Winter 12-5-2005

A Customer-Support Knowledge Network Integrating Different Communication Elements for an E-Commerce Portal Using Self Organizing Maps

Pradeep Kumar

Tapati Bandopadhyay

A Customer-Support Knowledge Network Integrating Different Communication Elements for an E-Commerce Portal Using Self Organizing Maps

Tapati Bandopadhyay & Pradeep Kumar
ICFAI Business School, IDPL Complex, Old Delhi Jaipur Road, Dundahera,Gurgaon (Haryana) 122016, INDIA
Phone : 91-124-2455485
Fax : 91-124-2455484
pkgarg@ibsdel.org , tapati@ibsdel.org
Track Area
Customer Relationship Management
Correspondence Author
Pradeep Kumar (pkgarg@ibsdel.org)

Abstract: Successful e-commerce portal organizations focus intensely on customers. They try to consider every bit of information that flows from the customer to their system as an input for analyzing and identifying their needs precisely and catering to them. Being mostly 'click and mortar' or completely e-enabled, they have a lot of operational flexibility to address customer requirements in a more personalized and customized way than their brick-and-mortar counterparts with more operational rigidity and resource constraints.

Managing diverse range of channels is a challenge because of exponential and sometimes disruptive growth of diverse technologies that are used for supporting high-volume e-commerce operations. Customers are bouncing between phone, email and the web with greater fluidity than ever and therefore, fragmented, 'stove-pipe' communications, in such situations, can create problems as they loose out the holistic view on the basic nature of the problems and customer priorities. Therefore, the use of a common knowledge base across all channels is a dire necessity for an e-commerce portal, especially the ones which do not have a 'brick-and-mortar' back-end. The customer-support knowledge network as proposed in this paper addresses these issues. Using Self Organizing Maps(SOM), the network becomes incrementally self learning representing various groups of communication instances at any point of time. The advantages include the integration of all communication elements and an assimilation of all the customer communication issues into a reusable form of self-learning network. It adds an immense value for a customer-focused e-commerce company for identification of generic issues, better understanding of customer concerns and priorities and designing products/ services/ promotions accordingly, to ensure an overall better success of business.

Keywords: Self Organizing Maps(SOM), Customer Relationship Management (CRM), E-Commerce, Knowledge Network.

Proceedings of the Fifth International Conference on Electronic Business, Hong Kong, December 5-9, 2005, pp. 71 - 77.

I. Introduction

I. 1 Importance of Effective Customer Communication-Based Knowledge

Successful e-commerce organizations treat customers as the only way of their survival as well as gaining competitive advantage. They focus intensely on customers, creating customer profiles, designing promotions and offers focusing onto particular customer segments and online communities. They try to consider every bit of information that flows from the customer to their system as an input for analyzing and identifying their needs precisely and catering to them. Being mostly 'click and mortar' or completely e-enabled, they have a lot of operational flexibility to address customer requirements in amore personalized and customized way that their brick-and-mortar counterparts with more operational rigidity and resource constraints. Big e-commerce portal companies like Yahoo and Amazon.com invest heavily in systems like CRM (Customer Relationship Management), processes and often in capturing inputs also to improve their products and provide better services. They focus on retaining customer loyalty as well as on attracting new customers. In Roffey Park research (2004), high performance has been shown to be strongly correlated to customer-focused purposes. Internally also, a significant high proportion of employees in HPOs think that their organization is highly customer-centric or even customer-driven. So, ignoring customer needs can be completely detrimental to sustainable growth of any organization. The most effective vision for change is customer focus. [3].

To achieve customer focus, one of the crucial and valuable sources for any company is the customer communication. It is all the more important for e-commerce portals which often do not have any backend brick-and-mortar interface for handling customer communication, and thereby consequently have to maange a diverse range of communication channels for customer interfacing. Most of the e-commerce or e-business portal companies have a lot of ways to gather customer feedback, e.g. through online chatting with a support-resource, or though e-mails or form-based interfaces on the net, because it gives the customer the flexibility to communicate in an asynchronous

domain (which is not the case with the telephone calls) and also gives them a platform to communicate in writing which is a more convenient way as perceived by people for putting the problems or thoughts in a more structured fashion. Also it reduces the anomalies in understanding as everything is in writing, and it also helps the customer or the company to keep a track of the communication that has taken place for example in terms of a series of request-response mails. Even in case of e-commerce-based companies which completely rely on digital medias for customer interfacing, customers communicate in a variety of ways. These are usually mentioned and supported as the communication channels between the customer and the company in the company website, e.g. the contact email_ids, the phone numbers or online forms. So, consequently, the customers call. They send email. They visit web sites and use the self-service portal facilities or chat services or online complaint register forms etc. Managing each of these communications channels as effectively as possible is a real challenge. It has now become even more critical to manage these channels collectively, because, with the exponential and sometimes disruptive growth of diverse technologies that are used for supporting high-volume e-commerce operations, customers are bouncing between phone, email and the web with greater fluidity than ever. They send emails about problems that they have already discussed by phone, and then visit the web site to double-check the information they received on the phone. Fragmented communications, in such situations, can create problems. The portal's service representatives get blindsided by calls from frustrated customers. They repeatedly answer redundant questions, as they do not have any unified or holistic view of all the communication instances of the same customer on a same issue. Neither can they easily add new knowledge items to the web site or the portal's back-end knowledge resources. Worst of all, e-commerce companies with "stovepipe" communications channels e.g. isolated calls, email and web self-service systems can not adequately understand or address customers' top service issues, because they loose out the holistic view on the basic nature of the problems and customer priorities. Use of a common knowledge base across all channels is therefore a dire necessity for an e-commerce portal, especially the ones which do not have a 'brick-and-mortar' back-end equipped with actual people or real human faces to interface with the customers. The business impact of this unified approach is substantial. A unified knowledge network for managing customer responses can deliver competitively superior service at dramatically reduced cost, because each channel becomes more efficient and, over time, more and more customer interactions are driven to the least expensive and most scalable channel that is the web. This also provides companies with clearer insights into their customers' top concerns. These insights are enormously helpful for improving customer care, driving the development of successful products, promotion schemes, special offers focusing on particular customer segments and services, and formulating high-impact marketing strategies.

I. 2 The Customer-Support Knowledge Resources

When an e-commerce portal handles customer interfacing in a stove-piped situation:

- The phone channel is managed by a call center platform.
- The email channel is managed with an email management system.
- The web channel is managed with a content management system.
- Chat, if provided at all, is managed by a chat system.

All these systems, when disconnected, on heterogeneous platforms without any integration, communication or sharing, create information/ communication islands of their own without any connecting road-maps. So, naturally, the portal company tries to manage each of these communications channels as effectively as possible. But it has now become critical to manage these channels collectively. That is primarily because customers are bouncing between phone, email and the web with greater fluidity. Any interaction with a customer on one channel is completely divorced from any other interaction with that same customer on any other channel, because the interactions do not flow down to any common knowledge repository. When customer service representatives (CSRs) are on the phone with a customer, they have no idea what that customer's emails looked like. The information that a customer gets in an email may not exactly match the information on the company's web site. Each channel is isolated from the others—to the detriment of both the company and its customers. Subsequently, the problems that companies and their customers experience every day because of stovepipe communications channels are exemplified in situations as described below:

- The "Didn't you see my email?" syndrome

A customer with a problem takes the time to carefully describe that problem in an email message and then calls for a follow-up. The support person taking the call has no clue of the original email exchange. So instead of just asking a follow-up question, the customer has to explain the whole situation all over again. The customer gets annoyed, the CSR spends an unproductive time.

- The endless repetitive calls syndrome

Companies with conventional "stovepipe" customer communication channels have great difficulty developing effective self-service content for their web sites. When the support representatives encounter a new support issue, they have no fast and easy way to *capture* that issue and *publish* it as a knowledge item on the web. So instead of answering questions only once on the phone and then having them answered automatically in the future with relevant online knowledge items, "stovepipe" communication channels answer the same questions over and over.

The overall negative business impact of these problems can be that inconsistent information can be provided by each channel and customers are more frequently annoyed,

angered and confused. "Stovepipe" channels also deprive companies of insight into the issues that most concern their customers. It is difficult to get that insight if the company can not get a holistic view of customer interactions i.e. they can not look at communications activity across all of the customer service communications channels.

I. 3 Customer Communication Elements

Customer communications with an e-commerce portal can be broadly classified into following categories based on the purposes of communication:

- Enquiries e.g. on new products/ services/ schemes
- Feedback.
- Complaint
- Support

There are also the various ways customers communicate to the organizations using technology/ media combinations at different levels of complexities. From Figure 1, it can be derived that the customer input/ feedback sources are not only varied in terms of their use of various channels or media, or their degree of structured-ness and unstructured-ness, but also in terms of the underlying technologies like the communication domains, which also encompasses the network elements, the server technologies, data repositories like warehouses or operational RDBMS.

Using all possible combinations of these varieties, capturing customer knowledge from these combinatorial resources in a unified form is a daunting task. The ideal situation in this case would be as shown in Figure 2:

This is precisely the solution that we are proposing in this paper. Towards this end, we have first developed a generic process model, explained in the next sections, based on SOM for dynamically gathering communication inputs from all channels and then converting them into a knowledge network based on the unsupervised learning mode of the neural networks.

I. 4 Existing Tools and Techniques to Capture Customer Communication

Generally, in the context of capturing knowledge or discovering reusable patterns from customer inputs, data mining, pattern recognition techniques have been abundantly used. Data warehousing and on-line analytical processing (OLAP) have typically been used to solve data extraction, transformation, data cleaning, storage, and mining issues. Regarding handling of natural-language based unstructured information sources; previous efforts have used document-based technologies and supported document-level functions such as full text search, document classification, and so on. Business practitioners have developed automated tools to support better understanding and processing of information. Researchers have also developed advanced analysis and visualization techniques to summarize and present vast amount of information. [4] Despite recent Improvements in analysis capability [4], there is still a long way to go to assist qualitative analysis effectively. Most tools that claim to do analysis simply provide different views of collection of information {e.g.. comparison between different products or companies).Lin [8] identified various display formats for handling multi-dimensional data e.g. hierarchical displays-an effective information access tool for browsing, network displays, scatter displays [10]. There are other methods for document representation and visualization also which primarily concerns the task of getting insight into information obtained from one or more documents [12]. Most processes of document visualization involve three stages i.e. document analysis, algorithms, and visualization [10]. Web content mining treats a web document as a vector of weights of key terms [1]. He et al. [5] proposed an unsupervised clustering method that was shown to identify relevant topics effectively. The clustering method employed a graph-partitioning method based on a normalized cut criterion. This method we are using in this paper to extract knowledge from customer communication resources.

II. Creation of a Knowledge Web Using Customer Communication Elements

As explained in Figure 1 regarding various elements in customer communication, there is a need for integrating the knowledge captured from these communication resources. Customers sending mails or leaving voice messages are generating valuable knowledge assets. Voice recordings can be converted into text documents. Now, treating these text messages (source being emails/ form based text inputs/ voice messages converted) as unstructured documents, we can use co-occurrence analysis to find the similarities and then consequently the dissimilarities between the messages/ text contents. For this paper, herein after, we are using messages as the unit of customer communication and thereby are referring to any of the communication elements as messages. Messages which are very similar in terms of their contents i.e. many of the identified key-terms (i.e. Terms excluding the general terms like pro-nouns, prepositions, conjunctions etc.)are same, can be clubbed up together to form a cluster. Dissimilar message/ text bodies can be created as other clusters. These clusters can then form a network using hierarchical and partitional clustering method to form a graph which we are mentioning in this paper as a Knowledge Web based on customer communications.

II. 1 Self Organizing Maps

In this paper, we are proposing a generic model for developing a knowledge network to integrate various customer interaction/ communication elements. This network is developed using SOM(Self Organizing Maps) proposed by Kohenen[6].

If we treat team the customer communication resources broadly as textual(which most of them are, as is apparent from their description above in Figure 2, there are various classification schemes applicable to them. Most popular text-data classification techniques include clustering. There

are two main clustering methods: statistical and neural network approach. In the serial statistical approach, automatic text-data classification involves two processes: 1) a text representation structure, 2) a method to determine similarity between such text data. One the class identification is done, the hierarchical clustering of text data can be done divisively or agglomerative. In divisive clustering, one cluster is further broken down into smaller segments, whereas in agglomerative clustering, a gluing operation based on the similarity between clusters takes place to form groups. There is also a method of conceptual clustering. [11]. Algorithms for clustering involve co-occurrence analysis for similarity measurements, discovering conjunctive features and clumping relationships based on most frequently occurring data, and then a pair-wise comparison.

The neural network approach is a connectionist approach. Algorithms in neural networks are mainly parallel where multiple connections among the nodes allow independent, parallel, and consequently faster comparisons. These techniques are also of two types: supervised and unsupervised. In the supervised mode, a training data set is first presented in the training phase, based on which the network calculates the input weights , compares it with the desired output weights for as particular input, then derives the weights and minimizes errors incrementally. In unsupervised method, an input vector is given to the network. Through various network learning rules, these inputs vectors are analyzed for their statistical or similarity properties, and the network consequently dynamically adjusts the weights of each node forming in the graphical representation of the cluster. SOM is an unsupervised learning method for the neural networks. This is based on the associative properties concept of human brain functioning. The network contains layers like the input layer and the mapping layer in the form of a 2-D grid. The output layer can be thought as a distribution layer. SOM has been applied successfully in many research efforts concerning document classification. Kohenen[6] applied this model for finding out logical similarities between words. Lin et al [7]used it for information retrieval. Orwig[9] and Chen [2]] used it as SSOM or Scalable SOM to classify electronic brainstorming outputs.

We intend to use SOM developing a knowledge network with inputs from various customer communication resources as shown in Figure 2. These resources are to be extracted from heterogeneous sources i.e. mail servers, chat servers, IVR(Interactive Voice Response) systems, web-based forms, self-service portals etc. and then can be given as inputs to the knowledge network creation module. SOM appears to be a best candidate for choice in this context primarily because:

- Neural network methods being able to work in parallel can be much faster, provided adequate hardware resources are there, to handle complex, and voluminous text data resources, which is the typical case in the context of this paper, i.e. IS development projects.
- SOM in particular is an unsupervised learning method, so this can effectively self-enhance itself and keep the temporal consistency and integrity of communication elements i.e. can self-enhance based on the latest customer inputs

III. The Generic Model for Integrated Customer Communication-Based Knowledge Network Creation

This model takes all heterogeneous inputs from various sources as the inputs to the SOM module, as shown in Figure 3 below.

We explain this model with an example as below. Suppose a customer has bought a high-value electronic gadget from an e-commerce portal. The gadget has been installed but is not functioning properly. So, the customer has to communicate about this with the portal company. Suppose he/she makes the communication as follows:

- He/ she first tries the self-service facilities that provides information on trouble-shooting of that particular gadget. But, the customer does not succeed in solving the problem following the instructions or information that were given in the self-service portal.
- So, he/she calls up the portal's help-desk. An IVR complaint recording is done. The complaint_ID gets auto-generated by the IVR back-end and is given to the customer though the IVR system.
- The customer is worried because it is a brand-new high-value item. So he again calls for a follow-up. This time he chooses to speak to a service rep. He gives the IVR-given complaint-ID to the service rep.
- The service rep checks up on the status of the complaint handling process and gets back to the customer saying that it would be better if the customer could send a detailed e-mail on what is precisely the problem, as the voice recorded message is short and not fully explanatory.
- The customer sends a long e-mail explaining the whole problem in detail including his/her experience in trying to fix it by himself/herself using the self-service portal, and then giving reference of the same complaint-ID.
- The problem is sorted out next day when a mechanic visits, equipped with the necessary parts and tools.

Now, this sequence of communication has the following elements:

- A self-service session: can be identified by a transaction_ID on the log-file records of the portal.
- An IVR interaction and a voice-recorded message: identifiable with a complaint_ID
- A telephonic interaction with a service rep: identified and referred with the same complaint_ID
- An e-mail with the complaint_ID and referring to the transaction with the self-service protal.

According to the generic model, the unsupervised SOM-

based knowledge network module would take thse inputs in an integrated fashion as follows:

- The network would receive messages from
  - the IVR back-end database,
  - the call recording
    - converted into digital contents using voice analyzer software,
  - e-mail messages from the e-mail server
  - transaction log-files from the self-service portal server
- The network will create the first level of clusters:
  - In the example case, first by locating the corresponding transaction_ID and associating it with the complaint_ID
  - Then by grouping the entire sequence of communication as:

CommunicationInstance[i] {Complaint_ID, [elements of communication]}

Otherwise, this primary clustering canmbe also based on:

- Customer_IDs e.g. for new product/ service queries/ propomotions
- Query_IDs etc., depending on the purpose of communication which is explained in section 2

- This level of cluster would give an integrated view of all communication elements for a particular sequence or communication instance.
- Now, these communication instances will be fed into the SOM module for creation of a reusable customer communication knowledge network, the steps being depicted below:
  - Collate all elements, say $\sum_{i=1}^{N}[E_i \ldots E_j]$ from N communication instances, over a time t
  - group them by automatic parsing for a reference information i.e.
    - Class_reference numbers for complaint classes in the Complaint_IDs
    - Error/ exception no.s
    - Transaction Ids
- Create an input vector of N keywords for each group, e.g.:

GroupInstance1[reference index(Complaint_Class/ TxnID), string_of_terms_used (nouns(a..n), verbs(a…n), qualifiers(a..n),)]

- Input an initial group vector e.g. GroupInstance1 to the SOM module
- Initialize the input nodes, output nodes and connection weights: Represent each string as input vectors as shown above. Create a 2-D grid based on the N input vector elements and M output nodes.
- Input all the GroupInstance vectors for all the groups and the grid gets created by dynamically adjusting the weights to the input nodes based on the input vector string terms
- Compute distance between all nodes e.g. D_j between the input and each output node j as follows:

$$D_i = \sum_{i=0}^{N-1}(x_i(t) - w_{ij}(t))^2$$

where x_i (t) is the input to node I at time t and $w_{ij}(t)$ is the weight from input I to output node j at time t.

- select wining node j^* (i.r. the node with the minimum distance) and update weights to node j^* and it's neighbors based on reducing distances.
- Label the regions in the map identifying the typical group features.
- Go to step 1

The SOM network that is formed this way will become a self learning incremental knowledge network representing various groups of communication instances at any point of time t.

IV. Conclusion

There are two main advantages of this customer-support knowledge network for an e-commerce portal:

1. With the primary clustering, all the communication elements regarding one sequence or one problem are integrated. So, the problem of stove-piped communication is resolved.
2. With the knowledge network, all the customer communication issues are assimilated into the self-learning network in a reusable from, which again would add immense value for a customer-focused e-commerce company, for identification of generic issues, better understanding of customer concerns and priorities and designing products/ services/ promotions accordingly, to ensure an overall better success of business.

There are various possibilities of extending the premises of this paper like creating a knowledge network of support teams themselves and/or the support aspects of an organization.

References

[1] Bowman. C.M, Danzig. P.B., Manber. U.; Schwartz, F'. Scalable Internet resource discovery: Research problems and approaches. *Communication of the ACM. Vol* 8 (1994). pp 98-107.
[2] Chen, H; Schuffles, C; Orwig, R, Interenet categorization and search: a self-ornizing pproach, jornal of visual communications and image representation, 7, 1 (March 1996), 88-102
[3] Ellsworth,R.R, *Leading with Purpose: The New Corporate Realities*,Stanford University Press,Stanford,2002.
[4] Fuld, L.M.: Singh. A.: Rothwell. K.; and Kim, J. *Intelligence Software Report™ 2003: Leveraging the Web*. Cambridge. MA: Fuld & Company, 2003.
[5] He. X.; Ding. C; Zha. H.; and Simon, H. Automatic topic identification using Webpage clustering. In X. Wu. N. Cercone, TY. Lin, J- Gehrke. C. Clifton. R. Kotagiri. N. Zhong. and X. Hu (eds,). *Proceedings of the 2001 IEEE International Conference on Data Mining.* Los Alamitos. CA: IEEE Computer Society Press. 2(X)I. pp, 195-202.
[6] Kohenen, T., Self organizing maps, Berlin: Srpinger-Verlag, 1995
[7] Lin X.; Soergel D., and Marchionini G., A self organizing semantic

map for information retrieval, prioceedings of the Fiftennth Annual International ACM/SGIR Confefernce in R&D in Informaiuton retrieval,Copenhagen, 1992, pp 32-50

[8] Lin, X. Map displays for information retrieval. *Journal of the American Society for Information Science. 4H.* 1 (1997), 40-54.

[9] Orwig R., Chen, H., Nunaaker J. F., A graphical selkf organizing approach to classifying electronic meeting output,Journla of American sociery of Informaiton Scuiences, 48, 2 (1997), 157-170

[10] Spence, R. *Information Visualization.* New York: ACM Press, 2001.

[11] Stepp R., Coinecpts in conceptual clustering, Porceedings if 10th Internaitonal Conference on AI, Milan 1987.

[12] Wise, J.A.; Thoma, J.J.: Pennock. K.; Lantrip, *D.:* Pottier, M.; Schur, A.; and Crow, V. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of IEEE Symposium on Infonnation Visualization.* Los Alamitos, CA: IEEE Computer Society Press, 1995, pp. 51-58.

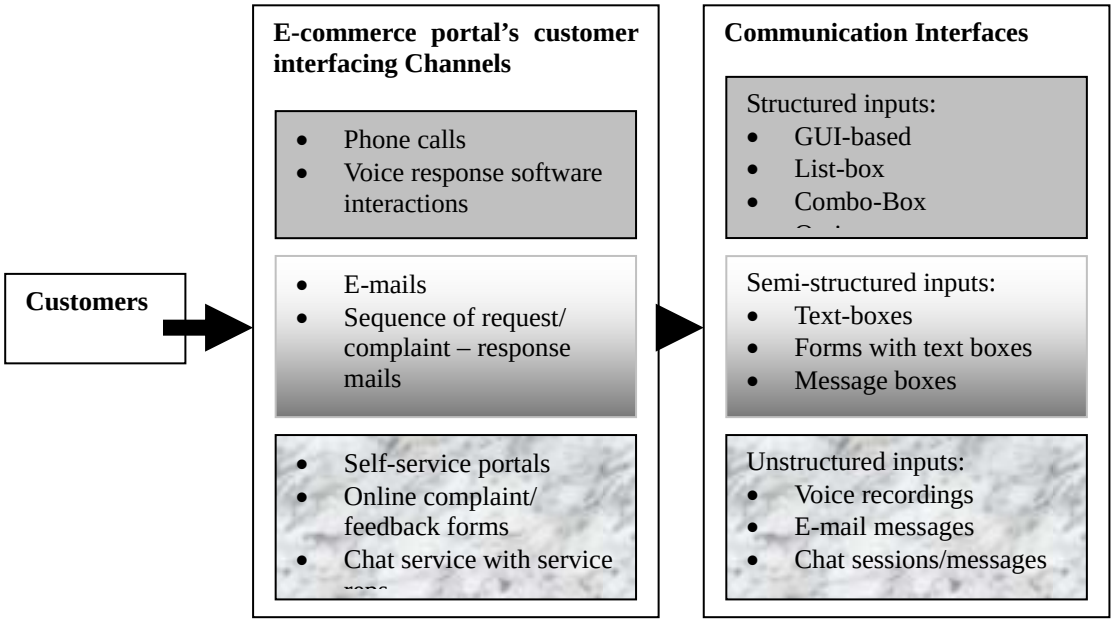

Figure 1: Elements of Customer Communication

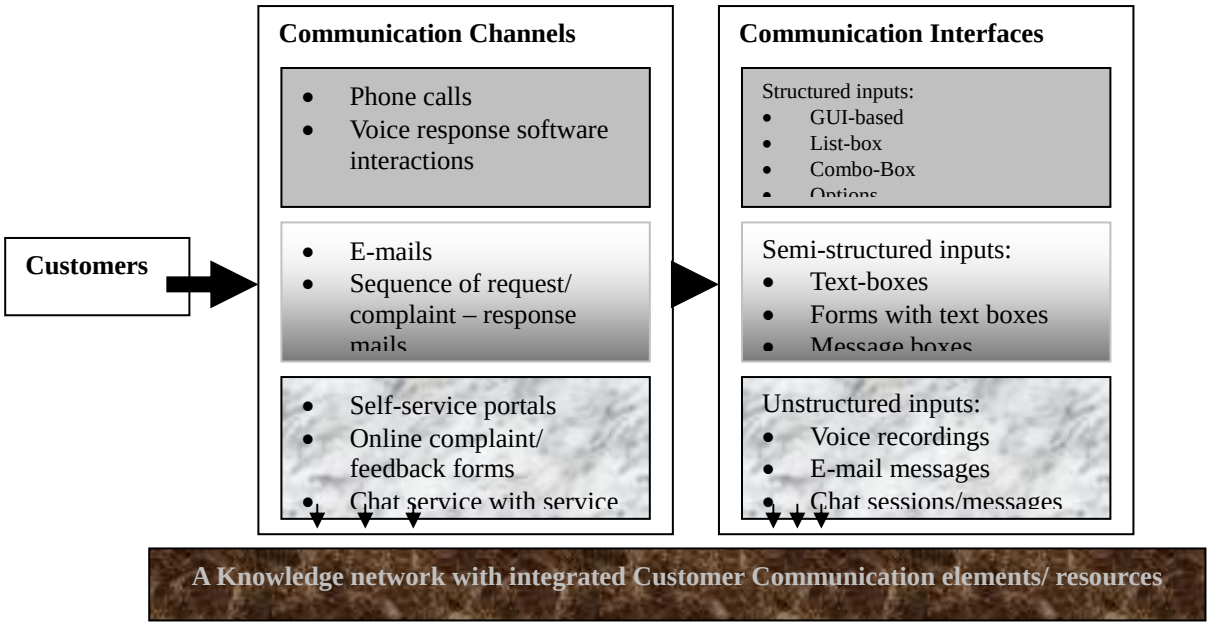

Figure 2: The integrated Communication Situation

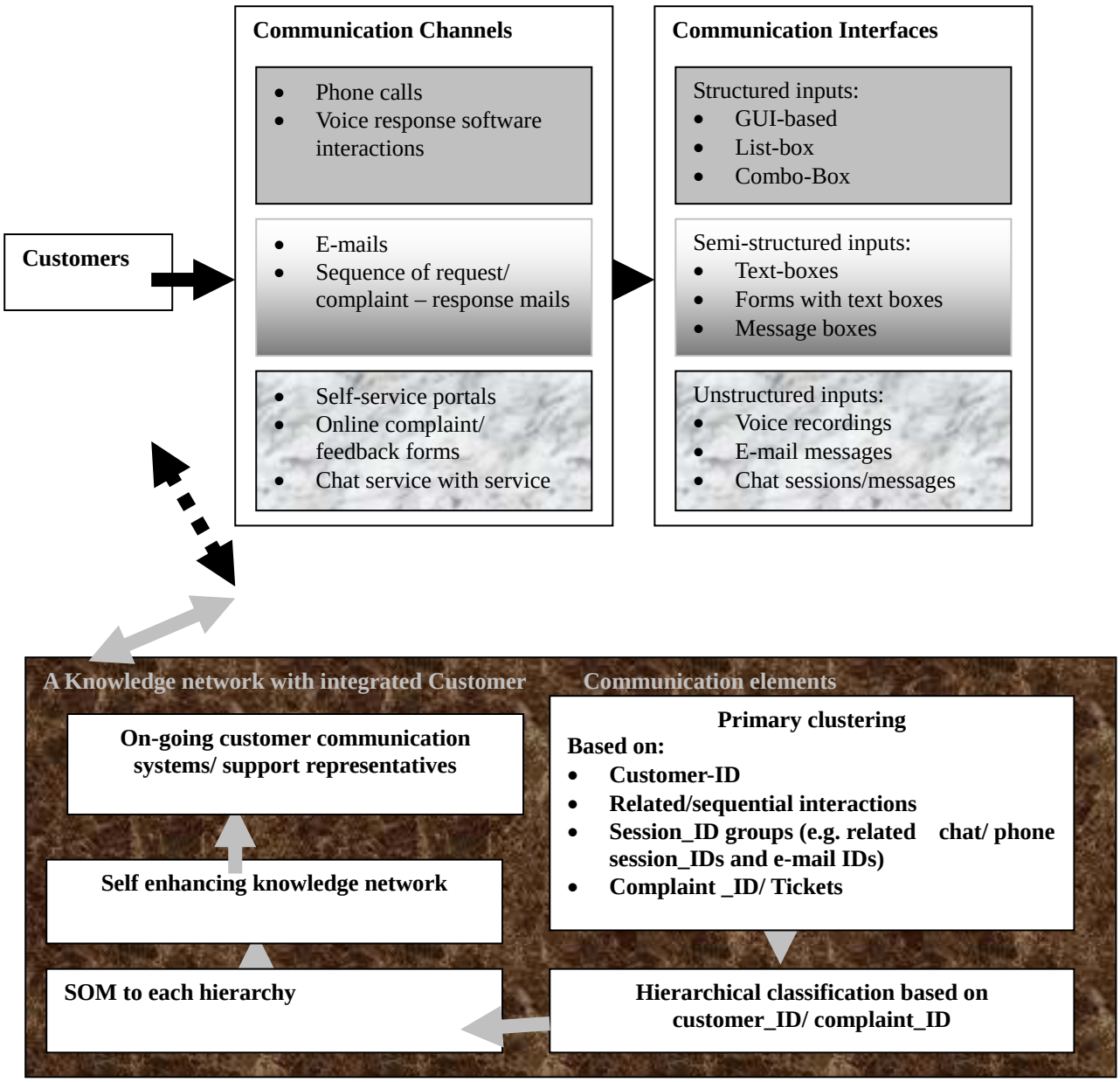

Figure 3: The generic model for customer communication knowledge network creation