

“Sistema de búsqueda en línea de sitios Turísticos dentro de la costa ecuatoriana utilizando un modelo de predicción de minería de datos”

Fabricio Echeverría¹, Silvana Andrade Ramírez², Flor Macías Bustamante³, Iván Ochoa Brito⁴.

¹ Director de Tópico, Magíster en Sistema de Información Gerencial, 2006, Profesor de ESPOL; e-mail: pechever@espol.edu.ec

² Ingeniera en Computación Especialización Sistemas Tecnológicos, 2006; e-mail: sandrade@espol.edu.ec

³ Ingeniera en Computación Especialización Sistemas Tecnológicos, 2006; e-mail: fmacias@fiec.espol.edu.ec

⁴ Ingeniero en Computación Especialización Sistemas Tecnológicos, 2006; e-mail: iochoa@fiec.espol.edu.ec

Resumen

El presente trabajo está basado en el desarrollo de un “Sistema de Búsqueda en Línea de Sitios Turísticos en la costa ecuatoriana utilizando modelos de predicción de minería de datos”, por medio del cual se pretende mejorar la difusión de los atractivos turísticos en la costa ecuatoriana, ayudando a reconocer al turista cuál es el mejor pareja al que puede ir, basado en las actividades y características del lugar que desea visitar.

Un modelo predictivo de minería de datos, usando el algoritmo de Naive Bayes, se logra creando matrices de probabilidades, las cuales sirven para establecer una estimación de probabilidades, con lo que realizamos la predicción. Para construir este modelo se utilizaron las características de los diferentes atractivos turísticos que hay en la costa ecuatoriana.

Palabras Claves: Minería Datos, Modelo Predictivo, Predicción, Naive Bayes, Sitios Turístico, Atractivos Turísticos.

Abstract

The present work is based on the development of a "On-line Search System of Tourist Sites in the Ecuadorian coast using a predictive model of data mining", by means of as it is tried to improve diffusion of the tourist attractive in the ecuadorian coast, helping to recognize the tourist wich is the best place than he can visit based on the activities and characteristic of the place that he wants to visit.

A predictive model of data mining, using the Naive Bayes's algorithm is obtained creating probabilities matrix, which serve us to create an estimation of probabilities, with which we made the prediction. In order to construct this model we used the different characteristic of the tourist attractive in the ecuadorian cost.

1. Introducción

En la actualidad el crecimiento del turismo en la costa ecuatoriana ha generado gran expectativa entre los turistas dentro y fuera del país. El Ecuador es muy afortunado por poseer una diversa gama de atractivos turísticos, los cuales muchas veces no son explotados o conocidos por sus propios habitantes.

Los pocos medios de difusión en línea y la falta de un inventario turístico ha complicado la tarea de dar a conocer la costa ecuatoriana.

En base a estos requerimientos se ha decidido crear un sistema de búsqueda de atractivos turísticos de la costa ecuatoriana basados en el perfil de los visitantes.

Esta modalidad de búsqueda es considerada, porque; en muchas ocasiones, el turista no sabe que lugar contiene las actividades o características que a él le interesa conocer. Esto se va a lograr usando un modelo predictivo de minería de datos.

2. Definición del Problema y aplicación

El problema consiste en determinar cual es la mejor ciudad que se acomode al perfil del turista, lo que nos induce a conocer previamente cuales son las necesidades o requerimientos del turista.

Para conocer las necesidades de los turistas se realizaron encuestas a diferentes personas, las mismas que nos indicaron cuales eran sus preferencias y las actividades que practican cuando realizan turismo.

Con este análisis se noto la relación que existe entre las actividades y los diferentes tipos de atractivos que existen en la costa ecuatoriana. De esa manera encontramos nuestros patrones y generamos la predicción de la mejor ciudad.

Una vez que el usuario ingresa al sistema a través de la dirección web www.turismocosta.com.ec y decide ir al módulo de búsqueda por perfil, el sistema le permitirá ingresar las características del sitio requerido como factores para determinada búsqueda por perfil y aplicar la predicción como son: lugar para comer, parqueo, seguridad, limpieza, y concurrencia. Y además seleccionar las actividades que servirán para predecir el sitio más adecuado dependiendo de las preferencias del usuario y que se han clasificado como: aventura, cultura y ciencia, deportes, deportes extremos, diversión nocturna, ecoturismo, gastronomía, recreación, y otras actividades.

Una vez realizada las operaciones necesarias, el usuario podrá observar la mejor predicción (ciudad) que se perfila a sus necesidades, junto con todos los atractivos que puede encontrar en el sitio. Además de otras predicciones (ciudades) más que se acercan a su perfil.

Por ejemplo, si el usuario desea practicar algún deporte extremo y a la vez visitar un museo, se le indicará cual es la ciudad más apropiada en la que puede quedarse para desde ahí visitar los atractivos turísticos en dónde podrá realizar las actividades seleccionadas.

Finalmente, el sistema proporcionará un ranking de todos los atractivos que hayan sido calificados por los usuarios, donde el usuario podrá informarse de los sitios que tengan mejor calificación.

3. Algoritmo de Predicción y su Aplicación en la búsqueda por perfil.

Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés conocidas como variables objetivos o dependientes, usando otras variables o campos de la base de datos a las que llamamos variables independientes o predictivas.

En el sistema aplicamos un método bayesiano para generar un modelo predictivo que indique cual es el sitio apropiado de preferencia para el turista.

Para el caso de la búsqueda de sitios como se explicó previamente utilizaremos el algoritmo de Naive Bayes [3].

Naive Bayes – predicción del mejor atractivo

Este algoritmo supone tomar atributos independientes como variables, en la primera fase de utilización del algoritmo utilizamos las actividades que se realizan en un atractivo, y como variable clase, el tipo de atractivo. Estos datos se ingresan discretizados, es decir, como una matriz de ceros (si es que no se realiza esa actividad en el atractivo) o 1 (si se realiza la actividad en el atractivo).

Con estos datos se procede a crear la red bayesiana, la cual genera un vector de probabilidades por cada variable (actividad) asociados a los tipos de atractivos. Estos vectores son luego utilizados para la predicción del tipo de atractivo, tomando aquellos con más alta probabilidad según las actividades escogidas

Naive Bayes - predicción mejor ciudad

Para la selección de la mejor ciudad, en la segunda fase se utilizan las características de los atractivos (limpieza, lugar para comer, parqueo, seguridad y concurrencia) como variables para generar el modelo, teniendo como variable clase, la ciudad a la que pertenece el tipo de atractivo que se predijo en el modelo anterior, todas del tipo discretas (contienen datos numéricos tipo entero).

Una vez obtenidos todos los datos del repositorio, se procede a la creación de la red bayesiana, generando un vector de probabilidades por cada una de las características de los atractivos, asociado a las ciudades, para determinar cuál es la ciudad mas apropiada para el turista.

Funciones lineales discriminantes - ranking mejor atractivo

En el ranking se realiza una clasificación de los atractivos según sus calificaciones. Luego de esto se procede a seleccionar el mejor atractivo (con mejores puntajes) dentro de cada grupo.

Se utiliza un método de función discriminante con información a priori¹ para realizar el ranking. Inicialmente se agrupan las calificaciones de cada atractivo y se obtiene el promedio de calificaciones de cada uno, para luego aplicar la función discriminante:

$$y_k = \alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \dots + \alpha_k x_{k,i} + \dots + \alpha_n x_{n,i}$$

Se efectúa la evaluación de las funciones Y_k $\{y_1, y_2, \dots, y_k, \dots, y_n\}$ para el elemento i -ésimo, verificando cuál es la calificación más alta. En base a esto, se determina a que clase pertenece dicho elemento, usando:

$$y_{MAX} = MAX\{y_1, y_2, \dots, y_k, \dots, y_n\}$$

Donde el índice de y_{MAX} , indica la clase a la que pertenece el elemento i -ésimo. Una vez realizado este procedimiento se agrupan todos los x_i (atractivos) pertenecientes a cada clase, seleccionando el mayor para determinar el mejor de cada clase.

4. Resultados Obtenidos

El resultado al final del recorrido de la red bayesiana es un vector de probabilidades de cada tipo de atractivo (la clase para este modelo), del cual se selecciona el que tenga la mayor probabilidad. Este dato sirve como una de las

variables para el algoritmo de predicción de mejor ciudad. El entrenamiento descrito, muestra los patrones de las diferentes actividades que generalmente se realizan en un determinado tipo de atractivo.

Las características de los atractivos y el tipo de atractivo se relacionan con las ciudades, estos son los patrones que resultan al final del modelo.

Con los patrones obtenidos finalmente lo que recibe el usuario es una predicción de la ciudad más apropiada al que puede acudir para desde ahí desplazarse a los sitios turísticos donde pueda realizar las actividades de su preferencia.

Los datos generados por las calificaciones de los usuarios hacia los diferentes atractivos, van a servir para realizar el ranking.

Al evaluar el modelo se obtienen los atractivos con mejores calificaciones de cada una de las tres clases existentes (atractivo, concurrencia y limpieza).

Mediante la aplicación del modelo predictivo se pretendía reconocer los patrones de las actividades que se realizan en un tipo de atractivo específico, lo cual fue conseguido gracias a la utilización del método bayesiano Naive Bayes y a la disposición de los datos [2].

5. Conclusiones

Basados en los resultados y en la experiencia obtenida en la realización de éste sistema podemos concluir lo siguiente:

- 1) El sistema constituye un gran medio de difusión, e información turística a través del cual se puede promocionar la belleza de nuestros parajes, brindándole al usuario la oportunidad de tomar decisiones al realizar turismo dentro del país.
- 2) Gracias al uso de redes bayesianas Naive Bayes, se logró desarrollar el modelo predictivo generando las ciudades y sus atractivos de una manera comprensible y eficiente.

Referencias

a) Referencias de Internet

- [1]<http://www.latino-bi.com/servicios/business-intelligence/datamining.htm>
- [2]www.inpeau.ufsc.br/ivcoloquio/anais/completos/Diego%20D.%20Gregoraz%20-%20Explotaci%F3n%20de%20la.doc
- [3]<http://www.icaen.uiowa.edu/%7Ecomp/Public/Bayes.pdf>

¹A priori: (Del latín, 'lo que viene antes de'), que se tiene información previa al proceso

b) Referencias de libros

- HERNÁNDEZ, J.; RAMÍREZ, M.J. y FERRI, C. Introducción a la Minería de Datos, Prentice Hall, España, primera edición, 2004.
- HAIR, J.F.; ANDERSON, R.E.; TATHAM, R.L. y BLACK, W.C. Análisis multivariante, Prentice Hall, Madrid, 1999
- SILVA, X. Guía de Carreteras, Turismo y Ecoturismo 2006, Dinediciones, Ecuador, 2005.
- JIANG L.; ZHANG, H.; y SU, J. Learning k-nearest neighbor naive bayes for ranking, China, 2005