

Mars Express Power Challenge

CC71Q - Introducción a la Minería de Datos

Departamento de Ciencias de la Computación
Facultad de Ciencias, Física y Matemáticas
UNIVERSIDAD DE CHILE

Gabriel De La Parra

25 de abril de 2016

1. Introducción

El *Mars Express Orbiter* (MEX) es un satélite de exploración científica de la Agencia Espacial Europea, lanzado en junio del 2003 y colocado en la órbita de Marte. Recolecta información científica sobre la ionósfera y la composición de la sub-superficie del planeta rojo.

El MEX está provisto de siete sensores, entre estos, una cámara estéreo que ha permitido capturar imágenes sin precedentes sobre el planeta. Mucha de la información recolectada por el MEX ha sido crucial para las misiones posteriores de exploración de Marte (*Curiosity*, *Opportunity*, etc.). La mayoría de los sensores en la nave deben mantenerse en un rango de 10-20 °C. Adicional a esto tiene dos sensores: el PFS (*Planetary Fourier Spectrometer*) y OMEGA (*Visible and Infra Red Mineralogical Mapping Spectrometer*) que deben mantenerse a -180 °C.

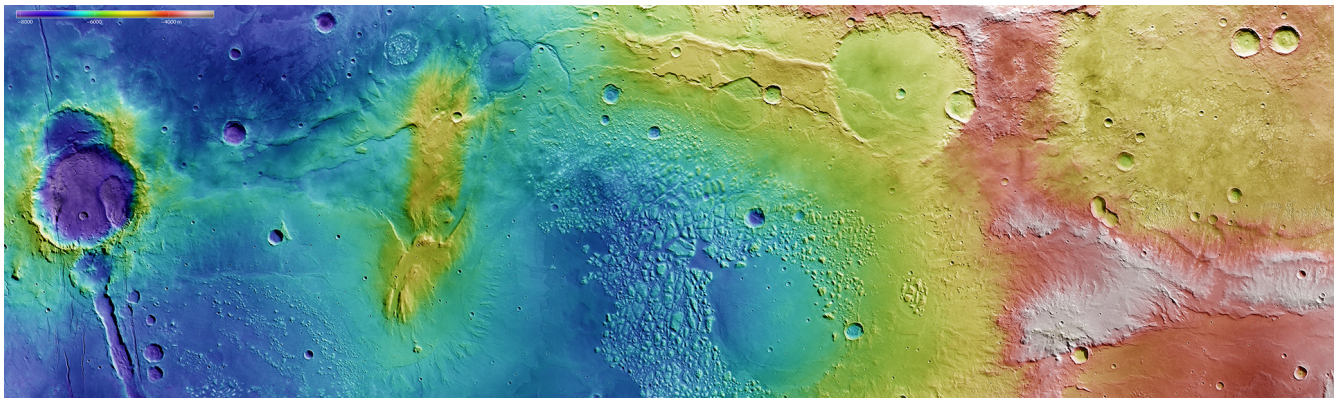


Figura 1. Imágen con cotas del valle *Ancient Atlantis*

El MEX es operado por la ESA desde su centro de operaciones en Darmstadt, Alemania. Desde ahí se analizan los datos recibidos así como también se monitorea cuidadosamente el estado del satélite, con el fin de planificar mejor las misiones de observación y evitar problemas en el suministro eléctrico.

Los operadores del MEX mantienen registros de su consumo de energía termal. La nave utiliza energía eléctrica, proveniente de paneles solares (o baterías, durante eclipses) no solo para alimentar los sensores y cámaras, sino también para mantener la temperatura del satélite dentro de rangos óptimos de funcionamiento. El resto de la energía que no se ocupa se puede derivar para realizar las transmisiones de datos.

$$E_{Telecomunicaciones} = E_{Generada} - E_{Sensores} - E_{Climatizacion}$$

El Mars Express Power Challenge se enfoca en predecir el consumo de energía destinada a la calefacción. El concurso provee datos de tres años (marcianos) como un set de entrenamiento y un cuarto año para pruebas. El éxito del concurso permitirá al MEX entregar información por periodos más largos de tiempo.

2. Descripción de los datos

Los datos para este proyecto son entregados por la Agencial Espacial Europea en formato CSV (2 GBytes) para los últimos cuatro años marcianos. Cada año marciano corresponde a 687 días terrestres.

Los tres primeros años están destinados para entrenamiento de un sistema predictivo y tienen información sobre el consumo eléctrico del sistema termal. El cuarto año no tiene información sobre el consumo eléctrico del sistema termal.

Contextos

Los datos se encuentran separados según el tipo (o contexto) de información que contienen. Así, existen cinco tipos de contextos de datos: Solar Aspect Angles (SAAF), Detailed Mission Operations Plan (DMOP), Flight Dynamics Timeline (FTL), Other Events (EVTF), Long Term Data such as Sun-Mars Distance (LTDATA).

Cada uno de los tipos de contexto tienen diversas categorías (o columnas).

SAAF: Aspectos solares

Los ángulos entregan información sobre la incidencia de los rayos del sol sobre los paneles solares. Los ángulos se miden con respecto a la línea Sol-MEX.

DMOP: Detalles de la planificación de operación

Los archivos DMOP entregan información sobre los comandos que han sido activados en cada sistema. Cada comando incluye el nombre del sistema y el nombre del comando. Debido a la cantidad de comandos, estos no se explican, sin embargo cada uno tiene distintos efectos sobre la temperatura en los distintos subsistemas, y por esto afectando el sistema termal. En estos comandos se encuentran el encendido y apagado del sistema de telecomunicaciones y de los instrumentos científicos.

FTL: Eventos de la trayectoria de la nave

Los eventos listados en estos datos pueden impactar el comportamiento del satélite ya que pueden afectar los ángulos de incidencia, así como también tener relación con el encendido o apagado de algunos instrumentos.

EVTF: Otros eventos

Este set de datos contiene varios eventos. Contiene información sobre eventos del FTL y la complementa con información sobre otros eventos, como por ejemplo eclipses y sus fases (Penumbra y umbra).

LTDATA: Información de periodos extendidos

Incluye información sobre la distancia entre el Sol y Marte, distancia a la Tierra, duración de los eclipses, etc.

Consumo energético

En estos sets se incluye información sobre la corriente medida en los 33 circuitos eléctricos de la nave. Cada medición se realiza cada 30 o 60 segundos.

El objetivo de la competencia es predecir el consumo promedio por hora del satélite, por lo que la información sobre el consumo eléctrico solo está presente en los datos de entrenamiento.

3. Exploración inicial

Como se mencionó anteriormente, el objetivo de la competencia es predecir el consumo energético de los sistemas termicos de la nave a partir de los otros datos que se tienen. La exploración inicial de los datos consistirá en analizar los datos de consumo energético y relacionarlos con los otros datos de estado de los sistemas, incidencia solar y posición del satélite.

Luego de importar los datos en R, lo primero que llama la atención es la cantidad de datos. El conjunto de datos que intuitivamente contiene más información es el de los datos del consumo energético. Éstos se muestrean cada 30 o 60 segundos. Para el primer año existen 1.830.121 registros.

Al graficar este conjunto de datos se puede apreciar que existen niveles de comportamiento. Por la cantidad de datos que existen, el gráfico se aprecia como una mancha continua con secciones de rangos. Debido a esto se procede a graficar en rangos más pequeños.

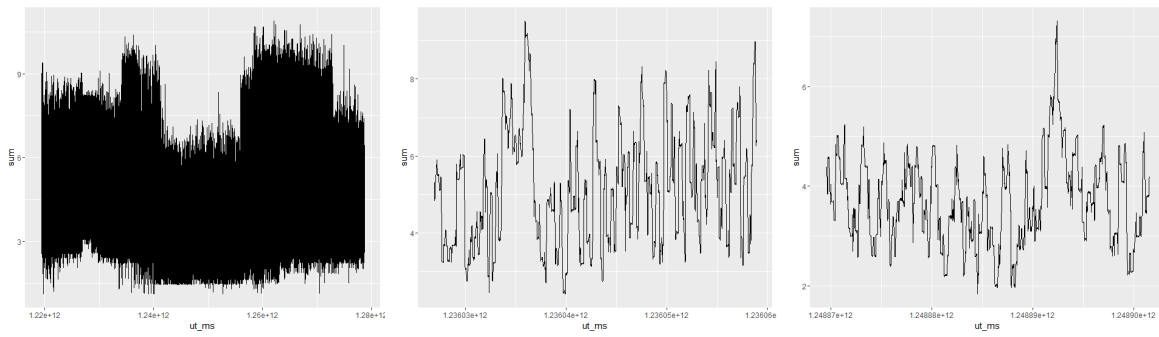


Figura 2. Consumo energético del sistema termal para el primer año
 Figuras 3 y 4. Consumo energético del sistema termal en intervalos menores

Al graficar en rangos de mil datos, se puede apreciar mejor el comportamiento de los datos a lo largo del tiempo. Cabe notar que los datos anteriores corresponden a la suma total de los datos de consumo. Se intuye que al revisar los datos del consumo, se encontrará evidencia de la existencia de clusters de circuitos, es decir, de los 33 circuitos, deben haber varios grupos, divididos entre refrigeración/calefacción y entre los distintos sensores y secciones del satélite.

Una parte importante del trabajo consistirá en encontrar estos clusters. Una vez agrupados los datos se podrá proceder a relacionarlos con comandos enviados remotamente para el encendido y apagado de interfaces y así determinar una relación directa entre estos.

De la misma forma, de la primera observación de los datos, se prevee que pueda existir una relación con la distancia con el sol, similar a las estaciones del año. Esto podría influir en el consumo energético de forma significativa.

Con respecto a los otros set de datos, se considera importante el ángulo de incidencia solar.

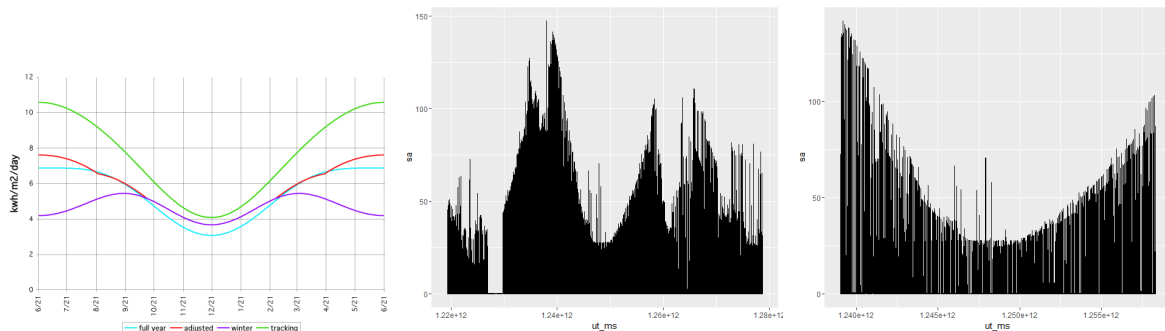


Figura 5. Curvas de incidencia solar en la Tierra

Figura 6. Datos de incidencia solar del satélite para el primer año

Figura 7. Datos de incidencia solar del satélite en un intervalo menor

Al graficar el total de datos de ángulos de incidencia sobre el tiempo, se puede apreciar que el comportamiento es aparentemente similar a una senoide. Al compararlo con gráficas de ángulos de incidencia en la Tierra (figura izquierda), se infiere que el comportamiento debería ser, efectivamente, una senoide.

Se puede apreciar que los registros de incidencia solar presentan una laguna significativa y varias irregularidades. Se pretende buscar alguna relación entre estos datos y los comandos recibidos remotamente. De la misma manera, se espera que exista una relación directa con los distintos eventos como eclipses que puedan haber sucedido.

El resto de los datos corresponden a eventos puntuales, salvo los datos sobre información de vuelo y de distancia con el sol. Estos se utilizarán posteriormente para relacionar los valores de consumo y de incidencia solar.

4. Propuesta de trabajo

Debido al tamaño de los datos, se considera poco viable analizar toda la información en conjunto. Lo anterior principalmente debido a que al intentar visualizar todas las mediciones de los sistemas térmicos, se tuvo problemas

en la capacidad computacional del equipo en donde se realizaba la tarea (i7-4600m Quad-2.9GHz, 8GB RAM).

Se considera entonces la opción de realizar segmentación de la información por rangos según los eventos visibles en cuanto a la recepción de luz. A partir de esta segmentación se propone buscar la superposición entre los datos para verificar si existen relaciones y posteriormente procesar los datos de forma más depurada.

Considerando lo anterior, se propone realizar las siguientes tareas:

- Preparar y segmentar la información de incidencia solar y compararla con la información sobre eclipses, distancias con el sol, dirección de la nave y comandos enviados.
- Preparar la información de consumo energético para realizar clustering. El proposito de esto será encontrar los sistemas que sean de calefacción/refrigeración y así también separar idealmente los 33 circuitos térmicos según los 7 sensores que lleva la nave.

Lo anterior permitirá generar una estimación de relación de los datos. Se entiende que esta estimación puede no ser correcta debido al poco conocimiento específico del dominio del problema. Debido a lo anterior, se propone iterar con secciones de los datos para buscar la validación de las relaciones.

Si bien, se es consciente de la falta de experiencia en el proceso, se pretende que posterior a lo anterior sea posible introducir los datos a WEKA o R y se probará con distintos algoritmos para buscar una clasificación adecuada y construir un modelo que pueda enviado al concurso para su evaluación.

5. Estado del Arte

Tanto la Agencia Espacial Europea como la NASA han lanzado anteriormente varios concursos públicos de investigación. Entre algunos de estos hay algoritmos de optimización, algoritmos de detección planetaria, prototipos de robots de exploración, desarrollo de aplicaciones móviles, etc.

En relación a los ángulos de incidencia solar, existe una amplia bibliografía sobre la incidencia solar sobre la superficie terrestre. Según el Ngram de Google, este tiene su peak alrededor de 1980, por lo que se considera que estos conceptos tienen una madurez suficiente.

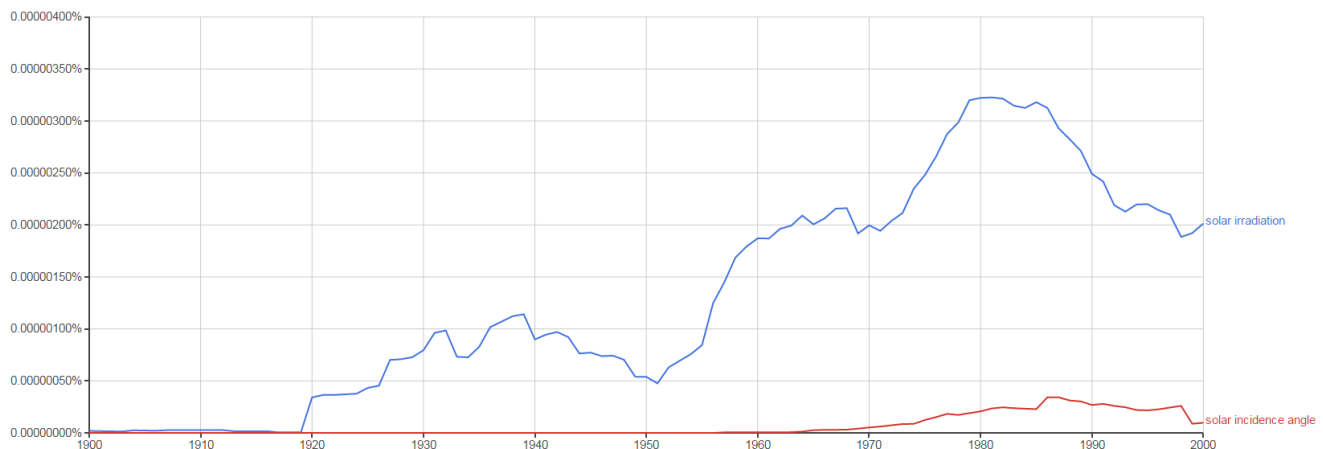


Figura 8. Resultados de búsquedas de incidencia solar e irradiación solar

Para el caso de este proyecto, la incidencia debe ser calculada sobre un satellite, por lo que se prevee se puedan extender los conceptos anteriores.

Con respecto a la información de los sistemas a bordo de la nave, sus comandos y el funcionamiento de los mismos, es poca la información disponible en la página del concurso. Se prevee que el entendimiento de estos sean de poca relevancia, sin embargo se considera que en la medida en que se profundize sobre los datos, se tenga una mejor comprensión sobre los datos anteriores.