

Two Approaches to Assertion Classification

Özlem Uzuner^{1,3}, Ph.D., Xiaoran Zhang², Tawanda Sibanda², M.Eng.

¹State University of New York, Albany, NY 12222; ²MIT CSAIL, Cambridge, MA 02139;

³Middle East Technical University, Northern Cyprus Campus, Kalkanli, Guzelyurt, Cyprus

Abstract

We present a study of two approaches to assertion classification: one of these approaches, Extended NegEx, extends the rule-based NegEx algorithm to cover alter-association assertions; the other, SNegEx, is a machine learning approach and explores the contribution of lexical and syntactic context to assertion classification. Both approaches determine whether a problem, as asserted in a patient record, is present, absent, or uncertain in the patient, or associated with someone other than the patient.

We present the two approaches and study their strengths. We show that Extended NegEx is a general algorithm that can be directly applied to new

corpora. However, despite being based on machine learning, SNegEx can achieve similar generality. SNegEx can classify assertions by utilizing the specific syntactic and lexical context of the target, i.e., the word to be classified with an assertion type, in each corpus. Among the features it has been trained with, SNegEx benefits the most from information found in the ± 4 word window of the target. This finding generalizes to both discharge summaries and radiology reports. The specific patterns learned within the ± 4 word window and the rest of the context features of one corpus also generalize from discharge summaries to radiology reports.