

Keith Cassell – Research Interests

Overview

My current research concentrates on the intersection of artificial intelligence and software engineering, where I use clustering techniques to help decide how to refactor object-oriented classes. I want to expand my research within this area by more intelligently combining semantic and structural knowledge. I also want to expand my use of clustering to other application domains, including military applications, social networks, and biology. The following sections describe my research history from when I got my Master's degree until the present, followed by anticipated directions of future research.

Research History

I have had an interesting and diverse computer science career, beginning in the early 1980s. In addition to academic research, I have done research for defense companies, and application programming for both commercial and defense companies. This background has given me useful insights into how to balance interesting ideas with useful applications. The rest of this section gives a history of my work, with emphasis on the research areas. My complete work history is available from my CV.

Masters degree

I earned my Masters degree in Computer Science from the University of Texas. My coursework emphasized artificial intelligence. While studying for my Master's, I worked part time at MCC, a research consortium in Austin, where I designed and developed software to analyze the static relationships and run-time behavior of Prolog programs. This work at MCC became the basis for my thesis, "Tools for the analysis of large Prolog programs" [1].

Research in industry

My Prolog background proved useful in getting a research scientist position at a Unisys research lab, designing and coding parts of a computer configuring expert system in Prolog [2]. Later work involved designing and developing a frame-based knowledge representation system in Ada [3]. This frame-based system, in combination with a rule-based system, provided the foundation for a reusability library framework [4]. The underlying idea was to use the frame-based system to organize the software artifacts, and to use the rule-based system to provide the user heuristic knowledge and advice for locating software.

After Unisys, I joined a Lockheed AI research group in Austin on a project for validation and verification of expert systems [5]. I contributed several features to the system, including one that checked for cases of incompleteness in the knowledge base. When that research lab closed, I began working in another part of Lockheed that was using AI techniques for battlefield command and control. Thus began twenty years as a software engineer, typically as a lead engineer for a small programming team. Some of the work was in a commercial setting; some was in defense. Much of the defense work involved the application of AI techniques (clustering, search, blackboard systems, social network analysis) to battlefield command and control systems.

Doctoral degree

I decided to pursue a long-time goal and get my PhD. As a software engineer, I was impressed with how refactoring tools eased my job; however, there were a number of things that they could not do. In particular, existing refactoring tools generally lacked the ability to indicate what needed to be refactored and how. I decided to investigate which clustering techniques were most useful for determining how to refactor object-oriented classes [6–8]. This idea formed the basis for my thesis – “Using clustering techniques to guide refactoring of object-oriented classes”, submitted in December, 2011. My open source research environment is self-created and makes heavy use of other open source software. As such, it can be used as the foundation of additional work into the structure and semantics of object-oriented software.

Future Research

Extensions of current research

In the near future, I want to expand my research by introducing more knowledge and semantics to the representation of software, and I am interested in how to balance the different sources of information.

Military applications

Much of my career in industry was as a software engineer for defense companies. Many of these jobs entailed applying artificial intelligence techniques to solving military problems. In fact, some of the clustering techniques that I used for my thesis work on software engineering, I had previously used for defense applications. In particular, graph-based clustering techniques like betweenness clustering [9] are particularly applicable to analyses of communications and can be used to analyze military communications and social networks, among others.

Possible collaborations

I am also interested in doing some interdisciplinary work in biology. I have a BS in Biology, and I would like to cooperate with researchers in genetics or ecology. Clustering is already heavily used in genetics, and I can see applications of certain clustering algorithms to ecological problems as well. I am also interested in developing research opportunities with defense companies.

References

- [1] K. A. Cassell, "Tools for the analysis of large Prolog programs," Master's thesis, University of Texas - Austin, 1985.
- [2] D. B. Searls and L. M. Norton, "Logic-based configuration with a semantic network," *The Journal of Logic Programming*, vol. 8, no. 1-2, pp. 53-73, January.
- [3] K. Wallnau, J. Solderitsch, M. Simos, R. McDowell, K. Cassell, and D. Campbell, "Construction of Knowledge-Based Components and Applications in Ada," *Intelligent Systems Review*, vol. 1, no. 4, 1989.
- [4] R. C. McDowell and K. A. Cassell, "The RLF Librarian: A Reusability Librarian Based on Cooperating Knowledge-Based Systems," in *Proceedings of the 4th Annual Rome Air Development Center Knowledge-Based Software Assistant Conference*, 1989.
- [5] C. L. Chang, J. B. Combs, and R. A. Stachowitz, "A report on the Expert Systems Validation Associate (EVA)," *Expert Systems with Applications*, vol. 1, no. 3, pp. 217-230, 1990.
- [6] K. Cassell, P. Andreae, and L. Groves, "A Dual Clustering Approach to the Extract Class Refactoring," in *Proceedings of the 23rd International Conference on Software Engineering & Knowledge Engineering*, Miami, FL, 2011, pp. 77-82.
- [7] K. Cassell, P. Andreae, L. Groves, and J. Noble, "Towards automating class-splitting using betweenness clustering," in *24th IEEE/ACM International Conference on Automated Software Engineering*, Auckland, NZ, 2009, pp. 595-599.
- [8] K. Cassell, C. Anslow, L. Groves, and P. Andreae, "Visualizing the Refactoring of Classes via Clustering," in *Proc. Australasian Computer Science Conference (ACSC 2011)*, Perth, Australia, 2011, vol. 113, pp. 63-72.
- [9] M. Girvan and M. E. Newman, "Community structure in social and biological networks.," *Proc Natl Acad Sci U S A*, vol. 99, no. 12, Jun. 2002.