

Sistemas Inteligentes

Trabajo 1 Etiquetado – Ubicación geográfica de Usuarios de Plataformas cQA

Profesor: Alejandro Figueroa

alejandro.figueroa@unab.cl

Ayudante: Jean Contreras

j.contrerasleyton@uandresbello.edu

Horario: miércoles 10:20-12:00

Jueves 14:00-15:40

Fecha de Publicación: miércoles 2 de agosto de 2017

Fecha de Entrega: **lunes 28 de agosto de 2017**

Lugar: Horario de clases/ayudantía/con Haydeé Vidal

Aspectos Generales

- El trabajo es individual.
- La entrega del informe impreso debe ser realizada de manera presencial, de acuerdo a lo señalado en el cuadro superior.
- Lea atentamente las indicaciones esbozadas en el syllabus del curso.

Comunidades de Pregunta-Respuesta

Las comunidades de Pregunta-Respuesta (cQA) son sitios en donde, a grandes rasgos, sus miembros responden a preguntas realizadas por otros usuarios. Algunos ejemplos de estas pueden ser StackExchange, Quora o Yahoo! Answers. Los elementos centrales de una cQA, podrían ser identificados como; preguntas, respuestas, categorías, calificaciones hechas a la calidad de las respuestas, etc. Una de las razones que hace importante el uso de estas comunidades es que los usuarios pueden obtener respuestas específicas a sus preguntas, adecuadas a su necesidad puntual que no necesariamente son encontradas fácilmente mediante una búsqueda en Internet. Por ende, este tipo de comunidades son atractivas para una gama amplia de personas. En estas comunidades existen distintas formas de visualizar, agrupar y encontrar cosas en común entre sus miembros. Una de ellas es estudiar patrones de acuerdo a su ubicación geográfica.

Unas de las características de algunos de estos sitios es que sus miembros pueden ingresar una descripción personal. Muchos de estos usuarios proveen antecedentes de su ubicación geográfica o de los lugares a los cuales ha estado arraigado.

2do Semestre 2017 - Sistemas Inteligentes

Para el ejercicio de esta tarea, el estudiante deberá etiquetar manualmente un conjunto de perfiles de miembros cQA. Para ésto, deberá identificar si la descripción no provee ninguna pista de algún lugar geográfico relacionado al usuario (undetermined), si el usuario está relacionado con sólo lugares fuera de USA (non-US), si el usuario está relacionado con lugares dentro y fuera de USA (world), y si está relacionado sólo con lugares dentro de USA (USA only). Además, deberá proveer una lista de los lugares a los cuales está relacionado (los strings del texto). Nótese que el valor de la variable debe estar explícitamente o implícitamente en el texto.

Well, let's see. I'm a physics and biology teacher, and a sincere Christian. I have always thought the best way to honor Christ was to love and cherish the people and the world He made. That means loving the unlovable, among other things. Hmm. .. what else? I love kids, I'm a published author of 3 novels, and I guess I have a goofy sense of humor. I live in Texas , been to Europe a couple of times, and pretty much always manage to have a good time. That's about it, I guess. Peace, y'all. :)	USA only Texas
I LOVE PAINTBALLING!!! 16yrs old washington single in love highschool working B+ student loves seattle and raves!<33	USA only washington
I live in Tennessee. I am so obsessed with the twilight saga it can't be healthy!! I think Robert Pattinson a.k.a Edward Cullen is SOOOO FINE!!! I love the cast of twilight!! I love to talk!! I mainly am on beauty and style or entertainment and a lot of other things!! So fell free to email me also!!	USA only Tennessee
My name is Charlotte, I'm a certified nursing assistant at a skilled nursing and rehabilitation center.	undetermined
Hey the names Kitty Im 17 [nearly 18 woo] and Im from Scotland!! Im currently awaiting going to college and my impending trip around the world - in August! I like photography, boxing and exporing/ learning new cultures ! Im hoping to apply to be a fireman in the near future and get a house ! well thats all for now! Tattie bye !	Non-USA Scotland

2do Semestre 2017 - Sistemas Inteligentes

my name is jessica AKA la chiva. i jus turned 17 which is sooo cool=) i year closer 2 being able 2 watch rated R movies leagally lol. i live in Oklahoma city, OK. im a preety laid back person...i like 2 laugh and have fun...i go 2 DOVE (okc) i own a 1year old lab named shadow which i love very much=) i think datz all *sorry 4 my horrible spelling but datz...*thats* just the way i write=)*	USA only Oklahoma city, Ok.
Born in W. Germany , love to party,travel,meet new people, hang out with friends,play video games like guitar hero 2 on xbox,eat, extrem sports like rock wall climbing, my favorite color is blue,love dolphins, added to ice cream(think goodness whoever invented that),and thats bout it! :)	Non-USA W. Germany
Hey. My name is Tora (Victoria) and I live in Wollongong Australia. I was born in England and moved to Aus when I was eleven. I am seventeen :)	Non-USA Wollongong Australia England
안녕하세요~ So like..Sup, I'm 순규 :> I'm Korean But I love China, especially Hong Kong. I wish I was more interesting.	undetermined
I am currently retired, but I pass my days working as a maintenance officer at a local church. Originally I'm a Florida Cracker but have been a nomad for quite some time. I've traveled throughout the US, the Carribbean and in Europe. Most recently, I lived in Southern Oregon and worked at Crater Lake National Park, then at Diamond Lake Resort. I operated my own business for a time in Oregon. My interests are travel, photography, my Harley-Davidson motorcycle and playing competitive billiards. I love to fish and to cook.	USA only Suthern Oregon
What's up I get on this site when I'm bored as fuck,,, Im 17 and live in Tennessee	USA-only Tennessee
I am a Christian. I love God and support Israel.	Undetermined
I am the Information Systems Specialist for a retail	USA-only

2do Semestre 2017 - Sistemas Inteligentes

company in the State of Maine. We have 5 stores across the state. I graduated from college in 2010 with a BS in Computer Information Systems and am currently working on my MCITP.	Maine
I am a 30 year old female and live in New Zealand in the South Island. Came across the answers section when bored at work one day and am almost hooked! is great to be able to help so many people or to share stories. I am in a happy defacto relationship and currently aiming to travel with him at the end of the year to the USA and UK. Currently working in Administration in retail.	Non-USA New Zealand
South Carolina	undetermined
I lived in UK, NY then Chicago. Previous owner of a Range westminster, Audi S4 and 08 TT. Presently owns an X5 07, Discovery 03 and RR sport SC 09. My passion is all about cars and travelling. Glad to be of help.	World UK NY Chicago

Objetivos

La tarea apunta a confrontar al alumno con la etiquetación de un conjunto de datos, y los problemas que conlleva esta tarea. Además, a revisar conceptos aprendidos en asignaturas anteriores como el de probabilidad. Sumado a esto, se pondrán en práctica conceptos que serán discutidos en clases como el de entropía.

Desarrollo

A cada estudiante se le asignará un número de 2000 usuarios donde deberá identificar su ubicación geográfica de acuerdo a los términos señalados anteriormente. El alumno deberá entregar un informe donde:

- Describe extensamente su experiencia de etiquetado: ¿Qué observó? ¿Qué dificultades encontró? ¿Qué observa de las preguntas que fueron en la clase “undetermined”? Discuta esto en extenso y ejemplificando sus observaciones.
- ¿Qué características tienen los usuarios relacionados con cada ubicación geográfica (las más prominentes)? ¿Cuál es la distribución en su conjunto de datos? ¿Cuál es la probabilidad de cada uno de los valores de variables?

2do Semestre 2017 - Sistemas Inteligentes

- ¿Hay relación entre algún lugar de fuera y dentro de USA?
- Calcule la entropía¹ de sus conjuntos de datos en términos de ubicación geográfica.
- Calcule la Información mutua para los valores de fuera y dentro de USA(world)².

Nótese que debe entregar un informe conforme al esquema descrito en el Syllabus del curso, es decir con introducción, descripción del problema, conclusiones, etc.

Entregable

El conjunto de datos finalizado el proceso de etiquetación. Este debe estar en el servidor.

<https://ba21e628.ngrok.io/SI/index.php>

El no término del etiquetado en los plazos estipulados, implica la nota 1.0. Un informe que desarrolle los puntos anteriormente nombrados. El alumno necesitará este conjunto de datos para tareas posteriores.

Bonus

El alumno podrá adquirir un punto de bonus para la tareas, mediante el etiquetado de una porción adicional de 2000 miembros. Su análisis deberá abarcar los 4000 miembros. El caso de sobrepasar el 7.0, el resto podrá ser cobrado en alguna tarea futura.

Estructura de los Informes

Sino se expresa explícitamente lo contrario, los informes deben contar con una introducción (5%), descripción del problema (10%), conclusiones (15%), presentación y análisis de resultados (60%). Además, la ortografía y redacción cuentan con un 10%.

¹http://gerardnico.com/wiki/data_mining/entropy

²<http://nlp.stanford.edu/IR-book/html/htmledition/mutual-information-1.html>