

Text Mining to Support Foreign Language Learning

フラナガン, ブレンダン ジョン

<https://doi.org/10.15017/1807065>

出版情報 : 九州大学, 2016, 博士 (情報科学), 課程博士
バージョン :
権利関係 : 全文ファイル公表済

氏 名 : フラナガン ブレンダン ジョン

論 文 名 : Text Mining to Support Foreign Language Learning
(外国語学習支援のためのテキストマイニング)

区 分 : 甲

論 文 内 容 の 要 旨

The ability to express intention and emotion are fundamental to human communication, and these skills are usually learnt during early childhood. This can be seen not only in a child using their first words, but also in an adult who is using their limited known vocabulary of a foreign language that they have just started to learn. A fundamental problem in language learning is the difficulty of expressing intention and emotion in a coherent sentence regardless of whether the language being learnt is a foreign or native language. Unintended syntactic errors can alter the intention of an expression, making the meaning of an expression different from the intended meaning, such as the word order error in the following sentence: “Here are *new my* temple pictures” where the error makes the intended meaning of the sentence unclear. Expressing emotion in a foreign language can be a daunting task for learners, and has parallels with other circumstances where a set of special emotion expressions that are not used in everyday communication, such as: the expression of complex aromas, flavors, and many other characteristics used in wine tasting notes. In this work, we investigate text mining techniques to support foreign language learners from the perspective of intention and emotion.

The first half of this thesis focuses on the expression of intention, starting with the automatic classification of 15 writing error categories by SVM analysis of data collected from a language learning SNS. This initial analysis serves as the basis for investigation into the characteristic differences between learner native language groups from the perspective of their writing errors. In particular, we examine the error co-occurrence profiles and the distance between 5 native languages, and how a learners’ native language can be predicted from automatic writing error analysis. We also examine the classification of a learners’ proficiency based on their usage of words. The contribution of these researches is the ability to automatically profile learner characteristics that affect the expression of intention, and predict other characteristics based on common trait association. Also, we address the problem of insufficient sample data, which is often faced when analyzing automatic writing error detection as it is costly and time consuming to create learner writing corpora. To overcome this problem, we propose a method for extracting word order errors, which are a known problem for Japanese learners of English, and automatically compile a balanced corpus from the logs of a language learning SNS.

The second half of this thesis focuses on the expression of complex emotion, starting with the analysis of wine tasting notes by SVM models from 4 sensory viewpoints. These models are then used to extract characteristic words that can be used in emotive sensory expression, and to visualize the characteristics of collections of wine reviews from regions. We then examine the change in the use of emotional expression by analyzing differences in wine tasting notes over a period of 14 years. This research looks in particular at the changes of expression from the perspective of adjective antonym pairs that represent opposites of emotional meaning attributes by visualization. Finally, we examine in detail the complex adjective antonym meaning dimension of THICK and THIN, where the optimal quality is not found at the polar opposites of the dimension. The contribution of these researches is the ability to automatically analyze expressions of emotion, and provide methods to understand complex emotive expressions through visualization.