

Procesamiento de lenguaje natural aplicado a la evaluacin de personalidad por medio de test proyectivos de frases incompletas

Darwin Omar Alulema Flores
Universitat Oberta de Catalunya
Universidad de las Fuerzas Armadas

A thesis submitted for the degree of
Doctor of Information Technologies and Networks

2015

1. Revisor: David Garcia Solorzano

2. Revisor: Enrique Gaus Termens

Dia de la defensa: 28 / 06 / 2015

Firma del jefe del comite de doctorado:

1

Introduccion

1.1 Introduccion

La investigacion aborda una de las areas de la Inteligencia Artificial, que es el Procesamiento de Lenguaje Natural, ya que se pretende obtener, a partir del analisis del discurso de un paciente, informacion acerca de sus emociones y sentimientos, producto de sus contenidos psicologicos. De esta manera la investigacion aborda una de las tecnicas psicologicas proyectivas, como es el caso de fases incompletas de Sacks, donde el paciente completa oraciones y a partir de la experiencia del terapeuta y el protocolo de evaluacion, se determinan los rasgos generales de personalidad.

Con los resultados, se obtiene rasgos caracteristicos de las personas, sobre como es su comportamiento, lo cual en investigaciones futuras permitirian que los sistemas adapten sus interfaces a cada una de las personas.

1. INTRODUCCION

2

Propuesta

La investigación aborda una de las áreas de la Inteligencia Artificial, que es el Procesamiento de Lenguaje Natural, ya que se pretende obtener, a partir del análisis del discurso de un paciente, información acerca de sus emociones y sentimientos, producto de sus contenidos psicológicos. De esta manera el sistema propuesto aborda una de las técnicas psicológicas proyectivas, como es el caso de fases incompletas de Sacks, donde el paciente completa oraciones y a partir de la experiencia del terapeuta y el protocolo de evaluación, se determinan los rasgos generales de personalidad. Con los resultados, se obtiene rasgos característicos de las personas, sobre como es su comportamiento, lo cual en investigaciones futuras permitirían que los sistemas adapten sus interfaces a cada una de las personas.

2.1 Justificación

Las investigaciones en psicología no han cubierto la posibilidad de realizar estas evaluaciones de forma automática, lo cual se ve reflejado en que, para que un terapeuta realice valoraciones objetivas, necesita un periodo de entrenamiento previo, además se debe considerar que es muy difícil eliminar la subjetividad lo que distorsiona los resultados. Por otro lado las investigaciones hechas con respecto al comportamiento, se han enfocado a factores fisiológicos, para determinar el comportamiento de las personas, dejando de lado los aspectos psicológicos. Es así que la investigación pretende abordar un tema que ha sido escasamente estudiado y que traería como resultados, procesos de valoración más rápidos y objetivos.

2. PROPUESTA

2.2 Estado del Arte

La investigación aborda una de las áreas de la Inteligencia Artificial, que es el Procesamiento de Lenguaje Natural, ya que se pretende obtener, a partir del análisis del discurso de un paciente, información acerca de sus emociones y sentimientos, producto de sus contenidos psicológicos, para determinar la personalidad.

Cabe destacar que aun no existe una gran bibliografía sobre desarrollos y proyectos similares, pero el tema está parcialmente presente en trabajos relacionados con la minería de opinión, análisis de sentimientos, análisis de redes, seguridad digital, lingüística forense, diagnósticos psicológicos, exploración en web, análisis de ontologías y estudios de lenguaje.

2.2.1 Procesamiento de lenguaje natural

La inteligencia artificial está en la intersección de varias disciplinas: la psicología, las matemáticas, la computación, la ingeniería, la biología, la lingüística y otras. Todas ellas, sin excepción, se benefician con los resultados obtenidos; y reciprocamente, los adelantos en inteligencia artificial son resultado del estudio multidisciplinario y la interacción de los respectivos avances (Sonta, 1972)

El Procesamiento de Lenguaje Natural, como una rama de la Inteligencia Artificial pretende determinar el proceso de generación y entendimiento del lenguaje, relacionándose con la Psicología, considerando que involucra factores como el pensamiento, cognición, aprendizaje, personalidad, emociones, entre otros.

Existen en la actualidad dos tipos de líneas de investigación (Sonta, 1972):

- Modelos de sistemas individuales normales y anormales, que consisten en una gran estructura de datos sobre conceptualizaciones del individuo y una secuencia de instrucciones para procesar la información.
- Programas entrevistadores, capaces de conducir diálogos con individuos reales para recopilar datos respecto de las relaciones interpersonales, con la finalidad de separar la subjetividad de los entrevistadores humanos que introducen cierta variación y sesgo.

Algunas de las aplicaciones que utilizan Procesamiento de Lenguaje Natural son:

- Traducción automática: Systran, Google Translate
- Clasificación de texto: Span filtering, topic classification, sentiment analysis
- Recuperación de información: Google search, Lucene, MongoDB, CouchDB
- Extracción de información: Scraping
- Entendimiento y generación de lenguaje natural: Question and answering, chatbots
- Corrección automática
- Compiladores e intérpretes

Las investigaciones de Kenneth Colby, de la Universidad de Stanford, se enmarcan en conseguir que las máquinas aprendan a simular ciertas convicciones y complejos del individuo entrevistado, tal es el caso del programa Modelo de personalidad neurotica, que consistía de un nivel inferior, que almacena ciertos conceptos básicos como el yo, padre, madre, hermano, amor, entre otros; el nivel superior almacena relaciones entre los conceptos mencionados, representando convicciones y creencias; pudiendo ser estas convicciones compatibles, independientes o contradictorias.

Para la ejecución el operador induce al programa a relacionar algunas de las convicciones; en caso de que no existan contradicciones graves, las expresa; pero si existe conflicto una de las convicciones esta se modifica para hacerla compatible con las demás, por un proceso análogo al mecanismo de defensa psicoanalítico. Si después de sucesivos intentos no se consigue establecer la transformación necesaria para compatibilizar las convicciones, aumentará el valor de la variable que representa la ansiedad; que llegado a ciertos valores hace posible que la máquina rehuse continuar pensando en el argumento generador del conflicto. (Sonta, 1972)

Jhon Loehlin, de la Universidad de Texas, desarrolló un programa llamado Aldous que simula la personalidad humana, con capacidades de reconocer la configuración que se le presenta, seleccionar acciones a ejecutar ante la configuración, capacidad de introspección y capacidad de aprendizaje, además posee tres posibles reacciones emocionales, como son el temor, ira y atracción; consiguiendo después de un cierto número de pruebas aprender a asumir actitudes generales o específicas ante los objetos, pudiendo

2. PROPUESTA

reaccionar con rapidez o con vacilacion, de acuerdo con las experiencias precedentes y su estado en el instante considerado. (Sonta, 1972)

En el Institute of Living, en Connecticut, se automatizo la prueba Minnesota Hartford Personality Assay - MHPA, que provee medios sistematicos de obtencion y registro de juicios clinicos, para medir el nivel de adaptacion del paciente y los cambios resultantes de la intervencion terapeutica, el sistema utiliza la tecnica de Q-sort, para la cuantificacion y evaluacion del estado mental del paciente. Ademias se automatizo el Minnesota Multiphasic Personality Inventory MMPI que permite crear un diagnostico mental del paciente a partir de 9 caracteristicas: hipocondria, depresion, histeria, desviaciones psicopaticas, grado de masculinidad y femineidad, paranoia, psicostenia, esquizofrenia e hipomania.

Las investigaciones N. M. Amosov, por su parte abordan el tema de automatias que sienten, determinado 6 tipos de programas como determinadores del comportamiento humano (Sonta, 1972):

- Un programa de sensaciones fisicas, percepcion y procesamiento de informacion referente al sistema fisico
- Programas de percepcion de influencias del contexto y para el reconocimiento de las cualidades del contexto
- Un programa para la accion, es decir, para impartir energia e informacion sobre el contexto
- Un programa de comunicacion
- Un programa para el sistema de la conciencia
- Un programa para la creatividad y descubrimiento

Joseph Weizenbaum, del Massachusetts Institute of Technology, desarrollo Eliza1 que reproducia las habilidades de conversacion de un psicologo, recogiendo patrones de informacion de las frases del usuario y aplicando reglas de concordancia, generando asi respuestas como si de una entrevista se tratara, aparentando empatia con el individuo. Las rutinas de Eliza1 estaban sustentadas en los principios de la psicologia de Carl Rogers, en forma de frases hechas que se utilizaban para que el individuo siguiese conversando.

Plutchik y Kellerman propusieron varios modelos para representar las emociones, que se basan en categorías emocionales, para reconocer alegría, tristeza, ira, etc, el cual fue mejorado en modelo Circumplex, que utiliza una circunferencia con dos ejes que representan características emocionales, como (Garcia y Alias, 2008).

Para la clasificación de textos, Turney y Littman, se basan en la polaridad del texto, en cual para saber la polaridad de cada palabra del texto, calculan la diferencia entre su similitud con un conjunto de palabras positivas y otro de palabras negativas.

(Vybornova, y otros, 2011) se enfoca en la detección de la tensión y la intención en grupos sociales, a partir del análisis de las redes sociales, foros, blogs y comentarios en noticias, a partir de la sintaxis y la semántica de los mensajes, considerando reglas psicológicas, sociológicas y psicolingüísticas de la región rusa, como una herramienta para detectar señales de intenciones de la gente y en particular planes para expresar descontento que podrían desembocar en disturbios sociales.

(Mitchel) propone que cualquier sistema que se deriva de un análisis idiosincrático del texto distorsiona el mensaje y sugiere que para garantizar resultados aceptables, se debe determinar como las personas procesan las oraciones, analizando propias introspecciones e intuiciones, realizando observaciones con expertos lingüistas y pruebas empíricas con gente común.

Para la representación semántica de los textos escritos (Kats, 2014) se basa en la entropía para el procesamiento del lenguaje natural, en aplicaciones de diagnóstico médico en personas con condiciones mentales atípicas, en las cuales el nivel de entropía relativa puede caer por debajo de parámetros normales y la distribución de la entropía a través del texto puede mostrar patrones inusuales lo que contribuye a la evaluación del estado mental de un sujeto, aunque como asegura el autor esta evaluación puede extenderse a evaluaciones políticas o culturales.

El estudio realizado por (Wei, Wu, Lin, y Li, 2014) se enfoca en factores psicológicos como la personalidad y las emociones para reconocer el estilo de interacción en la conversación hablada, destacándose la introversión y extraversión, así como la tristeza y la neutralidad.

Además Deb Roy, ha enfocado sus investigaciones en los robots y la inteligencia artificial para identificar los factores que determinan el desarrollo del habla en las personas, a partir del procesamiento de información de video y audio, con lo que identifica el contexto en el cual se aprende el lenguaje.

2. PROPUESTA

2.2.2 Herramientas de Software

El desarrollo de las herramientas para el Procesamiento de Lenguaje Natural, empezó en la década del 50 del siglo XX, con sistemas de traducción automática basados en diccionarios, hasta la época actual en la que incorporan teorías lingüísticas, modelos cognitivos y enfoques de ingeniería. Las herramientas se pueden clasificar en las siguientes categorías (Manaris, 1994):

- Sistemas de Traducción
- Interfaces de usuario de lenguaje natural
- Procesamiento de texto
- Sistemas de generación de texto
- Sistemas de comprensión de voz
- Sistemas de generación de voz
- Asistentes inteligentes de redacción

El Departamento de Arquitectura y Tecnología de Sistemas Informáticos (DATSI) de la Universidad Politécnica de Madrid (UPM) y del Departamento de Informática de la Universidad Carlos III de Madrid, desarrollaron COES: Herramientas para Procesamiento de Lenguaje Natural en Español el cual es un conjunto de reglas gramaticales españolas y su aplicación para probar distintos tipos de corrección en documentos escritos en Español. (DATSI, s.f.)

El Grupo de interés temático en procesamiento del lenguaje natural con R (GIT-PLN) se enfoca en la investigación de herramientas alternativas para el análisis de textos en español, ya que es un idioma en el que resulta más complicado que en el inglés determinar el lema de las palabras debido a la presencia de desinencias (género, número) y conjugaciones verbales.

Una de las herramientas más utilizadas es NLTK (kit de herramientas de lenguaje natural), el cual está integrado por un conjunto de bibliotecas para el procesamiento del lenguaje natural simbólico y estadístico sobre Python, que incluye demostraciones gráficas y datos de muestra, en lingüística empírica, ciencias cognitivas, inteligencia artificial, recuperación de información, y el aprendizaje de la máquina.

2.2.3 Test Psicologicos

Son instrumentos que permiten evaluar una característica psicológica específica, o los rasgos generales de la personalidad de un individuo, ya que el comportamiento individual que los reactivos de la prueba provocan es comparado al de otros individuos.

La mayor clasificación de los test se da en dos líneas (Retto Vera y Cabana Shapiama, 2010):

- Test Psicométricos: asignan un valor a determinada cualidad o proceso psicológico (inteligencia, memoria, atención, funcionamiento cognitivo, dao cerebral, comprensión verbal, etc.), y se dirigen a actividades de evaluación y selección, como también al diagnóstico clínico
- Test Proyectivos: se basan en la individualidad de la respuesta de cada persona para así presumir rasgos de su personalidad, la actividad proyectiva puede ser gráfica (Bender, DFH Machover, Koppitz, HTP), o narrativa (TAT, Symonds, CAT, Rorschach, Sacks), su uso suele ser más en el ámbito clínico, forense e infantil

2.3 Hipotesis

El reconocimiento de los contenidos psicológicos de una persona proyectados en un discurso, pueden ser analizados por medio del Procesamiento de Lenguaje Natural, empleando el test de Sacks de frases incompletas para determinar rasgos caracterológicos.

2.3.1 Objetivos

2.3.2 Objetivo General

Disear una aplicación que emplee procesamiento de lenguaje natural para determinar los rasgos caracterológicos de una persona por medio de test proyectivo de frases incompletas.

2.3.3 Objetivo específico

- Analizar las herramientas de software para el procesamiento de lenguaje natural en lengua española

2. PROPUESTA

- Determinar el protocolo de evaluacion psicologica del Test de Frases incompletas de Sacks
- Elaborar el repositorio de preguntas del Test de Sacks
- Diseñar un escenario de prueba del sistema

2.4 Metodologia de investigacion

La estrategia de investigacion de Diseo y Creacion, es la mas adecuada para el proyecto ya que se enfoca en el desarrollo de un nuevo producto tecnologico, enmarcandolo en la categoria de metodo porque se pretende establecer un algoritmo para la evaluacion psicologica de la personalidad de forma automatica.

Para el proyecto se pretende implementar un prototipo off line para evaluar los rasgos caracterologicos de la personalidad de un sujeto, para lo cual es necesario una investigacion documental sobre teorias psicologicas que sustentan la validacion a las respuestas.

Ademas es necesario realizar evaluaciones del sistema sobre varios sujetos, los cuales deben ser contrastados con la evaluacion de un Terapeuta, con la finalidad de obtener datos cualitativos y cuantitativos que puedan ser analizados matematicamente dando a conocer el porcentaje de validez del proyecto.

2.5 Plan de Investigacion

Para el desarrollo del proyecto en la Figura 2.1, se detalla el plan de investigacion.

2.5 Plan de Investigacion



Figure 2.1: Plan de investigacion -

2. PROPUESTA

3

Director de tesis

3.1 Propuesta de director

David Garcia-Solorzano recibio sus licenciaturas en Multimedia e Ingenieria Informatica de la Universidad Ramon Llull, Barcelona, Espaa, en 2005. En la misma universidad, obtuvo sus grados de EM en Multimedia e Ingenieria Informatica en 2007 y 2008, respectivamente. En 2013, obtuvo un doctorado de la Universitat Oberta de Catalunya (UOC), Barcelona, Espaa.

3.2 Relacion con la UOC

En la UOC, ha sido profesor e investigador en las herramientas interactivas para el grupo de Aprendizaje en linea (ITOL) desde septiembre de 2008. Sus intereses de investigacion incluyen diferentes temas relacionados con el e-learning: entornos virtuales de aprendizaje, analisis de aprendizaje y autorregulacion.

3. DIRECTOR DE TESIS

4

Bibliografía

- Calzada, J. (2004). La tecnica de las frases incompletas: revision, usos y aplicaciones en procesos de orientacion vocacional.
- DATSI, D. d. (s.f.). COES: Herramientas para Procesamiento de Lenguaje Natural en Español. Obtenido de <http://www.datsi.fi.upm.es/coes/>
- Garcia , D., y Alias, F. (2008). Identificacion de emociones a partir de texto usando desambiguacion semantica. *Procesamiento de Lenguaje Natural*, 75-82.
- Jaen, S. (s.f.). Diseo e implementacion de un sistema para el analisis y categorizacion en Twitter mediante tecnicas de clasificacion automatica de textos. Madrid: Universidad Carlos III de Madrid.
- Kats, Y. (2014). *Semantic Search and NLP - Based Diagnostics*.
- Manaris, B. (1994). *Natural language processing tools and environments: the field in perspective*.
- Mitchel, D. (s.f.). *Natural Leanguage processing - a psychological perspective*.
- Pea, A. (2006). *Lenguaje Natural: Descripcion de las etapas para su tratamiento*. Mexico: Instituto Politecnico Nacional.
- Retto Vera, E., y Cabana Shapiama, P. (2010). Desarrollo de un sistema de informacion aplicado al proceso de orientacion vocacional en zonas rurales del departamento de Lima. Lima: Pontificia Universidad Catolica del Peru.

4. BIBLIOGRAFIA

- Sonta, E. (1972). Temas de Inteligencia Artificial. Buenos Aires: PROLAM.
- Vybornova, O., Sochenkov, I., Kiselyov, A., Tikhomirov, I., Chudova, N., Kuznetsova, Y., y Osipov, G. (2011). Social Tension Detection and Intetion Recognition Using Language Semantic Analysis. European Intelligence and Security Informatics Conference.
- Wei, W.-L., Wu, C.-H., Lin, J., y Li, H. (2014). Exploting Psychological factors for interaction style recognition in spoken conversation. IEEE/ACM transactions on audio, speech, and language processing.