Introducción Antecedentes Metodología Validación Conclusión Material de apoyo

Desarrollo de una herramienta de visualización de redes sociales mediante un enfoque difuso

Natalia Marín Pérez Tutor: Carlos González Alvarado

Instituto Tecnológico de Costa Rica

Noviembre 27,2017



Tabla de contenidos

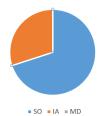
- Introducción
 - Objetivos
- 2 Antecedentes
 - Clustering: fuzzy c-means
 - Análisis de redes sociales
- Metodología
 - Análisis y pre-procesamiento de datos
 - Análisis de distancias
 - Algoritmo difuso
 - Características de la visualizaciíon
- Walidación
 - Visualizacion
- Conclusión
- Material de apoyo



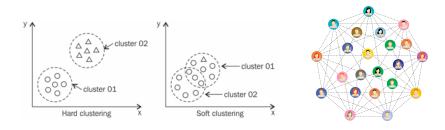
Introducción



Áreas de investigación: Investigador 1



Importancia de la difusidad en redes sociales



Objetivos

Desarrollar una herramienta de visualización de redes sociales utilizando un enfoque difuso. Evaluando los datos de migración en Costa Rica, así como la representación de cómo interactúan las personas con la tecnología.

- Analizar y comparar diferentes distancias con el fin de escoger cuál se adapta mejor al problema
- Desarrollar una metodología que permita el análisis de clústeres difusos en una red social
- Implementar una herramienta que represente, de forma difusa, la conexión entre objetos en diferentes particiones
- Validar la herramienta con diferentes conjuntos de datos para medir su efectividad

Algoritmo difuso: fuzzy c-means

Fórmula fuzzy c-means

$$J_m(U, V) = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ij}^m ||Y_k - v_i||^2$$

$$Y = y_1, y_2, ..., y_n \subset R^n = datos,$$
 $c = numero de clusteres en $Y; 2 \le c < n,$
 $m = exponente en peso; 1 \le m < \infty,$
 $U = particion difusa de $Y; U \in M_{fc},$
 $v = v_1, v_2, ..., v_c = vectores de centros,$
 $v_i = (v_{i1}, v_{i2}, ..., v_{in}) = centro de cluster i$$$

Concepto de redes sociales y difusidad



Figure: a. Fuerza de atracción nodo a nodo b. Fuerza de repulsión nodo a nodo

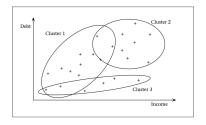
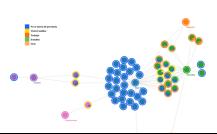


Figure: Ejemplo de representación de cluster difuso

Ejemplo de difusidad en redes sociales

Figure: Movimientos internos en Costa Rica durante una semana-según encuesta realizada en mayo 2017



Movimiento interno

$$t = tiempo$$

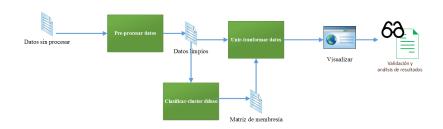
$$i = provincia_i$$

$$j = provincia_i$$

$$t_i + t_j = 7$$

$$porcentajes_membresia = \frac{t_i}{7}y\frac{t_j}{7}$$

Proceso general



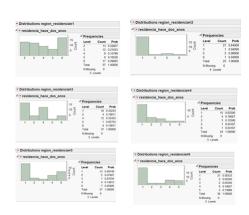
Caso 1: Datos de migración en Costa Rica (INEC)

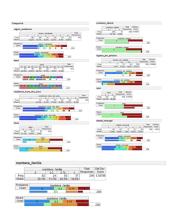


Table: Atributos a analizar

Atributos				
Región de	Condición de			
residencia actual	actividad			
Región residencia	Título			
hace dos años	Titulo			
Mantiene familia	Estado conyugal			
Ingreso por	Ocupación			
persona neto	Ocupación			

Caso 1: Análisis de datos de migración en Costa Rica





Caso 2: Análisis de datos de interacción de personas con la tecnología

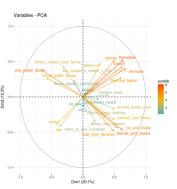


Table: Atributos a analizar

Atributos					
Uso de internet	Tiene teléfono				
(general)	inteligente				
Uso de internet	Tenencia de				
en dispositivo móvil	computadora				
Uso de internet en	Frecuencia en				
el lugar de residencia	redes sociales				
Nivel de conexión					

a internet

Pre-procesamiento de datos

- Limpieza de datos
- Transformación de los datos: normalización, discretización
- Obtención de muestra: solo personas migrantes (interna), mayor de edad.



Distancia Ahmad and Dey: comparación de dos datos categóricos

- Distancia Ahmad and Dey: comparación de dos datos categóricos
- Distancia Euclideana:no aplica para datos categóricos

- Distancia Ahmad and Dey: comparación de dos datos categóricos
- Distancia Euclideana:no aplica para datos categóricos
- Distancia Hamming: recomendada para análisis de tipo binario

- Distancia Ahmad and Dey: comparación de dos datos categóricos
- Distancia Euclideana:no aplica para datos categóricos
- Distancia Hamming: recomendada para análisis de tipo binario
- Distancia Gower: Permite el análisis de variables mixtas: tanto númericas como categóricas

Fórmula Gower

$$S_{ij} = \frac{\sum_{k} w_{ijk} \times S_{ijk}}{\sum_{k} w_{ijk}}$$



Distancia de gower

```
Algorithm 1 Pseudocódigo distancia de gower

 suma_elemento_ij = 0

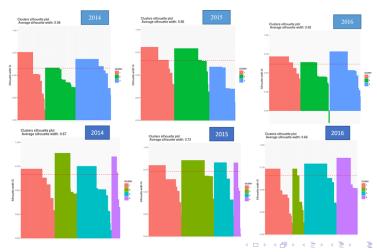
 2: suma_peso_i = 0
 3: for c en columnas do
      objeto_{-}ij = 0.0
      peso_i = 0.0
      if tipo_de_datos[c] es numero then
                            absoluto(vi[c] -
 7.
        objeto_ij
        v_i[c])/(datos\_mixtos[c])
        peso_i = peso[c]
 8:
 9:
      else
        suma\_elemento\_ij = [1,0][vi[c] ==
10:
        vi[c]
11:
        peso_i = (peso[c])
      end if
12:
13:
      suma\_elemento\_ij+=(peso\_ij*objeto_ij)
      suma\_peso\_ij+=peso\_ij
14:
15: end for
```

16: return suma_elemento_ij/suma_peso_ij

Proceso algoritmo fuzzy c-means

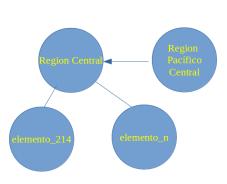


Evaluación del agrupamiento



Estructura de archivo para generar visualizacion

```
"nodes": [{"id": "Region Central",
 "inmigracion total": 71.
 "emigracion total": 99.
 "children": [
     "id": "element 214".
     "titulo": "secundaria completada".
     "ingreso por persona": "medio bajo".
     "condicion laboral" : "Ocupado".
     "membership": [ {
       "percent": 0.06953217343210001
       "percent": 0.88512507829
        "percent": 0.045342748278
  "links": [{
   "source": "Region Pacifico Central",
   "length": 500.
   "target": "Region Central",
   "linkDistance": "strong".
   "preferredCluster": 0.
   "width": 13
```



Comparación: Enfoque difuso(soft) y k-means (hard)

Figure: Comparación de análisis de agrupación utilizando algoritmo k-means con respecto al algoritmo de fuzzy c-means

Nivel socioeconomico medio bajo Nivel socioeconomico bajo Nivel socioeconomico medio-alto Nivel socioeconomico (moderadamente) alto





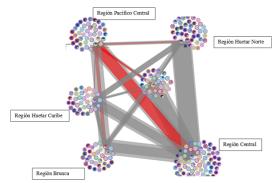




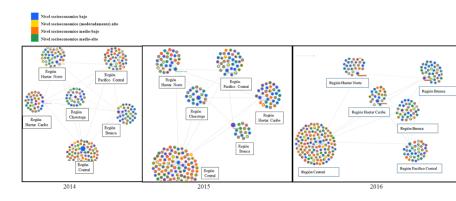


Visualización de enlace a partir de regiones específicas

Figure: Movimientos y densidad de migración a partir de la región Pacífico Central

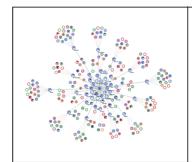


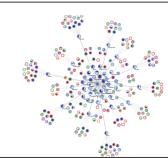
Caso 1: Migración interna en Costa Rica



Caso 2: Interacción de personas con la tecnología en Estados Unidos

Desconectado tecnologico
Tecnologia inmebil (no smartphone)
Inactivo en redes sociales, usuario medio de tecnologia
Usuario medio: promedio uso de varios recursos
Altamente tecnologico
Tecnologico mobil





Conclusión

- Evolución en el tiempo de datos agrupados con un enfoque difuso para una red social
- Comparación del análisis difuso con respecto a uno binario
- Visualización propuesta permite una manera intuitiva de inferir la membresía y relaciones de manera dinámica
- El algoritmo force-directed para agrupar los nodos de acuerdo a su membresía
- Casos analizados: Migraciones internas en Costa Rica y encuesta sobre la interacción de personas con la tecnología.



Introducción Antecedentes Metodología Validación Conclusión Material de apoyo

¡Muchas gracias por su atención!

¿Preguntas?



Introducción Antecedentes Metodología Validación Conclusión Material de apoyo

Material de apoyo

Diapositivas de apoyo



Trabajo futuro

- Probar la metodología con otros estudios como:
 - el movimiento de las personas a través del tiempo influido por las nuevas tecnologías
 - La influencia de las relaciones en la migración de personas tanto de manera interna como externa
- Se puede mejorar la transición de la visualización para que no se note de una manera más natural a través del tiempo

Similitudes en el conjunto de datos

Table: Par de elementos más similares del conjunto de datos Migración Interna en Costa Rica

Título	Condición laboral	Ingreso		Calidad vivienda	Mantiene familia
0.7	0	1	1	0	1
0.7	0	1	1	0	0.75

Disimilitudes en el conjunto de datos

Table: Par de elementos más disimilares del conjunto de datos Migración Interna en Costa Rica

Título Condición laboral	Ingreso			Mantiene	
		conyugal	vivienda	familia	
1	0	1	1	0.9	1
0	0	0.2	6	0	0

Código creación de nodos

```
function drawPieChart(nodeElement, percentages, options) {
  var radius = getOptionOrDefault('radius', options);
  var halfRadius = radius / 2;
  var halfCircumference = 2 * Math.PI * halfRadius;
  var classo = getOptionOrDefault('classo', options);
  var percentToDraw = 0:
  var color_arc =0;
  var range = ["#1A5FFF"."#FFD500"."#FF7100"];
  var cluster = ["cluster1", "cluster2", "cluster3"];
  for (var p in percentages) {
    percentToDraw += percentages[p].percent:
    nodeElement.insert('circle', '#parent-pie + *')
    .attr("r", halfRadius)
    .attr("fill", 'transparent')
    .attr("class", classo)
    .style('stroke', range[color arc])
    .stvle('stroke-width', radius)
    .stvle('stroke-dasharrav'.
           halfCircumference * percentToDraw
           + halfCircumference):
    color_arc++;
```