

報告番号	※甲	第	号
------	----	---	---

主 論 文 の 要 旨

論文題目 Quantifying the mental image of visual concepts
 (視覚的概念がもつ心像性の定量化)

氏 名 KASTNER Marc Aurel

論 文 内 容 の 要 旨

The semantic gap is defined as the lack of coincidence between the information one can extract from data and the interpretation of that same data. It is an yet unsolved issue for content retrieval and multimedia applications, usually describing issues regarding word choice problems and selecting correct retrieval results, and so on. For example, in applications like image tagging, image captioning, or machine translations, it is often difficult to select the best fitting wording out of a group of candidates.

To create a measurement for perceived differences between concepts and such to quantify the semantic gap of different candidates for word choice problems, this thesis proposes the idea of measuring the visual variety of concepts referring to image data. Abstract or vague input words which have a broad mental image due to being less visually defined would result in a broad feature space, while concrete or visually defined input words result in a rather narrow visual feature space. A system is created which regresses a perceived visual variety score for an input word using visual data analysis. The resulting score describes the input word in its visual variety, approximating the perceived abstractness of that word as a number. For this, two approaches are proposed: firstly looking at the relative differences of closely-related words, and secondly as an absolute measurement on a dictionary-level comparison of words.

The first research topic presented in this thesis analyzes the relative visual variety differences of related concepts in a narrow domain by means of a data-driven approach. In this research, existing datasets are reconfigured to create imagesets which reflect the image variety of the real-world. Using the hierarchical relationship of concepts, imagesets for sub-ordinate concepts are aggregated and combined to create imagesets for their composite concepts. As a ratio, a popularity index based on content retrieval engines is used to determine the ratio of sub-concept images. Employing a clustering method on the resulting corpora, the visual feature is quantified to determine a visual variety score for each concept. A crowd-sourced survey is used to decide ground-truth scores for expected visual variety for

different closely-related concepts. Datasets using different popularity methods are compared to baseline corpora to evaluate the performance of the proposed method.

The second research topic presented in this thesis estimates the absolute visual variety by comparing the variety of visual characteristics across imagesets using an algorithm-driven approach. Using this information, imageability scores for arbitrary words on a dictionary-level are estimated by means of a machine learning model. Thus, in this research, the core assumption of using visual image data for human mental image prediction is applied for the concept of imageability. Imageability is a concept originating from Psycholinguistics which aims to provide word ratings on a Lickert scale from unimageable to imageable. A large image corpora crawled from Social Media services is analyzed using a mixture of six low- and high-level visual characteristics. Using the cross-similarity across all visual features, a model is trained to regress an imageability score from an input imageset. The corpora is evaluated using imageability dictionaries from Psycholinguistics as a ground-truth. The evaluations compare the proposed method to existing methods using textual analysis instead of image analysis.

As part of the appendix, two dataset visualization projects are outlined, each loosely connected to one of the two research topics introduced above. In these projects, visual datasets originating from either research topic are compared and analyzed regarding their visual characteristics. These projects complement the ideas from the research topics, looking into future directions and applications of the proposed ideas.

In summary, this thesis presents methods to analyze the mental image of words, targeting a way to quantify the semantic gap between vision and language. Chapter 1 gives an overview on the background of this research from various angles. Chapter 2 reviews existing work in the discussed fields thoroughly, giving a comprehensive analysis of the state-of-the-art in this field. The proposed methods for Research Topics 1 and 2 are discussed in Chapter 3 and Chapter 4, respectively. Chapter 5 compares the results of both approaches, outlining the upsides and downsides of each method for different applications. Lastly, Chapter 6 concludes this thesis by summarizing the research contributions and results found through my doctoral studies.