



Tema 5

Actividad 1

Autor

Juan José Méndez Torrero

Evaluar articulo cientifico

Introducción

Para esta actividad, se ha decidido evaluar: "[Pattern Mining for Extraction of mentions of Adverse Drug Reactions from User](#)", Azadeh Nikfarjam, Graciela H. Gonzalez. En este artículo se propone un nuevo método para automatizar la extracción de reacciones adversas a los medicamentos (*Adverse Drug Reactions* - ADRs) de usuarios usando técnicas de procesamiento de lenguaje natural que van más allá del emparejamiento léxico. Para ello, los autores aplicaron reglas de minería de patrones para extraer menciones de ADRs en los comentarios de los usuarios sobre los medicamentos que estaban tomando en las redes sociales de salud. Su hipótesis es que, a pesar de que el lenguaje sea usado de una manera muy informal, las personas escriben sus comentarios usando patrones que pueden ser identificados para facilitar la extracción de información interesante sobre su opinión sobre algún medicamento.

Resumen del artículo

En este artículo, la minería de patrones la modelan como si fuera un problema de minería de reglas de asociación, en la que las sentencias son transacciones y las palabras dentro de la sentencia son considerados como ítems dentro de la transacción. Para ello, usaron el algoritmo de patrones sobre un conjunto de datos que contiene un total de 6890 comentarios relacionados con 4 medicamentos: carbamazepina, olanzapina, trazodona y ziprasidona.

Antes de realizar la extracción de patrones frecuentes usando el algoritmo Apriori, los autores realizaron un preprocesamiento de los comentarios de los usuarios para intentar mejorar los resultados. Para ello, realizaron los siguientes pasos:

1. **Preprocesamiento:** Reemplazaron todas las menciones de ADRs en los comentarios por la palabra "ADR". Se eliminaron todos los espacios y signos de exclamación, algunos signos de puntuación menos importantes como las comillas y, los paréntesis.
2. **Tague POS:** Utilizaron el [paseador](#) creado por Stanford en cada una de las sentencias, donde las frases que contenían sólo un nombre fueron también consideradas.
3. **Generación de secuencias:** Se utilizó la parte de la oración como representante de todas las palabras, excepto los conectores (como "y", "pero") y unos pocos verbos seleccionados manualmente, como (hacer, causar, dar). Además, examinamos los lemas de los verbos en lugar de su etiqueta POS, utilizando la raíz del verbo en [Wordnet3](#).

Una vez realizado el procesamiento del conjunto de datos, se aplicó el algoritmo Apriori para la minería de reglas de asociación, centrándose sobre todo en reglas que contuvieran la palabra ADR en el consecuente.

Por último, se generaron los patrones frecuentes convirtiendo las reglas extraídas en posibles patrones, como se puede mostrar en la siguiente figura:

Rule	Possible Sequence	Patterns
PRP and make -> ADR	make PRP ADR and	make PRP (.*) and
	and make PRP ADR	and make PRP (.*)
	make PRP JJ and ADR	make PRP (?:[^]*)* and (.*)
VB RB and -> ADR	VB ADR RB and	VB (.*) RB and
	RB ADR and VB	RB (.*) and VB

Evaluación y resultados

Del conjunto de datos total, se recogieron 1200 comentarios etiquetados, entre los cuáles, sólo los comentarios que contuvieran alguna mención de ADR fueron usados para entrenamiento y los que no tuvieran esa mención, como test. Se ejecutó una validación cruzada 10-folds para evaluar el comportamiento del método. En cada ejecución, se usó el 90% para entrenamiento y el 10% para test.

Las métricas usadas para evaluar el método fueron: *Recall*, *Precisión* y *F-Measure*. Según los resultados obtenidos, los mejores fueron con una configuración de soporte mínimo de 4, un soporte máximo de 6 y un número mínimo de términos por reglas de 4. Estos resultados obtuvieron un 70% de *Precisión*, 66% de *Recall* y 67,96% de *F-Measure*. Además, en cada ejecución de la validación cruzada, el número de reglas frecuentes encontradas fue de 40 y un total de aproximadamente 300 patrones fueron generados en base a esas reglas.

Conclusiones

En mi opinión, este artículo es bastante interesante, ya que nos permite encontrar problemas dentro de comentarios de usuarios sobre algún medicamento y poder analizar esos resultados para poder entender mejor a los pacientes a la hora de realizar el seguimiento en un tratamiento médico.

Este tipo de métodos también podría aplicarse para poder encontrar patrones frecuentes dentro de los comentarios que tuviera algún paciente cuando le da un brote dentro de la enfermedad que tuviera. La información que se pudiera extraer de ahí, podría ser bastante interesante para que el doctor pudiera decidir entre un tipo de tratamiento u otro dependiendo de la opinión del paciente.