Available online at www.sciencedirect.com

ScienceDirect

Computer Speech & Language 50 (2018) 85–104

www.elsevier.com/locate/csl

Inferring stance in news broadcasts from prosodic-feature configurations[☆,☆☆]

Nigel G. Ward*, Jason C. Carlson, Olac Fuentes

Computer Science, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968, USA

Received 2 June 2017; received in revised form 10 November 2017; accepted 16 December 2017
Available online 24 December 2017

Abstract

Speech conveys many things beyond content, including aspects of appraisal, feeling, and attitude that have not been much studied. In this work, we identify 14 aspects of stance that occur frequently in radio news stories and that could be useful for information retrieval, including indications of subjectivity, immediacy, local relevance, and newness. We observe that newsreaders often mark their stance with prosody. To model this, we treat each news story as a collection of overlapping 6-s patches, each of which may convey one or more aspects of stance by its prosody. The stance of a story is then estimated from the information in its patches. Experiments with English, Mandarin, and Turkish show that this technique enables automatic identification of many aspects of stance in news broadcasts.

Keywords: Information retrieval; Attitude; Broadcast news; Prosody; American English; Mandarin; Turkish

1. Introduction

People use language not only to convey factual information but also additional information such as attitudes, opinions, feelings, judgments, categorizations, and so on. Aspects of this have been studied under many names, including sentiment, attitude, feelings, appraisal, and stance (Read and Carroll, 2012; Rambow and Wiebe, 2015; Chindamo et al., 2012; Biber and Staples, 2014; Elfardy and Diab, 2016; Mohammad et al., 2016). This article will use “stance” as an umbrella term for such information, including all the nuances and subtleties of attitudes and related functions that people display in the course of pursuing various communicative goals.

☆ The first author started this work at Kyoto University. We thank Elizabeth Shriberg, Diego Castan, and Andreas Tsiartas for discussion, for tidying the English-data filenames, and for demonstrating the need to consider more features, and Gerardo Cervantes for discussion. This work was supported in part by DARPA under the Lorelei program, by a Fulbright Award, and by the NSF through REU supplements to IIS-1449093. This work does not necessarily reflect the position of the US Government, and no official endorsement should be inferred.

☆☆ This paper has been recommended for acceptance by R. K. Moore.

* Corresponding author.

E-mail address: nigelward@acm.org (N.G. Ward), jccarlson@miners.utep.edu (J.C. Carlson), ofuentes@utep.edu (O. Fuentes).

URL: http://www.nigelward.com (N.G. Ward).

http://dx.doi.org/10.1016/j.csl.2017.12.007
0885-2308/ 2017 Elsevier Ltd. All rights reserved.

Stance is richly present in many genres of language use, and especially in spoken language. Such information can be useful for information retrieval and filtering, among other purposes. Previous work has explored how, in speech data, stance can be inferred from the speaker's prosody, but so far only for a handful of aspects of stance. Building on this work, this paper examines 14 aspects of audio documents as they occur in English, Mandarin, and Turkish radio news, showing how they can be automatically detected from prosody. The contributions[1] include:

(a) A description of 14 aspects of stance, extending the inventory of computationally modeled aspects of sentiment and attitude (Section 3).
(b) The finding that stance is commonly present in news stories (Section 4).
(c) A model of how prosody conveys stance (Sections 6 and 9).
(d) A demonstration that this model is able to infer stance, in three unrelated languages (Sections 6 and 11).
(e) Evidence that the prosodic expressions of stance include configurations of diverse prosodic features, and illustrations of such configurations (Sections 7 and 10).
(f) The finding that the significance of a prosodic-feature configuration can depend on where it occurs in a story (Section 11).
(g) The finding that the prosodic expressions of stance are extremely language-dependent (Section 12).

2. Motivation and related research

Stance can be useful for information retrieval, information filtering, and information extraction (Larson and Jones, 2012; Lee et al., 2015; Purver et al., 2007; Freedman et al., 2011; Wollmer et al., 2013). Although there has been a fair amount of research on stance, so far it has been mostly limited to expressions of sentiment, positive or negative, with also some work on presence of or strength of opinion (Liu and Zhang, 2012; Freese and Maynard, 1998; Pillet-Shore, 2012; Wilson et al., 2005; Freeman et al., 2015). In this paper, we examine many more aspects of stance.

Although most computational models of stance so far has been for text, there is also some exemplary work on stance in spoken language. For example, Morency et al. (2011) found, in product-review videos, that valenced utterances, in comparison to neutral utterances, have wider pitch range and fewer pauses. Using this information with lexical and video information in a HMM-based model enabled good classification of positive/negative/neutral opinions. Mairesse et al. (2012) similarly found and exploited prosody-sentiment mappings, such as a correlation between low pitch variability and negativity. Levow and colleagues studied stance in dialog and found, among other things, that *yeah* when expressing stronger stances tends to have higher intensity and pitch, and when expressing positive stance tends to be longer, quieter, and higher-pitched (Freeman et al., 2015; Levow and Wright, 2017). In this paper, we study more specifically how prosody expresses stance.

We were inspired to study stance in spoken language by a practical problem: making humanitarian assistance and disaster relief more effective. After a disaster, whether natural or anthropogenic, the international community often mobilizes to help. However, this involves many challenges, including, as described in the Lorelei scenario (DARPA, 2014), enabling a person planning a relief mission to obtain a clear picture of what is needed and what should be done. Today mission planners rely heavily on news broadcasts and social media communications to obtain relevant information. However, the large volume of such data makes this difficult, and there is the need for better tools to help filter and organize such inputs. A further challenge is that disasters can happen anywhere, and the information sources can be in any language, including low-resource languages, for which there may be neither tools (speech recognizers, machine translation systems, etc.) nor adequate resources for building tools. In such cases even imperfect filtering can be of value. Moreover, given the typically large volumes of raw information, mission planners may use information obtained from statistics and tendencies across many items, and for this they may use tools where the categorization of any specific item is only probabilistically accurate.

Stance is relevant to this domain because mission planners frequently need not only actionable information but also big-picture information relating to situational awareness (Verma et al., 2011)—such as indications of where the

[1] Compared to our earlier paper on this topic (Ward et al., 2017), the current paper adds full detail on the methods and experiments, additional observations and findings, and new contributions: e, f, and g.