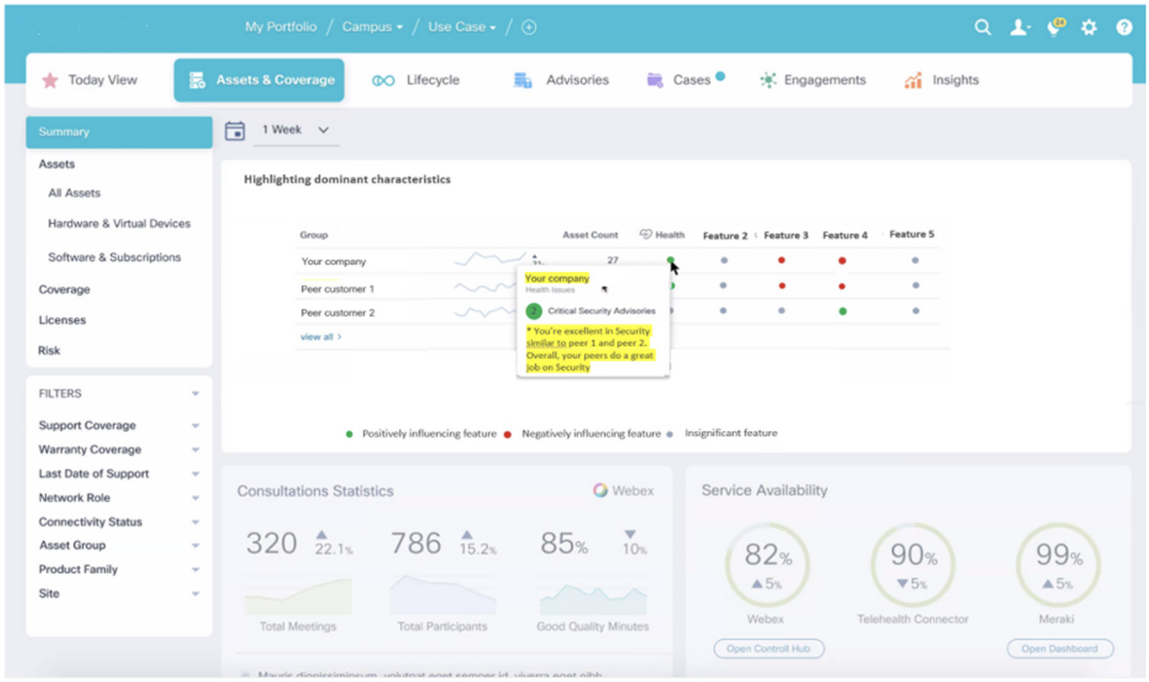

Figure 13: Exemplary Dominant Features for Similar Customers

The three use cases that were described and illustrated in the above narrative are exemplary only. It is important to note that aspects of the techniques presented herein can benefit many other use cases as well.

In summary, techniques have been presented that leverage ML models to resolve the two fundamental questions that were described above. Aspects of the presented techniques encompass an end-to-end system, which for convenience may be referred to as DeepSense, that resolves the entire lifecycle mystery of peer comparison. Additionally, aspects of the presented techniques employ an SVD algorithm to define similarity among customers in a way that is able to overcome the limitations that are caused by latent information. Further, aspects of the presented techniques leverage NMF to capture the dominant features which can influence the similarity among peers. And finally, aspects of the presented techniques support a user-friendly customer interface in real working production systems.