ABSTRACT

New age telecommunication sector (GSM) has been based on new customer acquisition plans and tactics. For this purpose, evaluation of the call traffic data from the data warehouses and making some predictions by data mining algorithms are so important. Data mining is a process for filtering massive amounts of data to find useful information to optimize decision making in the firms. Data mining methods use the Neural Networks as a tool for mining data from the data warehouses. In our case, we use Cellular Neural Networks to predict future usage of the GSM lines and data services and try to estimate some promotions to gain market opportunity in the Turkish market for the ARIA GSM Company. This work focus on for defining patterns and implementing Cellular Neural Networks to gain knowledge acquisition from the data warehouses and determining the customer profiles and behaviors.

KEY WORDS
Data Mining, Neural Networks, Cellular Neural Networks, Very Large Databases

1. Introduction

Previous data mining works, show us that neural networks can be used effectively in the classification process and give a lower classification error rate than the decision trees, but require longer learning time [1]. Also, classification rules can be extracted from a neural network based methods [2]. In reference [3], an intelligent system is proposed for identifying customers by neural networks and genetic algorithms [4]. Data mining algorithms can be used in customer relationship management by supporting with neural network algorithms [5, 6].

Cellular phone usage is very important for the GSM operators. Both voice and data services are basic revenue gaining items for them. Thus, GSM operators should have to predict their customer behaviors to keep them registered. Managers want to understand why some customers remain loyal while others leave from the telecom company. By this viewpoint, they can be constructed a model that derived from historical data of loyal and left customers. The model illustrates the customer loyalty and churn rate. This model consists of the all data mining activities.

For this reason, we use Cellular Neural Networks (CNN) to detect patterns in data, generalize relationships found in the data, and predict outcomes. First of all, we preprocess the raw data from the data warehouses, transform the cleaned data and put them into CNN model to train the neural network. Cellular neural networks share the best features of both worlds; its continuous time feature allows real-time signal processing feature makes it tailor made for different implementation purpose. Cellular neural networks are uniquely suited for high-speed parallel signal processing. From this viewpoint, we adapted the CNN to use in our data mining problem.

After the development phase, we use some partitions from our data provider’s data warehouse as a raw data for preprocessing and putting them into the KDD(knowledge discovery in databases) process.

Fundamental objectives of this research are

• Testing Cellular Neural Network usage in the real life problem to get some important responses from the data warehouse system efficiently then the other Neural Networks.
• Creating an alternative data-mining model for the GSM Operators.
• Arranging new tool for the data-mining purpose on the Oracle database.

Paper is organized as follows. Section 2 describes the data mining algorithms and focus on Neural Networks method. Section 3 defines the problem. In the 4th section, Cellular Neural Networks are used to modeling mining strategies. According to the defined model in the 4th section, creating and testing CNN is given in the 5th section. Section 6 covers summary and conclusions of the investigation.

2. Data Mining Algorithms

Information is derived from the databases by the statistical methods and approaches. Input and expected
outputs are correlated with formulas. But, these type
methods are working slow and taking so much time and
space cost according to the algorithm design step.

In the data mining process, data mining algorithms that
are categorized as supervised and unsupervised
techniques have queried data. Also, data mining is the
process of finding trends and patterns to discover
information. The benefits of data mining is to turn this
newfound knowledge into actionable results, such as
increasing a customer’s likelihood to buy, or decreasing
the number of customer churn in a GSM company.

A pattern is an event or combination of events in a
database that occurs more often than expected. Typically
this means that its actual occurrence is significantly
different than what would be expected by random choice.
A model is a description of the original historical database
from which it was built that can be successfully applied to
new data in order to make predictions about missing
values or to make statements about expected values.

The goal of a predictive model is to predict future
outcomes based on past records with known answers.
Two approaches are commonly used to generate models:
supervised and unsupervised.

Supervised or directed modeling is goal-oriented. The
task is to explain the value of some particular field. The
user selects the target field and directs the computer to tell
how to estimate, classify, or predict it. In unsupervised
modeling is used to explain those relationships once they
have been found. [7, 8]

2.1. Supervised Modeling Techniques
The two most common supervised modeling techniques
are classification and regression. If the data element under
investigation is discrete (meaning that it has a small
number of fixed values) the task is called classification.
On the other hand, if the data element is continuous
(meaning that it can take on a large number of values and
exhibits a common underlying unit of measure) the task is
called regression. Classification predicts class, or group,
membership. With classification, the predicted output (the
class) is categorical; a categorical variable has only a few
possible values, such as yes–no, high–middle–low, etc.
Regression predicts a specific value [9]. Techniques that
dominate the commercially available classification and
regression tools today include:
• Decision trees,
• Neural networks,
• Naïve Bayes and K-nearest neighbor.

2.1.1. Neural Networks
There are two main types of neural network models:
supervised neural networks such as the multi-layer
perceptron or radial basis functions, and unsupervised
neural networks such as Kohonen feature maps. A
supervised neural network uses training and testing data
to build a model. The data involves historical data sets
containing input variables, or data fields, which
correspond to an output.

The training data is what the neural network uses to
“learn” how to predict the known output, and the testing
data is used for validation. The aim is for the neural
networks to predict the output for any record given the
input variables only. A neural network starts with an input
layer, where each node corresponds to a predictor
variable. These input nodes are connected to a number of
nodes in a hidden layer. Each input node is connected to
every node in the hidden layer. The nodes in the hidden
layer may be connected to nodes in another hidden layer,
or to an output layer. The output layer consists of one or
more response variables. [10, 11]

<table>
<thead>
<tr>
<th>Table 1-Advatanges and Disadvantages of Neural Networks [12]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages of Neural Networks</strong></td>
</tr>
<tr>
<td>High Accuracy: Neural networks are able to approximate complex non-linear mappings.</td>
</tr>
<tr>
<td>Noise Tolerance: Neural networks are very flexible with respect to incomplete, missing and noisy data.</td>
</tr>
<tr>
<td>Independence from prior assumptions: Neural networks do not make a priori assumptions about the distribution of the data, or the form of interactions between factors.</td>
</tr>
<tr>
<td>Ease of maintenance: Neural networks can be updated with fresh data, making them useful for dynamic environments.</td>
</tr>
<tr>
<td>Hidden nodes, in supervised Neural networks can be regarded as latent variables.</td>
</tr>
<tr>
<td>Neural networks performance can be highly automated, minimizing human involvement.</td>
</tr>
<tr>
<td>Neural networks are especially suited to tackling problems in non-conservative domains.</td>
</tr>
<tr>
<td><strong>Disadvantages of Neural Networks</strong></td>
</tr>
<tr>
<td>Poor Transparency: Neural networks operate as “black boxes”.</td>
</tr>
<tr>
<td>Trial-and-error design: The selection of the hidden nodes and training parameters is heuristic.</td>
</tr>
<tr>
<td>Data hungry: Estimating the network weights requires large amounts of data, and this can be very computer intensive.</td>
</tr>
<tr>
<td>Over-fitting: If too many weights are used without regularisation, Neural network become useless in terms of generalisation to new data.</td>
</tr>
<tr>
<td>Neural networks are totally dependent on the quality and amount of data available.</td>
</tr>
<tr>
<td>Neural networks may converge to local minima in the error surface.</td>
</tr>
<tr>
<td>Neural networks lack classical statistical properties. Confidence intervals and hypothesis testing are not available.</td>
</tr>
<tr>
<td>Neural network techniques are still rapidly evolving and they are not yet robust.</td>
</tr>
</tbody>
</table>

3. Problem Definition

GSM operators always need to determine churns and
loyal customers to prevent missed customers to the other
GSM operators. For this purpose, we should find some
answers for the below questions.
- How many days has he used his GSM line since he was registered?
- How many minutes did he use his line?
- What is the usage trend of him?
- What is the level of the customer satisfaction?
- Is customer still with company (loyal) or no longer customer (lost)?

In this model, GSM Cell Company would like to determine current customer (OPEN), churned customer (CLOSED) and also hoppers (TENDENCY). According to the solutions, they try to prevent churns by new promotions.

3.1. Selecting Input and Dependent Variables
The field Gsm_Line_Status is selected as dependent variable from the Table 2 and it shows that whether the customer is still a registered (OPEN), churned (CLOSED) or have tendency to churn (HOPPER).

Churn Prediction is start with analyzing demographic data, subscription data, usage data and churn data from the data warehouse according to the Figure 1. Gsm_Line_Status field is selected as a dependent variable and therefore all the input data should give the solutions, as line is active, closed or tendency to close.

In the case of churned customers, the field values reflect conditions that existed at the time the customer left. Table 2 is detailed input and dependent (computed data) data.

### Table 2-Churn Management in the GSM Cell Company

<table>
<thead>
<tr>
<th>Column</th>
<th>Values</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSISDN</td>
<td>Integer</td>
<td>Unique phone numbers</td>
</tr>
<tr>
<td>State</td>
<td>Marmara, Ege, Akdeniz, Karadeniz, Orta Anadolu, Doğu Anadolu, Güney Anadolu</td>
<td>Cellular phone used area</td>
</tr>
<tr>
<td>City</td>
<td>City name</td>
<td>Cities in the Turkey</td>
</tr>
<tr>
<td>Marriage</td>
<td>Single, Married, Widowed, Divorced</td>
<td>Marital status of customer</td>
</tr>
<tr>
<td>Age</td>
<td>Integer: 18+, 25+, 35+, 45+</td>
<td>Customer age</td>
</tr>
<tr>
<td>Child</td>
<td>Integer</td>
<td>Number of child</td>
</tr>
<tr>
<td>Occupation</td>
<td>Professional, Manager, Executive, Worker, Other</td>
<td>Description of type of occupation</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, Female</td>
<td>Gender</td>
</tr>
<tr>
<td>Home</td>
<td>Rent, Own</td>
<td>Own home or not.</td>
</tr>
<tr>
<td>Month_Expense</td>
<td>Currency: 1bil.TL-, 1bil.TL+, 2bil.TL, 3bil.TL, 4bil.TL</td>
<td>Monthly expenses of customer</td>
</tr>
<tr>
<td>Month_Income</td>
<td>Currency: 1bil.TL-, 1bil.TL+, 2bil.TL, 3bil.TL, 4bil.TL</td>
<td>Monthly income of customer</td>
</tr>
<tr>
<td>Subscription_Len</td>
<td>Low: less than 6 months, Medium: less than 1 year, High: less than 3 years, Very High: more than 3 years.</td>
<td>Customer subscription period. Durations are segmented into four range.</td>
</tr>
<tr>
<td>Credit_Limit</td>
<td>Low, Medium, High, Very High</td>
<td>Customer credit limit. Segmenting into four range.</td>
</tr>
<tr>
<td>Cust_Segment</td>
<td>A,B,C,D,E,F,G</td>
<td>Customer segments derived from statistical analysis</td>
</tr>
<tr>
<td>Avg_Len_Call_Mon</td>
<td>Low, Medium, High, Very High</td>
<td>Average length of calls this month</td>
</tr>
<tr>
<td>Avg_Len_Call_1_3Mon</td>
<td>Low, Medium, High, Very High</td>
<td>Average length of calls one to three month</td>
</tr>
<tr>
<td>Avg_Len_Call_4_6Mon</td>
<td>Low, Medium, High, Very High</td>
<td>Average length of calls four to six month</td>
</tr>
<tr>
<td>Tariff</td>
<td>Standard, Basic, Business</td>
<td>Type of tariff</td>
</tr>
<tr>
<td>Count_Peak_Min</td>
<td>Integer</td>
<td>Number of peak minutes this month</td>
</tr>
<tr>
<td>Count_Peak_Min_1_3M</td>
<td>Integer</td>
<td>Number of peak minutes one to three months</td>
</tr>
<tr>
<td>Count_Roam_Call</td>
<td>Integer</td>
<td>Number of roaming calls this month</td>
</tr>
<tr>
<td>Count_Roam_Call_1_3M</td>
<td>Integer</td>
<td>Number of roaming calls one to three months</td>
</tr>
<tr>
<td>Count_Roam_Call_4_6M</td>
<td>Integer</td>
<td>Number of roaming calls four to six months</td>
</tr>
<tr>
<td>Total_spent</td>
<td>Currency</td>
<td>Total fair</td>
</tr>
</tbody>
</table>

In the data mining and knowledge acquisition works, start with data preparation and clearing data from the noise data. Derived fields as listed in Table 1, computed from
4. Cellular Neural Network Approach

Cellular Neural Networks (CNN) model have been defined firstly in 1988 by Chua. The basic circuit unit of cellular neural networks is called a cell. It contains linear and non-linear circuit elements. These are linear resistors, linear and non-linear controlled sources and independent sources.

The adjacent cell can interact directly with each other. An example of a two-dimensional cellular neural network is shown in Figure 3. Theoretically, it can be defined for any dimensions. [13, 14]

![Figure 3-A two-dimensional Cellular Neural Network](image)

In Figure 3 the circuit size is 3x3. The squares are the circuit units called cells. The links between the cells indicate that there are interactions between the linked cells.

In MxN dimensions cellular neural network has MxN cells arranged in M rows and N columns. The cell on the i th row and the j column denote by C(i,j) as in Fig. 3. A neighborhood of C(i,j) (r-neighborhood) define with this equation:

\[ N_r(i,j) = \{C(k,l) | \max(k-i,|l-j|) \leq r \} \]

where r is positive integer number.

Figure 4 shows 3 neighborhoods of the same cell with r=1,2 and 3. r-neighborhood contains \((2r+1)^2\) cells and this is symmetric. It can be not only in square structure but also in hexagonal structure.

An example of a cell circuit shows in Figure 5. C is linear capacitor. \(R_x\), \(R_y\) are linear resistors. I is independent voltage source. \(I_{xu}(i,j;k,l)\) and \(I_{xy}(i,j;k,l)\) are linear voltage-controlled current sources with the characteristics \(I_{xy}(i,j;k,l) = A(i,j;k,l)v_{yl} \) and \(I_{xu}(i,j;k,l) = B(i,j;k,l)v_{ul} \) for all \(C(i,j) \in N(i,j)\). \(I_{x} = (1/R_x) f(V_{xij})\) is a piecewise-linear voltage-controlled current source with its characteristics \(f(.)\) as shown in Figure 6. \(E_j\) is an independent voltage source.

All of the linear and piecewise-linear controlled sources used in our cellular neural networks can be easily realized using operational amplifiers.

Each cell of a cellular neural network has at most three nodes (Sometimes we will choose \(E_{ij}=0\) if \(B(i,j;k,l)=0\) for all cells in a cellular neural network. In this case there are only two nodes in a cellular circuit.) Since all cells have the same datum node, and since all circuit elements are voltage controlled, cellular neural networks are ideally suited for nodal analysis.
The only nonlinear element in each cell is a piecewise-linear voltage-controlled source \( I_y = \left( \frac{1}{R_y} \right) f(v_{xij}) \) with characteristic \( f(.) \) as shown in Figure 4. The equation of function is:

\[
f(x) = \frac{1}{2} \left[ \left| x + 1 \right| - \left| x - 1 \right| \right]
\]  

(3)

It is the activation function for CNN algorithm and outputs can be obtained according to this function.

Figure 6-The characteristics of the nonlinear controlled source

4.1. CNN Algorithm

Here, \( U \) is \( M \times N \) input, \( X \) is state and \( Y \) is output matrices. \( I \) is bias \( 1 \times 1 \) scaler.

\[
\frac{dX}{dt} = -X + A \times Y + B \times U + I
\]  

(4)

The equation can be shown as a state and output:

\[
X_{i,j}(n+1) = \sum_{(k,l) \in N_{i,j}} A_{i,j,k,l} \cdot y_{k,l}(n) + \sum_{(k,l) \in N_{i,j}} B_{i,j,k,l} \cdot u_{k,l}(n) + I_{i,j}
\]  

(5)

\[
y_{i,j}(n) = f(x_{i,j}(n))
\]  

(6)

\( A_{i,j} \) is feedback weight constant matrice and \( B_{i,j} \) is input weight constant matrice. The index \( i,j \) state the cell. The algorithm works as an iterative filter that converts \( x_{i,j}(n) \) to \( x_{i,j}(n+1) \).

In this work we used Recurrent Perceptron Learning Algorithm (RPLA). This algorithm uses 3x3 dimensions for \( A \) and \( B \) matrices but it can be used for different templates. For stability \( A \) and \( B \) should be symmetrical and \( A_{2,2} \) must be greater than 1.

\[
A = \begin{bmatrix}
A_{1,1} & A_{1,2} & A_{1,3} \\
A_{2,1} & A_{2,2} & A_{2,3} \\
A_{3,1} & A_{3,2} & A_{3,3}
\end{bmatrix} \quad B = \begin{bmatrix}
B_{1,1} & B_{1,2} & B_{1,3} \\
B_{2,1} & B_{2,2} & B_{2,3} \\
B_{3,1} & B_{3,2} & B_{3,3}
\end{bmatrix}
\]

For symmetry:

\[
A_{1,1} = A_{3,3} ; A_{1,2} = A_{3,2} ; A_{1,3} = A_{3,1} ; A_{2,1} = A_{2,3} \\
B_{1,1} = B_{3,3} ; B_{1,2} = B_{3,2} ; B_{1,3} = B_{3,1} ; B_{2,1} = B_{2,3}
\]

It can be defined \( w \) as a vector:

\[
w = [A_{1,1} \ A_{1,2} \ A_{1,3} \ A_{2,1} \ A_{2,2} \ B_{1,1} \ B_{1,2} \ B_{1,3} \ B_{2,1} \ B_{2,2} \ I ]
\]  

(7)

Error function shown as:

\[
\mathcal{E}[w] = \sum_{i,j} \sum_{(i,j) \in N} (y'_{i,j}(\infty) - d_{i,j})^2
\]  

(8)

Figure 7 shows dynamic structure of cell. So, this structure is the methodology of the CNN algorithm.

Figure 7-Dynamic Structure of Cell

5. Creating and Testing the CNN

In this research paper, we design a Cellular Neural Network Model to estimate subscriber behaviors that have tendency to rotate another GSM operator. We present them a new promotion campaign to prevent the rotation.

For this purpose, firstly, we get and clear the data from the data warehouse and prepare for the CNN network model. Then, we compute the weights by input and output parameters. And last step, we test the CNN network to other input data to predict the tendency of the subscriber. This cycle applied for every subscriber “call detail records” data to obtain his tendency.

Firstly, input data is selected from the subscribers demographic and gsm line usage fields by the clustering and preprocessing the real data. We can list them as follows; city, marriage, age, child, occupation, gender, home, month_expense, month_income, sub_len, credit_limit, cust_segment, avg_len_call_month, avg_len_call_3month, avg_len_call_4_6month, total_spent, avg_len_sms_month, avg_len_sms_3month, avg_len_sms_4_6month, gsm_line_status, tarrife.

Then, output matrices can be obtained by the average of the each input data. Input values that are exceeded the threshold are evaluated according to this method.

This data set is used to train CNN. Target values are known before and thus this training is supervised training application.

Learning method is selected as Recurrent Perceptron Learning Algorithm (RPLA).

According to the CNN equation, state outputs represented in equation (9).

\[
X_{i,j}(n+1) = \sum_{(k,l) \in N_{i,j}} A_{i,j,k,l} \cdot y_{k,l}(n) + \sum_{(k,l) \in N_{i,j}} B_{i,j,k,l} \cdot u_{k,l}(n) + I_{i,j}
\]  

(9)
In equation 9, A(3x3), B(3x3) and I(1x1) are CNN templates and they are selected as symmetrical for the computation simplicity. U represents the input vector and y shows the target vector in the equation 9.

After then, output values put into the activation function and x state outputs are computed by this way.

\[
y_{i,j}(n) = \frac{1}{2} (|x_{i,j}(n)+1| - |x_{i,j}(n)-1|)
\]

Error function \(\varepsilon[w]\) is calculated by equation 10.

\[
\varepsilon[w] = \sum \sum (y'_{i,j}(x) - d'_{i,j})^2
\]

Graphical plots show the relationships between input and outputs. Most effective fields for the output results are

- Occupation,
- avg_len_call_month,
- avg_len_sms_month,
- avg_len_sms_3month,
- avg_len_sms_4_6_month

In the figure 8, if the density is high, it shows the loyal customer potential. Otherwise, subscriber has tendency to change his service provider.

6. Conclusions

We trained input data and computed the weights of the CNN Networks model. Learning rate was selected as 0.1 and training was started with 500,000 calling detail records. Squared error rate is accepted as 0.01. Gini index for different models on the test data set are listed in below.

<table>
<thead>
<tr>
<th></th>
<th>Logistic regression</th>
<th>Tree</th>
<th>Nearest neighbour</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini index</td>
<td>0.445</td>
<td>0.451</td>
<td>0.512</td>
<td>0.524</td>
</tr>
</tbody>
</table>

The biggest Gini index indicates the best model. In our test results, CNN is better than the other models, except the learning phase.

We focused on to determine hopper customers that have tendency to use another GSM operator and offer them new promotion to keep them in the hand. Thus, our input data also included derived fields also, such as, their calling curves etc. By this way, we analysis calling detail records of the 100,000 subscriber and we found that, %27 subscriber doesn’t use his/her line. %36 percentage of customer has decreasing calling traffic and should present new promotion to them. Other subscribers have variable calling traffic, but their average calling rate is acceptable, thus we needn’t to offer a new thing to them. Actually, classification process gives a lower classification error, but requires longer learning time with CNN model. Thus, this model can be used for only non-online data mining processing issues.

References