Sociocognitive Language Processing – Emphasising the Soft Factors

Björn W. Schuller and Michael F. McTear

Abstract Sociocognitive Language Processing (SCLP) is the idea of coping with everyday language, including slang and multi-lingual phrases and cultural aspects, and in particular with irony / sarcasm / humour, and paralinguistic information such as the physical and mental state and traits of the dialogue partner (e.g., affect, age groups, personality dimensions), and social aspects. By that, multimodal aspects such as facial expression, gestures or bodily behaviour should ideally be included in the analysis where possible. At the same time, SCLP can render future dialogue systems more ‘chatty’ by not only feeling natural but being truly emotionally and socially competent, ideally leading to a more symmetrical dialogue. For that, the computer should itself have a ‘need for humour’, an ‘increase of familiarity’, etc., i.e., enabling computers to experience or at least better understand emotions and personality such that they have ‘a feel’ for these concepts. Beyond these ideas, the broader idea of SCLP includes verbal behaviour analysis, a closer coupling between language understanding and generation incorporating social and affective information, and new language resources to meet these ends. By that, SCLP unites expertise from psychology, social sciences, and (natural) language processing. Here, we give a short introduction.
1 Introduction

Sociocognitive Language Processing (SCLP)\(^1\) can be considered as a term to respect ‘soft factors’ in communication. Likewise, it can be seen as a specific kind of and arguably also as an extension to the more traditional and broadly defined field of Natural Language Processing (NLP), the idea of coping with everyday language, including slang and multi-lingual phrases and cultural aspects, and in particular with irony / sarcasm / humour, and paralinguistic information such as the physical and mental state and traits of the dialogue partner (e.g., affect, age groups, personality dimensions), and social aspects. By that, multimodal aspects such as facial expression, gestures or bodily behaviour should ideally be included in the analysis where possible. At the same time, SCLP can render future dialogue systems more ‘chatty’ by not only feeling natural but being truly emotionally and socially competent, ideally leading to a more symmetrical dialogue. For that, the computer should itself have a ‘need for humour’, an ‘increase of familiarity’, etc., i.e., enabling computers to experience or at least better understand emotions and personality such that they understand these concepts. Beyond these ideas, the broader idea of SCLP includes verbal behaviour analysis, a closer coupling between language understanding and generation incorporating social and affective information, and new language resources to meet these ends. By that, SCLP links NLP expertise more closely with such of psychology, the social sciences, and related disciplines.

In this short paper, we will exemplify the principle by focusing on the analysis side of NLP, known as Natural Language Understanding (NLU). Further, we will limit the example to spoken language understanding (SLU). The principles do, however, similarly apply to Natural Language Generation (NLG) in any form.

2 Example: Spoken Language Understanding

In SLU, the text output by the speech-to-text (speech recognition) component is analysed in order to determine its meaning. Meaning can take various forms. In spoken dialogue systems frame-based representations consisting of sets of attribute-value pairs are used to capture the information in an utterance that is relevant to the application [13]. Thus, the representation of an utterance such as book a flight to London on Friday would be something like: flight, destination = London, day = Friday. In other cases, a deeper semantic representation is required that reflects subtle distinctions in meaning that are conveyed by the syntactic form of the utterance [1]. For example, there is only one word that is different in the following sentences (S) but the meaning representations should be different:

- S1: List all employees of the companies who are based in the city centre
- S2: List all employees of the companies that are based in the city centre

\(^1\) As the field of Spoken Language Processing is usually abbreviated as SLP, we suggest SCLP as a short notation for Sociocognitive Language Processing.
The interpretation of S1 asks for a listing of employees who are based in the city centre while the interpretation of S2 asks for a listing of employees who are not necessarily based in the city centre but who work for companies based there. This difference can only be picked up by an analysis that reflects the difference between the use of who and that in such sentences. Up until the 1980s NLP was dominated by the knowledge-based (or symbolic) paradigm in which grammars were handcrafted and meaning was represented using logic-based formalisms. At this time the focus was on parsing written texts. In the 1990s a paradigm shift occurred in which probabilistic and data-driven models that had already been deployed successfully in speech recognition were now applied to language processing [31]. The availability of large corpora of spoken and written text led to an increasing use of machine learning techniques so that the previously handcrafted rules of the knowledge-based paradigm could be learnt automatically from labelled training data. At the same time, attention turned to SLU, for which statistical methods have proved to be more robust as they degrade gracefully with input that is previously unseen and potentially ill-formed. Currently, SLU focuses primarily on the textual form of a message and on its propositional content. However, additional information is conveyed in the prosodic features of a spoken utterance – its phrasing, pitch, loudness, tempo, and rhythm – that can indicate differences in the function of an utterance as well as expressing emotional aspects such as anger or surprise [29]. Other information to support the interpretation of an utterance may come from sensors that provide data about the environmental context, biosensors that report on the user’s physical and emotional state, and machine vision systems that can detect non-verbal accompaniments of speech, such as gestures and facial expressions.

3 The Sociocognitive View

Above (cf. Section 2), we introduced SLU and showed that, in principle, works in this field do consider prosody, and often also multimodal information such as considering the facial expression of speakers. One may thus ask what makes the term SCLP different or justified. In [10], the authors find in two experiments that it seems plausible to consider language understanding as “a special case of social cognition”. This is based on a model to predict an “interaction between the speaker’s knowledge state and the listener’s interpretation” [10]. Similarly, the authors in [8] attest the high relevance of “social, cognitive, situational, and contextual aspects” when dealing with language. Further, the fields of Affective Computing (AC) [21], Social Signal Processing (SSP) [20, 33] or Behavioural Signal Processing (BSP) [16], suggest consideration of affective or emotional and social cues for computing systems used, e.g., in human-computer interaction [17], or human dialogue analysis. In particular in speech analysis [24, 2, 6, 26] and synthesis [14, 15], such information was considered rather early. Further, sub-fields of NLP deal with such information, most noteworthy the discipline of Sentiment Analysis (SA) [18, 32, 11, 5, 28]. Including also acoustic speech feature information, Computational Paralinguistics (CP)
4 Conclusion

In this short contribution, we introduced the term of Sociocognitive Language Processing (SCLP). We further motivated its introduction reviewing key relevant literature in a constructive and synthetic manner with the aim to highlight borders between related existing disciplines and terms such as NLP and SA, and at the same time confine SCLP as compared to broader fields such as AC and SSP.

We exemplified the considerations by language analysis looking at NLU and leading to a sociocognitive view. In the same vein, in NLG, the sociocognitive view lends more weight on the ‘soft factors’ in communication, such as synthesising irony, or nonverbal fillers, and behaviour that fit these and accompany the linguistic content. We believe that, by integrating SCLP principles, one can render future dialogue systems more ‘chatty’ making them not only feel ‘natural’ but truly emotionally and socially competent (cf., e.g., [25, 7]). Ideally, this will lead to a more ‘symmetrical’ dialogue where both ends – humans and computer systems or intelligent machines – will integrate and comprehend soft factors in the communication.

Arguably, for that, communicative technical systems should themselves have a ‘need for humour’ [22, 4, 30], an ‘increase of familiarity’ during repeated or prolonged interactions, etc. In other words, it appears required for genuine SCLP to enable computers to experience and have or at least better understand emotions and personality, such that they have ‘a feel’ for these concepts (cf., e.g., [3]). For example, the degree of conscientiousness of a system might be the decisive factor
between taking S1 in our example in the sense of S1 or rather S2 in addition to its interpretation of ‘what the user is like’ (based also on increased familiarity) in relation to the best interpretation. This will, however, need to be further expanded upon in follow-up considerations and studies including ‘Sociocognitive Dialogue Processing’. Beyond these ideas, the broader idea of SCLP includes verbal behaviour analysis, a closer coupling between language understanding and generation incorporating social and affective information, and new language resources to meet these ends. By that, SCLP unites expertise from psychology and social sciences with NLP on the way to enable genuine conversational dialogue systems [12] or emotionally and socially aware computer-mediated communication [23].

Acknowledgements The first author acknowledges funding from the ERC under grant agreement no. 338164 (iHEARu), and the European Union’s Horizon 2020 Framework Programme under grant agreements nos. 645378 (ARIA-VALUSPA), 644632 (MixedEmotions), and 645094 (SEWA).

References