Chapter 15

যন্ত্র-না (Jantra-Na: Not-Machine) Can Only Feel যন্ত্রনা (Jantrana: Pain)!

Amitava Das
Norwegian University of Science and Technology, Norway

Björn Gambäck
Norwegian University of Science and Technology, Norway

ABSTRACT

Arguably, the most important difference between machines and humans is that humans have feelings. For several decades researchers have been trying to create methods to simulate sentimentality for machines, and currently Sentiment Analysis is the hottest, most demanding, and rapidly growing task in the language processing field. Sentiment analysis or opinion mining refers to the application of Natural Language Processing, Computational Linguistics, and text analytics to identify and extract sentimental (opinionated, emotional) information in a text. The basic task in sentiment analysis is to classify the polarity of a given text at the document, sentence, or feature/aspect level, that is, to decide whether the expressed sentiment in a document, a sentence, or a feature/aspect is positive (happy), negative (sad), neutral (memorable), and so forth. In this chapter, the authors discuss various challenges and solution strategies for Sentiment Analysis with a particular view to texts in Bangla (Bengali).

1. INTRODUCTION: SENTIMENT ANALYSIS (যন্ত্রানুভূতি)

The title of this chapter is inspired by the Bangla science-fiction writer Narayan Sanyal. One of his most popular Sci-Fi novels is *Nakshatraloker Debatatma* [নক্ষত্রলোকেরদেবতাত্মা] (1976), which was inspired by Sir Arthur C. Clarke's novel *2001: A Space Odyssey* (1968). Sanyal's book first describes the evolution of the human race all the way from primitive creatures to intelligent beings building civilisations and ruling the Earth. The book then takes the history further into the space age, with Jupiter exploration and the same

DOI: 10.4018/978-1-4666-3970-6.ch015

super intelligent computer, "HAL" as in Clark's work. Sanyal called HAL "*Jantra-Na*" (যন্ত্রনা), which in Bangla ambiguously means both 'not a machine'(যন্ত্রনা)and'pain'(যন্ত্রনা),metaphorically portraying the key difference between machines and humans: "The Feelings."

In the late 80s, researchers in Natural Language Processing (NLP) and Artificial Intelligence (AI) started to realize that machines should be able to understand and express sentiment to be intelligent. Since then researchers have attempted textual Sentiment Analysis (SA) for a range of different languages. Sentiment Analysis defines an overall problem, which addresses multiple sub-problems. It is without any doubt a challenging and enigmatic research task. Any scientific research needs to know the proper definitions of its problems in order to solve them. The essential question that is raised at the beginning of the sentiment analysis research is "*What is sentiment or opinion?*" Several researchers have tried to answer this question in the light of a range of research fields, such as Psychology, Philosophy, Psycholinguistics, and Cognitive Science, with many different researchers attempting to give their own definitions, going all the way back to Plato who interpreted opinion as being the medium between Knowledge and Ignorance.

Sentiment analysis research as such started as a content analysis problem in Behavioural Science. The General Inquirer system (Stone et al., 1966) was the first attempt in this direction. The aim was to gain understanding of the psychological forces and perceived demands of the situation that were in effect when a document was written. The system usually counted the occurrences of positive or negative emotion instances in any particular piece of text. The General Inquirer system and work by several researchers from the early 90s onwards (e.g., Wiebe et al., 1990; Hatzivassiloglou and McKeown, 1997; Turney, 2002; Pang and Lee, 2004) are milestones that mark the avenues to the current research trends of today. However, although sentiment analysis research started long ago, the question "*What is sentiment or opinion?*" still remains unanswered. It is very hard to define sentiment or opinion, and to identify the regulating or the controlling factors of sentiment. Moreover, it has not been possible to define a concise set of psychological forces that really affect the writers' sentiments, that is, the human sentiment, broadly speaking. Probably the question cannot be answered by the theories of Computer Science, and maybe the scopes of Medicine, Cognitive Science, Psychology, and other science fields have to be explored. Topically Relevant Opinionated Sentiment detection is better known as Subjectivity Detection (Wiebe et al., 1990). Janyce Wiebe borrowed the definition of opinion from Psycholinguistic research such as Quirk et al. (1985) which states that "an opinion could be defined as a private state that is not open to objective observation or verification."

Sentiment Analysis/Opinion Mining from natural language text is thus both a multifaceted and multidisciplinary AI problem (Liu, 2010). It tries to narrow the communication gap between the highly sentimental human and the sentimentally restricted computers by developing computational systems that can recognize and respond to the sentimental states of the human users. There is a perpetual debate about the best ways of collecting intelligence either by following the functional path of biological human intelligence or by generating new methodologies for completely heterogeneous mechatronic machines and defining a completely new horizon called electronic intelligence. Present research endeavors try to find the optimal solution strategies for machines that either mimic the techniques of self-organized biological human intelligence or can at least simulate the functional similarities of human sentimental intelligence.

Though it might even be impossible to formulate a complete analytical definition of sentiment (Kim and Hovy, 2004), the motivation behind the whole sentiment analysis research field is to