paper	Nos es de utilidad	Tecnología	Datos de entrada
Twittomender: Rec. usuarios	Profiling de usuarios	Lucene	ts + ers + ers.ts + ees + ees.ts
Prop. analysis: grafos	Conclusiones sobre topología	Ta. grafos	ts + ers + ees
Buzzer: Rec. artículos	Identificar breaking events	Lucene	ts
Influential users; LDA Model	find usuarios influyenyes	TODO	top users + ts + ers + ees
Homophily	7 tipos de relaciones	Teórico	…

1)Twittomender

Paper

Resumen Sistema de recomendación de usuarios basado en contenido y con filtros colaborativos.

Nos es de utilidad… técnicas para profiling de usuarios

Herramienta principal Lucene

Perfil de usuario documento para indexar con Lucene => genera un vector con pesos {word: weight}

Datos de entrada Tweets + ers + ers.tweets + ees + ees.tweets

Puntos interesantes

1. El peso de cada término en el documento para el perfilado de usuarios es el **TF-IDF** propocionado por Lucene: proporcional a la frecuencia de aparición en el perfil del usuario e inversamente propocional a la frecuencia en el resto de perfiles; un peso alto implica que se trata de algo común en el perfil del usuario pero inusual en el resto de la población.

2. Sistema de entrenamiento con casos de prueba de los que se conoce la solución.

    - Tomando como medida la *precisión*: la mejor opción es tomar los perfiles de los ees y la peor opción los tweets de los ees
    - Tomando como medida la *efectividad*: funciona bien los tweets de los ees pero tomando una K muy alta

3. Falla la medida de acierto (sólo se considera buena solución si ya era un folowee)

Cifras concretas

- "Últimos tweets del usuario" => últimos 100 tweets
- Trainig set de usuarios: 19.000
- Conjunto de usuarios para testear el sistema (conocemos su solución): 1.000
- Tiempo de prueba real: 1 mes
- Usuarios de prueba reales: 34
- Resultados reales: 6.9 tasa de acierto

2) Twitter properties analysis

Paper

Resumen Topología de twitter: teoría de grafos. Métricas para twitter

Nos es de utilidad… conclusiones relacionadas con la topología de twitter

Topología Usuarios = Nodos ; Relaciones = Aristas dirigidas (contrarías al flujo de información).

Datos de entrada Tweets + ers + ees

Puntos interesantes

1. Dos términos interesantes: ***Dynamics of the network*** (cambios en la estructura) y ***Dynamics on the network*** (interacción entre nodos y condicionamiento por vecinos)
2. ***Following ratio*** (followers / following)
    - ≃ 0 => spider coleccionando información sobre trending topics
    - < 1 => buscamos colectar información
    - 1 => standard
    - 1 => generador de contendio apreciado por sus propia comunidad
    - 10+ => nodos jefes Huge impact around general media
3. Los bots aparecen y desaparecen según las tendencias temporales.
4. Al llegar a 600 ers, la cifra se dispara a 100.000 ers.
5. **Rapidez en recibir la información** 30% le llegará la información en periodo t (casi instantáneo). La media está entre 0.22 y 0.3 de closeness

Cifras concretas

- Recrea el grafo con más de 14.000 nodos (usuarios)
- 80% usuarios han hecho 1500 tweets. La media es 9Tweets / día / persona
- 25% usaurios tiene 50 ers.
- 50% tiene un rating 1:1

3) Buzzer

Paper

Resumen Sistema de recomendación de artículos por contenido

Nos es de utilidad… Estado del arte de Buzzer; identificar *topical news stories*

Tecnología principal Lucene

RSS Identificar en twitter los breaking events para modificar una RSS

Datos de entrada Últimos tweets generados

Puntos de interés

1. Decripción detallada de la arquitectura y el funcionamiento de buzzer:
    - Dos conjuntos: artículos R y tweets T
    - Cada conjunto se indexa por separado con Lucene: MR y MT
    - Intersección t = (MR X MT)
    - Usamos ti € t como query para sacar el conjunto A de artículos que contienen t.
    - Cada articulo ai € A tiene una puntuación IDF
    - Sumatorio de puntuaciones para cada Aij
    - Resultado: actualizar los artículos de RSS
2. Estado del arte de Buzzer: *Digg.com* ; *Krakatoa Chronicle* ; *News dude*

Cifras concretas

1. Usuarios reales para las pruebas: 10

4) Influential users

Paper

> Nota: sin terminar

resumen técnicas para localizar los usuarios influyentes

Latent Dirichlet Allocation (LDA) Model

Puntos interesantes

1. El 72% de los usuarios cumplen que el 80% de sus ees son debido a la reciprocidad.

Dataset: S = top 1000 twitters S' = S {|s| s.ers + s.ees} |S'| = 6748 T = {tweets} |T| = 1.000.000

S' - no publicadores - robots

En el conjunto S en el que vamos a movernos hay 50.000 relaciones

Topic Distillation

Usa LDA Latent Dirichlet Allocation (LDA) model => Partimos de una bolsa de palabras. each topic is represented as a probabil- ity distribution over a number of words. => Cada documento es un vector-conteo de palabras. each document is represented as a probability distribution over some topics

5) Homophily in social networks

Paper

> Nota: sin terminar

Resumen

Puntos de interés

1. La información en la red se queda concentra en ciertos sub-grafos.
2. Clasificación de las relaciones personales entre usuarios
    - marriage
    - discussing important matters
    - friendship
    - career support at work
    - contact
    - knowing about

    - appearing with them in a public place

3. Clasifciación de *homophily*

    - *status*: race, ethnicity, sex, age, religion, education, occupation & social class.
    - *values, attitudes and beliefs*: internal states presumed to shape our orientation toward futurre behaviour

Cifras concretas

1. Ninguna

> Un dato que no conocía; Twitter creció un 2565% el año de su lanzamiento