Preface

Realizing expert-level automatic summarization has been a dream of computing. This work was initiated from the following self-observations when I was traveling in America in October 2012.

The first observation is that our minds often emerge image-like representations when reading texts relevant to experience such as news and novels and we talk and write according to the images in mind. The recall of the images in mind accompanies the reorganization of mental space. This indicates that human mind is likely to have the mechanism that can uniformly process various forms of representation and use one form of representation to interpret the other form of representation. The following question emerges when linking this observation to the problem of summarization: Can we realize a general summarization method through modeling human information processing mechanism?

The second observation is that we can use our own words when making summarization. This indicates that human make summarization with the knowledge for understanding and the knowledge of using language. The following question arises when linking this observation to the problem of summarization: Can we explore the summarization problem from some basic structures and behaviors involved in language use and understanding?

The third observation is that human can make creative summarization with knowledge and reasoning. The long-term experience in the physical space enables human to gain the ability to observe and think through different dimensions. People become more insightful when they can observe and think through more dimensions with deeper structure. The following question emerges when linking this observation to the problem of summarization: Can we explore summarization in a multi-dimensional space to enable summarization system to summarize from different dimensions?

The fourth observation is that human are living in a complex space consisting of cyberspace, physical space and social space, which evolves with human endeavors to develop sciences and technologies. The development of 3D/4D printers and 3D/4D display devices provides a new condition for displaying summary. The following question emerges when linking this observation to the problem of summarization: Can we explore the summarization problem in a cyber-physical-social space rather than in the text space to realize the Summarization of Things?

The major problems of traditional text summarization approaches are too specific to apply to various forms of text, and the rationality of the summarization processes are neglected. I was thinking to develop a rational approach such that summary can be derived from a given text through a reasoning process, based on user requirement and a set of rules. The advantage is that the summarization can suit any requirement and the summarization process is interpretable. However it is hard without formal definition of summarization, and this has been neglected in this research area.

It is critical for summarization research to identify boundaries.

The first research boundary is to develop a computing process that can generate a piece of text from the original text according to a set of predefined rules and constraints. This type of research just needs to concern the rationality of the rules and the efficiency of computing process. The computing result can be for human to read or for application systems to use. It is not necessary to evaluate the automatic summarization by comparing the result with the result of human summarization because the computing process is different from the human thinking process in nature. The result of summarization can be the input of various application systems or for human users to quickly know the interested contents.
The second research boundary is to develop an intelligent summarization system that aims at human-level summarization. Solving this problem requires the intelligent summarization system to have the same level of knowledge and cognitive ability as human experts. It is hard to reach this goal by just improving computing process.

Realizing an intelligent summarization concerns the following fundamental scientific problems:

1. Modeling human-level knowledge, understanding and representation.
2. Multi-dimensional semantic computing, which can zoom-in and zoom-out through multiple dimensions on the representations of different levels according to the rules of different dimensions.
3. The intrinsic relations between representation, semantics, computing and knowledge.

Solving these problems needs to coordinate the knowledge of multiple fields including language, cognition, and psychology, through the insight on the trend of computing. Research within this boundary needs to consider human dimension and other dimensions. Research concerns philosophical thoughts and methodologies but does not discuss whether the system has the same level of intelligence as human or not.

The third research boundary is the philosophy about knowledge, intelligence, dimension, summarization, and cyber-physical society. This book does not focus on research within this boundary.

I hope this book can inspire innovation within this area and beyond and make significant contribution to the transformation of research paradigm at the age of scientific, societal, industrial revolution.

Hai Zhuge
29 May 2016