Do Neural-Network Question Answering Systems Have a Role to Play in the Deployment of Information Systems?

Antonio Juarez Alencar
juarezalencar@br.inter.net

Renata Chaomei Wo
renata_wo@yahoo.com.br

Eber Assis Schmitz
eber@nce.ufrj.br

Priscila M. V. Lima
priscila@nce.ufrj.br

Armando Leite Ferreira
armando@coppead.ufrj.br

Institute of Mathematics and The Electronic Computing Center
Federal University of Rio de Janeiro
P.O. Box 68530
21941-590 - Rio de Janeiro - RJ, Brazil

Abstract

As Internet users become more numerous, experienced and skillful, and the number of companies doing e-commerce increases worldwide, so does the demand for on-line information about products and services. To satisfy this increasing demand for on-line information many companies have resorted to providing customer support services over a variety of on-line means of communication such as e.mail, chat services, voice on internet protocol (VOIP), etc.

This article presents a stepwise approach to the construction of hybrid question answering systems based upon neural network technologies and natural language processing. These special kind of information systems not only provide high speed answers to questions posed by customers, but they also allow customers to receive answers to their questions on a 24/7 basis, provide well conceived standard answers to those questions, allow for a precise recording of customer communication, and make the management of customer support services easier. All of this is made clear by a case study about the development of an automatic question answering systems to the "SEBRAE Challenge", a business game involving university students in seven different countries in South America.

1 Introduction

As Internet users become more numerous, experienced and skillful, and the number of companies embracing e-commerce increases worldwide, so does the demand for on-line information about products and services. The nature of the information required by Internet customers is varied, going from basic technical support to the description of product features, location of outlets and authorized repair services, product warranty coverage, requests for contact with commercial representatives, current status of suggestions and complaints they have made, etc. [12]

To maintain the balance between the increasing demand for information and their capacity to supply it, companies have resorted to a number of different on-line and off-line strategies, such as: expanding and outsourcing existing customer call centers, installing voice response units (VRUs), making lists of frequently asked questions available to customers, setting up web sites with information about products and services, and providing customer support services over a variety of computerized means of communication such as e-mail, chat services, voice on internet protocol (VOIP) etc. [1].

However, recent advances in applied computational intelligence and natural language processing have allowed for the development of question answering systems (QAS's), i.e., a special type of information systems that automatically detect the existence of requests for information, correctly identify their contents and, using a knowledge base, provide adequate answers to these requests [3].

When a request for information posed by a customer is beyond the answering capability of these systems three different courses of action may be taken: (i) provide an answer that is likely to fall short of satisfying the customer; (ii) let the customer know that the system is unable to provide an adequate answer to the request posed by them; or (iii) redirect the question to a human operator or a non-automatic system such as product manual or list of frequently asked questions.

When a question answering system seeks human assistance to help answering questions that fall beyond its knowledge base, it becomes a hybrid human-machine question answering system, or hybrid QAS for short.

This article presents a stepwise approach to the construction of web-based hybrid question answering systems based upon neural network technology and natural language processing. Such systems may be used to satisfy customers' needs for on-line information when conditions allow, with many advantages over other options.

2 Conceptual Framework

Since Allan Turing (1912-1954) came up with the concept of artificial intelligence while working as a leading cryptanalyst in Bletchley Park, England, during World War II, researchers have passionately pursued the goal of developing computer systems one could talk to in natural languages and get proper answers to questions [11].

However, among the multitude of computer related subjects in which research and commercial development is done, QAS has revealed itself to be a particularly challenging topic, where progress has not been made easily. This is partly due to the fact that QAS builds upon the advancements of several scientific fields, including natural language processing, information retrieval and human computer interaction. Nonetheless, Maybury [11] has identified several types of QAS's that are currently being deployed or upon which research is being done.

Differently from the large majority of QAS's that one may find in today's commercial and academic world, the QAS proposed in this article makes extensive use of artificial neural network in its inferential reasoning mechanism.

In the artificial neural network paradigm a mathematical structure composed of values and functions provides a general model for the neuron, a cell that serves as the basic construction block of the human brain and also the brain of many other living beings. However, while the human neuron receives electrical impulses through its dendrites, deals with them and propagates these impulses to connected neurons, depending on the occurrence of specific neurological conditions, the artificial neural-network neuron receives numerical values through its input nodes, processes these values using a combination of a summation and an activation function, and yields an output, depending on the input value themselves and the weights associated with them.

Diagram 1 presents the mathematical structure underlying an artificial neural-network neuron. In the diagram $x_1, \cdots, x_n$ are numerical values that represent input signals, $w_1, \cdots, w_n$ are the weights associated with each input, and f is the activation function, that receives the weighted sum of the inputs signals, i.e., $\sum_{i=1}^{n} x_i\, w_i$, as its parameter.

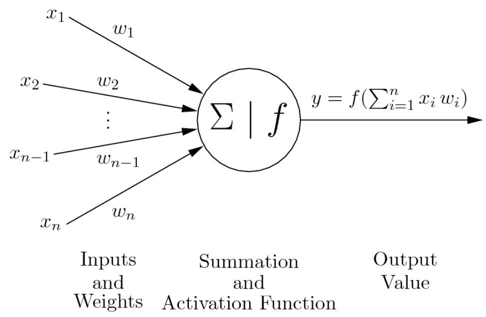

Figure 1. The mathematical structure underlying an artificial neural-network neuron.

Diagram 2, presents a feed-forward neural network that receives four input signals, i.e., $x_1, \cdots, x_4$. Each signal has a corresponding input node, whose function is simply to forward the input signals to the nodes in the next layer. In the hidden node layer there are three neurons, i.e., H_1, H_2 and H_3, which process the input signals and send a corresponding outputs to the next layer. In the output layer there is just one node, i.e., O_1, which processes the signals sent by the nodes in the hidden layer and generates an output, i.e., y.

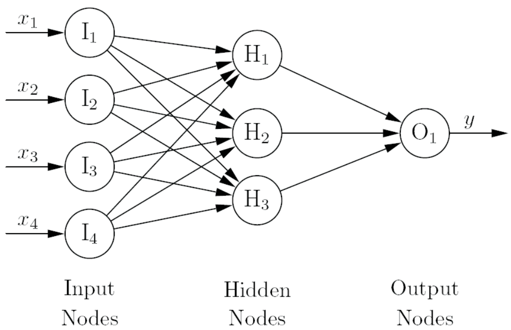

Figure 2. A symbolic representation of an artificial neural network.

Obviously neither the human brain nor the artificial neural network would be very useful tools, if they could not learn about the world around us and use the acquired knowledge to deal with new situations. In this sense, the most frequently used mechanism to train an artificial neural network is to choose randomly the weights used by the neurons and present the net to a set of observations about a specific event, together with its corresponding outcome.

The difference between the outcome of each observation and the result yielded by the network can then be successively used to adjust the weights in the hidden layers of neurons until an acceptable result is reached. This training strategy, called "back-propagation", was first proposed by Rumelhart *et al.* [15]. In the back-propagation training strategy the change in the weight w connecting unit i to unit j is given by $\Delta w_{j \leftarrow i} = \eta \, e_j \, y_i$ where a unit is either an input, hidden or output node, η is an arbitrarily chosen learning rate, e_j is the error derivative for unit j, and y_i is the output from unit i. The error derivative for unit j is given by $e_i = f'(y_i) \sum_{j=1}^{n} e_j \, w_{j \leftarrow i}$ where f' is the derivative of the activation function. A detailed discussion of back-propagation and other network architectures and training strategies is found in [7]

By choosing the right inputs, number of hidden layers, number of neurons in each layer, number of outputs, kind of activation function and training strategy one may design an artificial neural network that can precisely approximate any kind of function, hence the great usefulness of artificial neural networks [9]. Over the years a vast literature has been gathered around artificial neural networks, containing both theoretical advancements and reports on the development of practical applications. See Dreyfus [5] for a discussion about the many properties enjoyed by neural networks. Examples of recent application of neural network in many different areas is provided by Rabunal and Dorrados [13].

3 The Method

3.1 The SEBRAE Challenge

Among the several activities supported by SEBRAE one finds the "SEBRAE Challenge", a virtual business game that simulates decisions that executives face on a every day basis, including last words on the purchase of raw material, price formation, production, competition analysis, research & development, etc.

The game is played in rounds over the Internet by over 50,000 college students of seven different countries in South America, i.e. Brazil, Argentina, Paraguay, Uruguay, Chile, Colombia and Peru. In each round the participants have the opportunity to consider the current situation of their virtual companies, as well as the different forces that act upon the market where they sell products and services.

At the end of a round, which may last for several days, the participants must decide where to invest the resources they have at their disposal. These decisions have consequences upon the performance of their companies and the market where all the virtual companies do business. For example, as a result of those decisions some companies will consistently profit and prosper, while others will accumulate losses and eventually go bankrupt. Therefore, at the end of each round a certain number of low-performance companies will have eliminated themselves from the game. The process goes on until the winner is known.

Despite the extensive instructional material that SEBRAE makes available, participants of the game are free to contact the team of professionals responsible for its execution at any time via e.mail to resolve questions, make complaints and present suggestions. During the period in which the game is played, thousands of e.mails are sent to the management team. The number of e.mails is particularly high during the period that immediately precedes the deadline of each round.

To deal with the influx of e.mails SEBRAE has brought together a telemarketing service that answers these e.mails as quickly as possible, during office hours. This obviously falls short of satisfying the need for prompt answers to questions that might decide the fate of participants in the game, specially if these questions are addressed to SEBRAE outside office hours.

As result, with the view to provide answers to questions posed by the participants of the SEBRAE Challenge whenever they are needed, a hybrid QAS, based upon neural network, was built. Figure 3 presents the general structure of this system.

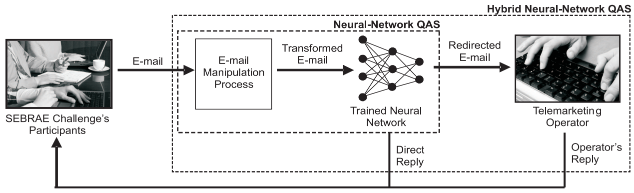

Figure 3. General structure of the SEBRAE Challenge's hybrid neural network QAS.

It should be noticed that there are circumstances in which the neural network question-answering system does not answer questions posed by the participants of the SEBRAE Challenge. Questions that are unlikely to be answered correctly are redirected to a telemarketing operator, who becomes responsible for providing the right answers.

3.2 The Data

In order to achieve the results presented in this article a considerable amount of data has been provided by SEBRAE. The data consist of a random sample of 1,531 messages sent by the Brazilian participants of the 2004 edition of the SEBRAE Challenge, together with the corresponding reply provided by the telemarketing operators. All messages were written in Portuguese.

The vast majority of those messages contained only a single question posed by participants of the SEBRAE Challenge, while a small number contained double questions. Table 1 presents the exact figures.

Message Type	Quantity	Percentage
Single question messages	1,454	95,0%
Double question messages	77	5,0%
Total number of messages and replies	1,531	100,0%

Table 1. Sample description.

The 77 double questions messages were later manually transformed into 142 single question messages that became part of the data set used to train the neural network. Therefore, the final training data set consisted of $1,454 + 142 = 1,596$ messages.

When examined closely the 1,531 replies provided by the telemarketing operators could be grouped together into 141 distinct answers without any loss of information. Further exam of the data sample revealed that the 141 distinct answers could be properly replaced by 24 new more general answers and that 5 of these new answers could be used to reply to 91.3% of all received messages. Table 2 and 3 summarizes these ideas.

Replies	Quantity	Messages Replied
Originally provided by SEBRAE	1,531	100%
Grouped replies	141	100%
New more general answers	24	100%

Table 2. Coverage of different groups of replies.

Quantity	Messages Replied
5	91.3%
19	8.7%
Total	100.0%

Table 3. Coverage of the 24 new more general answers.

3.3 Data Transformation

Before an e.mail can be presented to a neural network for both training and classification it has to go through a process of transformation. The reasons for this are quite simple: although e.mails are frequently composed of sequences of alphanumeric characters, neural networks are only able to deal with numbers. Moreover, data transformation helps

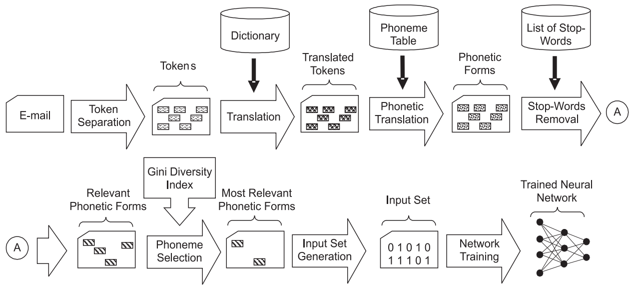

Figure 4. The neural-network question-and-answer building process.

to reduce the complexity of the problem of identifying the contents of e.mails.

The data manipulation process the e.mails provided by SEBRAE were subjected to is composed of six different activities that are executed sequentially, i.e. (i) token separation; (ii) translation of abbreviations and foreign language words; (iii) translation of words into their phonetic forms; (iv) stop-words removal; (v) selection of the most relevant words; and (vi) codification of words into binary numbers. Figure 4 places these activities into perspective.

3.3.1 Token Separation

Consists in the separation of the different tokens that compose an e.mal, i.e., different sequences of alphanumeric characters without spaces from which the punctuation signals have been removed. This is necessary because the back-propagation training process requires that observations are presented to the neural network together with their respective outcomes. In this case an observation is the set of all relevant tokens comprising an e.mail.

3.3.2 Translations

Although, at first sight, the ability to express the same idea in different forms may make a text more interesting to be read, it does add complexity to its understanding. For example, words in a foreign language and abbreviations are frequently used to convey the cultural aspect of a situation and make the text shorter respectively. However, in these circumstances, readers are required to master not only the meaning of words outside their natural tongue, but also the meaning of a considerable range of abbreviations.

Moreover, despite the wide availability of spell-check software on the market, the misspelling of words in e.mails is not an uncommon event, even when they are sent by university students. One way to reduce the extra complexity of having to identify the meaning of words with unstandardized spelling is to translate them into their phonetic form. In these circumstances, for example, “range”, “rrange”, “rannge” etc., may all be translated into “reindɜ”, its common phonetic form.

By translating abbreviations into their extended form, foreign words into a standard language, and, afterwards, any word in general in its phonetic form, one makes it easier for a neural network to recognize the contents of electronic messages.

3.3.3 Stop-Words Removal

Stop-words are either functional or connective words that lack discrimination power when used to search for information in a data base, classify texts, infer the meaning of electronic messages etc. Therefore, the removal of stop words facilitates the recognition of patterns in electronic messages at the same time that it does not imply the loss of relevant information.

Commonly used words in a given language are the most likely candidates to be listed as stop-words, although there are cases in which even infrequently used words suffer from the same absence of discrimination power [10]. The following are examples of words that are customarily listed as stop words in English: “a”, “an”, “the”, “in”, “of”, “on”, “are”, “be”, “if”, “into” and “which”. See [4] for lists of stop words in a variety of western languages, including the English language.

3.3.4 Selection of the Most Relevant Words

Despite all the transformations the e.mails provided by SEBRAE have been subjected to so far, they are likely to contain words with different discrimination power when it comes to the identification of which reply each e.mail requires. While some words may be highly relevant to the task, others will be almost as irrelevant as stop-words.

To determine the discrimination power of a word regarding a specific reply, it suffices to calculate how well this word can separate messages that should be properly answered with that reply from the others. This can be achieved with support of the widely used Gini diversity index. The index was initially proposed by Corrado Gini (1884-1965) and later adapted by Breiman *et al.* [2] for the development of classification methods.

In formal terms, for a given set of observations O containing *n* elements, and a class *J* of objects, Gini is given by $I(O) = 1 - S$, where $S = \sum_{i=1}^{n} P(j_i|O)^2$ for $j_i \in J$, and $P(j_i|O)$ is the probability of occurrence of objects j_i in *O*.

For example, consider the word “deadline” and the standard reply $\mathcal{C}$, which informs the participants of the SEBRAE Challenge about the closing date of the different rounds of the game. Suppose that among all the 1,596 e.mails provided by SEBRAE, 230 should be replied with $\mathcal{C}$. Table 4 summarizes these figures. In that table reply $\mathcal{A}$ stands for any other reply but $\mathcal{C}$.

Reply	E-mails	Percentage
$\mathcal{C}$	230	14,41%
$\mathcal{A}$	1,366	85,59%
Total	1,596	100,00%

Table 4. Distribution of replay $\mathcal{C}$ among the e.mails provided by SEBRAE.

In this circumstance Gini is given by $1 - (0,1441^2 + 0,8559^2) = 0,2467$. If there are only two objects in J, as is the case in Table 3, then Gini varies between 0 and 0.5, 0 indicating a perfectly homogenous set of observations and 0.5 indicating that each object in J has 50% chance of occurring. Therefore, a value close to the middle point of the variation interval like 0,2467 indicates that the data presented in Table 3 is reasonably disperse in regard to replies $\mathcal{C}$ and $\mathcal{A}$.

Now, consider the occurrence of the word “deadline”. Suppose that it occurs in a total of 206 e-mails and that out of these 184, i.e. 89.32%, should be replied with $\mathcal{C}$. Table 5 presents these figures. In this case Gini is given by $1 - (0,8932^2 + 0,1068^2) = 0,1908$, indicating that this is more homogenous set of values than the one presented in Table 4.

E-mails containing “deadline”		
Reply	**E.mails**	**Percentage**
$\mathcal{C}$	184	89,32%
$\mathcal{A}$	22	10,68%
Total	206	100,00%

Table 5. Distribution of reply $\mathcal{C}$ among e.mails containing the word “deadline”.

However, because the absence of a given word in a e-mail may also indicate that a certain reply should be used, one has to consider the situation where the word “deadline” is not present. Table 6 presents the corresponding figures. In this circumstance Gini is given by $1 - (0,0331^2 + 0,9669^2) = 0,0640$, indicating that the absence of “deadline” also leads to a set of values that is more homogenous than the one presented in Table 4.

Finally, one has to check how well the presence or absence of “deadline” in an e.mail indicates the occurrence of reply $\mathcal{C}$ and the occurrence of other replies. This can be achieved by subtracting the weighted average Gini of the e.mails with and without the word “deadline” from the diversity index corresponding to the whole set of e.mails. The bigger the different, the greater the discrimination

E-mails without "deadline"		
Reply	**E.mails**	**Percentage**
$\mathcal{C}$	46	3,31%
$\mathcal{O}$	1,344	96,69%
Total	1,390	100,00%

Table 6. Distribution of reply $\mathcal{C}$ among e.mails without the word "deadline".

power. In this case the proportions of e.mails falling in each case are used as weights, i.e. $\Delta I = 0,2467 - \left(0,1908 \times \frac{206}{1,596} + 0,0640 \times \frac{1,392}{1,596}\right) = 0,1663$

A positive difference in the value of Gini like 0,1663 indicates that the two sets of values resulting from grouping together messages with and without the word "deadline" are more homogenous on average then the set of values presented in Table 3. Therefore, "deadline" does have some discrimination power.

Because Gini tends to yield small numbers, its meaning is better understood when comparing different situations. For example, a word that appears in every e.mail inevitably leads to a situation identical to the one presented in Table 3, yielding exactly to the same Gini and, as a result, no improvement in diversity. Also, a word that occurs in just a few e.mails requiring a frequently used reply is certain to lead to an insignificant reduction in diversity.

By removing from the e-mails provided by SEBRAE words with reduced discrimination power, one makes it easier for an inferential engine to identify which reply should be used in different circumstances. One could rightfully argue that Gini can be used to remove stop-words. Although this can be done, it should be kept in mind that the process of calculating the diversity index is far more expensive than the task of looking for words in a list of well-known stop words.

Temple uses Gini to evaluate the benefits of economic growth [17]. Rokach and Maimon examine both Gini and other alternative ways of estimating the discrimination power of objects in a variety of different contexts [14].

3.3.5 Codification of Words into Binary Numbers

At this point in the data transformation process the e-mails provided by SEBRAE are left with words that are not stop words and that have been found to have some discrimination power. Although, each of these words has its own discrimination power, all of them are potential candidates to be presented to the neural network that serve as an inferential engine to the SEBRAE Challenge's hybrid QAS.

However, because neural networks can only deal with numerical values, these words must be transformed into numbers before they are presented to the network. A straightforward way of doing this is to indicate the presence or absence of a word with the numbers 1 and 0 respectively.

In this case, indicating the presence or absence of *n* different words in an e-mail may be accomplished with the support of *n* inputs to the neural network. If the occurrence of a given word *W* is associated with input x_i, for $1 \leq i \leq n$, then $x_i = 1$ indicates the occurrence of *W* and $x_i = 0$, its absence.See [18] for alternative ways of coding words into numerical values.

3.4 The Neural Network

A series of experiments where carried out in order to provide an adequate neural network to be used as the inferential engine that powers the SEBARE Challenge's hybrid QAS

All the networks used in these experiments share some common properties. For example, they all have exactly one fully-connected hidden layer of neurons, i.e., every neuron in the hidden layer are connected to all inputs and outputs. Also, following advice from both Sinha [16] and Jefferson *et al.* [8], the number of neurons in the hidden layer is set to be the integer immediately greater or equal to the geometric mean of the number of inputs and outputs[1].

To offer an unbiased estimation of the error committed by the neural network in each experiment, the e.mails provided by SEBRAE are divided into two different sets. While the first and larger set is used to train the network, the second and smaller set is used to estimate the number of correct answers yielded by the network. Table 7 presents the corresponding figures.

Training Set		Test Set		Total	
Qtd.	%	Qtd.	%	Qtd.	%
1,117	70,0	479	30,0	1,596	100,0

Table 7. Quantity of e.mails used to train and test the neural network.

Because only five standard replies can be used to answer the vast majority of all the e-mails provided by SEBRAE (see Table 2) and one can relay on the support of a telemarketing operator to answers messages for which a standard reply has not been properly provided, in all experiments the neural network was trained to deal with e.mails requiring six different kinds of standard replies.

While five of these standard replies are precisely those used to answer 91.3% of the e-mails provided by SEBRAE, the sixth provides a pseudo standard reply to the remaining 8.7% of the e-mails. The idea behind the use of a pseudo reply is that the actual reply to certain questions posed by

[1] If p e q are numbers, than $\sqrt{p \times q}$ is the geometric mean of p and q.

participants of the SEBRAE Challenge should be provided by a telemarketing operator in a customized manner. Therefore, when one of these remaining e-mails is detected, instead of indicating which standard reply should be used by the QAS, the neural network concedes that a telemarketing operator should be called in.

4 The Experiments and Related Results

In the manner that a human being can still understand the meaning of a sentence from which words or sounds have been removed, an artificial neural network can be trained to identify which reply should be used to answer an e-mail from a reduced subset of its contents. Therefore, the question that designers of neural-network QAS's face is to determines the smallest set of words that hold sufficient discrimination power to allow e-mails to be properly replied.

With regard to development of the SEBRAE hybrid QAS, five different experiment have been carried out with this objective in mind. In each experiment the set of most relevant word were gradually enlarged until no further improvement in the performance of the network was reached.

In order to avoid bias towards certain standard replies to the detriment of others, the same number of most relevant words have been selected for each of the six replies the network was trained to recognize when to use. Table 8 contains total number of relevant words used in each experiment.

For example, in the first experiment the twenty three most relevant words were selected to train the network. In the fifth and last experiment the fifty most relevant words have been selected. However, because some words have discrimination power to recognized the need to use more than one standard reply, the total number o relevant words in each experiment is smaller than six times the number of relevant words per standard reply.

Experiment	Most Relevant Words	
	Per Standard Reply	Total
1	5	23
2	8	33
3	10	38
4	12	44
5	15	50

Table 8. Quantity of most relevant words used in each experiment.

It should be noted that because each most relevant word corresponds to a neural network input, the number of inputs used in each experiments varies accordingly. However, the number of outputs is always the same, i.e., the number of replies the network is trained to recognize, six in this case.

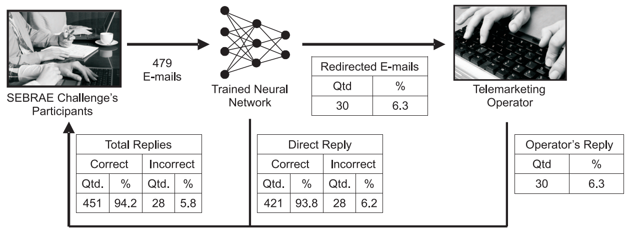

Figure 5. Summary of the last experiment in the pursue of a suitable inferential engine.

Table 9 presents the precision of the answers provided by the neural network in the different experiments. While in the first experiment the network was able to answer properly 87.8% of the e-mails in the test set, this figure grew to 92.5% in the last experiment, in which a larger set of most relevant words was used. Unfortunately, neither further increase in the number of the most relevant words nor in the number of hidden layers of neurons were followed by an increase in the number of properly answered e-mails.

Experiment	Replies					
	Correct		Incorrect		Total	
	Qtd.	%	Qtd.	%	Qtd.	%
1	397	91.9	35	8.1	432	100.0
2	405	92.5	33	7.7	438	100.0
3	412	93.2	30	6.8	442	100.0
4	414	93.7	28	6.3	442	100.0
5	421	93.8	28	6.2	449	100.0

Table 9. Precision of the answers provided by the neural network in each experiment.

In addition, that data presented in Table 9 reveal two important pieces of information. First, the precision of the answers provided by the network increases much slower than the increase in the number of most relevant words. This is due to the fact that as the experiments proceeded words with decreasing discrimination power were added to the set of most relevant words.

Second, not all e-mails were answered by the network. Only e-mail for which an actual standard reply has been provided are directly replied by the network. E-mails that does not fit in this category are forwarded to a telemarketing operator, who becomes responsible for providing the right answer. For example, in the last experiment the telemarketing operator was required to answer 30 e-mails, i.e., out of the 479 e-mails used for testing the neural network answered 449 and the telemarketing operator only 30. Figure 5 summarizes these figures.

5 Discussion

At the outset of this article we undertook to present a stepwise approach to the construction of hybrid question answering systems based upon neural network technology and natural language processing. Below we answer some key questions about the construction of such systems and discuss the implications of its existence for different dimensions of e-commerce and business-to-customer relationship.

5.1 Why should hybrid neural-network question answering systems be build?

Among all the many features that e-commerce sites have in common, one of the most popular is certainly the presence of an institutional e-mail through which customer support service can be contacted. In today's virtual market it is actually quite unusual to find a competitive e-commerce site that does not provide such facility.

If an e-commerce site is visited regularly by a considerable number of customers, a fraction of those customers contact the customer support service via e-mail, and customer communication is properly recorded, then, in the course of time, a large database of questions and answers will inevitably be available for analysis. Although, this is not the case for all e-commerce sites, it is surely the case for a large number of them worldwide.

Taking these databases as a starting point, companies can build hybrid neural-network question answer systems that (a) make it possible to deliver accurate high-speed answers to questions posed by customers; (b) allow customers to receive answers to their questions on a 24/7 basis and (c) provide well conceived standard answers to those questions.

All of this tend to increasing customer satisfaction, favoring repeated purchases of products and services. Moreover, well-conceived prompt answers to questions posed by customers may help to reduce cognitive dissonance[2], improving the quality of after-sale services and actions.

Therefore, not only the technology to build hybrid neural-network question answering system is available, but also its use has a potentially positive impact on customer satisfaction before and after a purchase is made and, as a result, on a company's overall sales initiatives.

5.2 Do Hybrid Neural-Network QAS's provide additional benefits for customers?

As organizations start replying to e-mails in a standardized and automatic way, in opposition to the relatively non-standard replies provided by telemarketing operators, they end up becoming potentially even more responsible before society and the law for the answers they supply. After all, mistakes made in an inadequate answer given by a telemarketing operator can always be attributed to human error, employee momentary dissatisfaction with an organization or even misunderstanding on the part of the consumer due to a reply badly formulated.

Fortunately, in general, organizations have more difficulty in exempting themselves from the responsibility for answers that can be revised in detail with great antecedence, as it is the case when using hybrid neural-network QAS's. The increase in responsibility naturally favors clarity and precision on the replies that are given. That combined with the promptness of responses provided by automated systems allows consumers to spend less time searching for answers to the questions they pose.

Finally, hybrid neural-network QAS's admit a precise electronic recording of customer communication. This recording capacity opens a new dimension in the support of marketing decisions as, in the course of time, each standard reply may be associated to changes in the attitude of customers towards the organization.

While some replies can positively influence customers's propensity to make new purchases, others may have the opposite effect. The capacity to distinguish between these two cases offers the possibility of making adjustments on the replies and policies that cause dissonance in the customer's mind, improving service quality and, as a result, favoring new sales.

5.3 Are hybrid neural-network QAS's hard to be build?

In order to build hybrid QAS's similar to the one described in this article one has to acquire competence in many different areas of expertise such as: artificial neural network algorithms, natural language processing, data storage and recovery, Internet technology, information systems development etc. Fortunately, all this knowledge is largely available in any well conceived undergraduate program in computer science. So, lack of specialized labor is far from being an obstacle to building hybrid neural-network QAS's.

Moreover, nowadays, neural-network software is widely available. While some full-fledged data analysis software containing a neural-network module may still be considered an expensive investment by some companies, there are plenty of quality standard-alone neural-network software at reduced price on the market. Some of them are even sold as adds-in to Microsoft Excel, i.e., can be run directly from an Excel spreadsheet[3].

[2] A state of psychological uneasiness arising when consumers try to reconcile conflicting states of mind. For example, the positive feeling of having decided to buy a product or service and the negative feeling of being disappointed with it afterwards.

[3] One can find neural-network software for as low as US$ 60 at www.download.com.

Finally, there is the question of providing standard answers to questions posed by customers. However, this is an activity that is routinely carried out in any well-structured telemarketing service, where operators are thoroughly trained to provide the same answer to a large variety of similar questions. Therefore, if one has the necessary data, there is no major obstacle preventing companies from building hybrid question-answering systems, but the willingness to do so.

5.4 Are hybrid QAS's hard to be maintained?

Another important point that favors the use of hybrid QAS's is that once they are in place, maintenance becomes a straightforward process, i.e. (a) revise the standard replies that are already being used, making the necessary adjustments; (b) questions that require human assistance to reply have to be routinely scrutinized in search of proper standard replies; (c) once proper replies have been found, those messages along with their replies should become part of the sample of messages and replies that are used to train the neural network; (d) adjust the network design parameters such as number of inputs, number of outputs and number of neurons in the hidden layers; and (e) properly train and deploy the network.

If there is a well-structured telemarketing service in place and communication with customers is recorded for further analysis, then the first three steps in the maintenance process of hybrid neural-network QAS's are closely related to the many tasks that telemarketing management carries out routinely. Therefore, one should not face any major difficulties in having them implemented.

The last two steps, however, are better performed by a professional with knowledge and experience in the use of neural network. Fortunately, such specialized labor is neither hard to be found nor difficult to be developed through professional training.

5.5 Can the choice of human-machine interface make maintenance easier?

Consumers natural tendency to utility maximization, i.e., to obtain the maximum satisfaction possible from any investment that is made (including time), favor the conception of e-mails where several questions are posed concurrently [6]. Such e-mails not only increase the complexity of inferring what answers should be given to questions posed by consumers, but also make the process of neural-network QAS maintenance more expensive and time consuming.

However, all this inconvenience may be avoided if customers are made aware that they are dealing with a hybrid neural-network QAS and that, as a result, they are expected to pose one question per e-mail. A proper design of human-computer interface may be used to enforce such minor restriction and make it more acceptable concomitantly.

5.6 What are the main advantages of having a neural network as an inferential QA engine?

When dealing with e-mails written in natural language it is important to keep in mind that this is frequently an imprecise form of communication, where questions are sometimes vague, noisy and only partly posed. Even when such e-mails go through a process of transformations with the view of making understanding easier, the final result may still suffer from some of these undesired properties.

Furthermore, even when a large number of e-mails are available for analyzes with the purpose of building a QAS, these are nevertheless only a sample of all the future e-mails that our customers are likely send us. Therefore, in order to be able to answer questions posed in natural language by automatic means it is advantageous to make use of an inferential engine that (a) has been widely tested in many different circumstances; (b) is able to identify patterns by example; (c) can extrapolate the acquired knowledge to handle unforseen situations; (b) deal well with noise; (e) recognize partners even if variables hold non-linear relationships among themselves and populations distributions are unknown; and (f) can be efficiently executed in a parallel computer architecture, if the problem in hand turns out to be highly complex. These are precisely some of the properties enjoyed by artificial neural networks, making this class of algorithms a very interesting option to be used in many circumstances as QAS's inferential engines.

6 Conclusions

One of the major differences between the traditional mortar and brick market of products and services we got used to and the virtual market that was created with the advent of the Internet is the easiness with which customers may compare the benefits of acquiring products and services from a multitude of different providers.

In the very competitive virtual market, where many products and services are offered in a global scale, it is becoming increasingly difficult for organizations to avoid providing services on any other basis but 24/7. After all, one of the main characteristic of the Internet is its around-the-clock availability .

During the process of product acquisition over the Internet, it is natural that potential buyers may want to get in touch with customer support services virtually in search of information that they could not find in a web site. Failure in providing such information promptly may lead to customer dissatisfaction and cognitive dissonance, that may impact negatively on customer-to-business relationship and, as result, on sales.

In this article we have demonstrated the viability of building hybrid question answering systems that are able to properly answers questions posed by customers via e-mail in natural language. These systems, that use artificial neural networks as their inferential engine, are not difficult to be built and maintained, yielding accurate results even in the presence of noise.

If the necessary data is available, there is no major obstacle preventing companies from enjoying the benefits of hybrid question-answering systems, but the desire to do so. As a consequence, it should not come as a surprise if in the years to come an increasing number of organizations elect to build neural-network hybrid question-answering systems, with a positive impact not only on the intensity of the relationship with customers, but also on the quality of services they provide.

References

[1] Jon Anton and Mike Murphy. *Managing Web-Based Customer Experiences: Self-Service Integrated with Assisted-Service*. The Anton Press, August 2003.

[2] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. Chapman & Hall/CRC Press, January 1984.

[3] Curt Champion. Taking advantage of web self-care to meet client needs. *Customer Interaction Solutions*, 22(5):48–50, November 2003.

[4] Institut Interfacultaire d'Informatique - University of Neuchatel. Have a look at the CLEF site (european languages) or NTCIR (asian languages) providing other information about multilingual retrieval. Information available on the Internet at www.unine.ch/info/clef/, Switzerland, 2005. Site last visited on February 8^{th}, 2006.

[5] Gérard Dreyfus. *Neural Networks*. Springer-Verlag, March 2005.

[6] Gordon R. Foxall, Jorge M. Oliveira-Castro, and Teresa C. Schrezenmaier. The behavioral economics of consumer brand choice: patterns of reinforcement and utility maximization. *Behavioural Processes*, 66:235–260, 2005.

[7] Madan M. Gupta, Liang Jin, and Noriyasu Homma. *Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory*. John Wiley & Sons-IEEE Press, 1^{st} edition, April 2003.

[8] M. F. Jefferson, N. Pendleton, S. B. Lucas, and M. A. Horan. *Artificial Neural Networks in Cancer Diagnosis, Prognosis, and Patient Management*, chapter 5: The Use of Genetic Algorithm Neural Network (GANN) for Prognosis in Surgically Treated Nonsmall Cell Lung Cancer, pages 39–53. CRC, 1^{st} edition, June 2001.

[9] V. Kürková. Kolmogorov's theorem is relevant. *Neural Computation*, 3(4):617–622, 1991.

[10] Sai Ho Kwok. P2P searching trends: 2002-2004. *Information Processing and Management*, 42:237–247, 2006.

[11] Mark T. Maybury. *New Directions in Question Answering*, chapter 1: Question Anserwing: An Introduction, pages 3–14. AAI Press / The MIT Press, 2004.

[12] Solomon Negasha, Terry Ryanb, and Magid Igbariab. Quality and effectiveness in web-based customer support systems. *Information & Management*, 40:757–768, 2003.

[13] Juan R. Rabunal and Julian Dorrado. *Artificial Neural Networks in Real-life Applications*. Idea Group Publishing, November 2005.

[14] L. Rokach and O. Maimon. Top-down induction of decision trees classifiers: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 35(4):476–487, November 2005.

[15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, chapter Learning internal representations by error propagation, pages 318–362. The MIT Press, Cambridge, MA, 1986.

[16] A. K. Sinha. Short term load forecasting using artificial neural networks. In *IEEE International Conference on Industrial Technology 2000*, pages 548–553, Mumbai, India, 19-22 January 2000. Jaico Publishing House.

[17] Jonathan R. W. Temple. Growth and wage inequality in a dual economy. *Bulletin of Economic Research*, 57(2):145, April 2005.

[18] Craig A. Wendorf. Primer on multiple regression coding: Common forms and the additional case of repeated contrasts. *Understanding Statistics*, 3(1):47–57, 2004.