Analysis of patients' diabetes compliance using opinion mining of clinical notes

Adetomiwa Oguntuga, Ky Nguyen, Elmer Bernstam MD, Craig . W. Johnson
The University of Texas School of Biomedical Informatics, Houston TX

Abstract

Introduction

Controlling for age and subjectivity, does race have an impact on a diabetic patient's' compliance with a physician's instructions regarding their diabetic condition. Within the context of this study, compliance is measured in terms of the patient's HbA1C levels and the polarity of the clinical note written by a physician regarding a clinical visit for their condition.

Opinion mining is the process of computationally identifying and categorizing the opinion expressed in text to determine its polarity (positive, negative or neutral). Subjectivity within the context of opinion mining refers to whether the opinion is subjective or objective.

Methods

The researchers carried out opinion mining on clinical notes for 464 diabetes patients, contained in the clinical data warehouse (CDW) at the School of Biomedical Informatics at the University of Texas Health Science Center at Houston. There are over 6000 clinical notes for diabetes patients in the CDW saved in Tagged Image Format File (TIFF), Rich Text Format (RTF) and HyperText Markup Language(HTML). 1134 notes were stored in HTML format, ranging from 2010-2015, the clinicians extracted the most recent clinical note per patient.

The Semantic Knowledge Representation was used to extract diabetes relevant sentences from each note, the rule-based method of TextBlob was used to carry out opinion mining on the extracted sentences and SPSS was used for statistical analysis. This research was deemed "exempt" by the committee for the protection of human subjects (CPHS; the local institutional review board).

Results

ANOVA Result for HbA1C as Dependent variable

Dependent Variable	Independent Variable	Covariate	Sig.
HbA1C	Ethnicity	Age(N/A)	0.000039

ANCOVA Results for Polarity as a Dependent Variable

Dependent Variable	Independent Variable	Sig
Polarity	Subjectivity (Covariate)	0.000001
Polarity	Ethnicity (Independent Variable)	0.012

Conclusion

The final ANOVA performed with HbA1C as the dependent variable and ethnicity as the independent variables showed that ethnicity had a significant impact on HbA1C levels. The final ANCOVA performed with polarity as the dependent variable, ethnicity as the independent variable and subjectivity as a covariate showed that while ethnicity did not significantly impact polarity, subjectivity did.