ARGOS: Dynamic Composition of Web Services for Goods Movement Analysis and Planning (Project Highlights 2007)

José Luis Ambite
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
ambite@isi.edu

Genevieve Giuliano, Peter Gordon
University of Southern California
School of Policy, Planning and Development
Lewis Hall 312
Los Angeles, California 90089
{giuliano,gordon}@usc.edu

ABSTRACT

Poster. We highlight the latest developments in the Argos project. First, we describe our approach to automatic workflow composition. Second, we discuss the validation of the freight estimation workflow. Third, we discuss some extensions of the model to residential land value analysis.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*; H.2.5 [**Information Systems**]: Database Management—*Heterogeneous Databases*

Keywords

Web Service Composition, Information Integration, Workflow, Transportation, Real Estate.

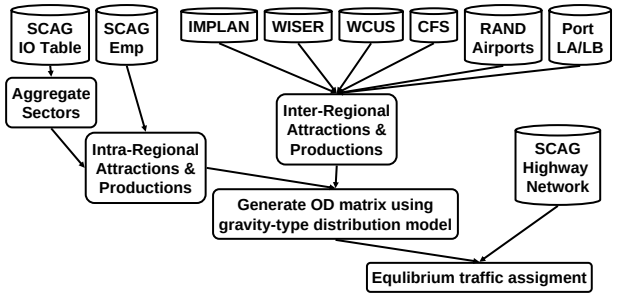
1. INTRODUCTION

The Argos project was funded under the Digital Government (DG) research program. The DG program seeks to focus computer science research on government problems. In this case, the government problem is transportation planning for goods movement. Adequate models for metropolitan freight transportation planning are not yet developed. In this research we explore the potential of web services workflows for developing intra-metropolitan freight flow models that are easily updated and based on widely available data sources. Figure 1(a) shows our abstract workflow for the Los Angeles metropolitan region.

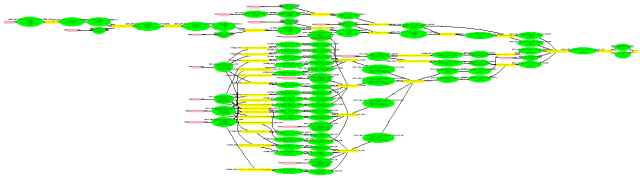
Research under this grant seeks to make significant contributions in computer science and transportation planning, while developing a computational tool that may be adopted in practice. We highlight this year's accomplishments in both computer science and transportation planning.

2. WORKFLOW COMPOSITION

We have developed a general approach to construct data processing workflows, where the data sources and data processing operations are represented as web services. These services consume and produce relational tables, thus are able to represent general computations. We describe the input/output signature of each service as relational formulas in an expressive logic (PowerLoom) using terms from an ontology of the application domain. These logical descriptions allow for a precise understanding of the data and enable the Argos planner to automatically construct a computational workflow in response to a user data request. Figure 1(b) shows the structure of the automatically-generated workflow, which is comprised of 54 services (17 sources and 37 data processing services).



(a) Abstract Data Processing Workflow



(b) Automatically-Generated Workflow

Figure 1: Estimating Commodity Flows in the Los Angeles Metropolitan region (LACMSA)

The Argos planner not only selects the relevant sources and services, but can also automatically insert adaptor services to connect the input and output of existing services. We developed a set of domain-independent adaptor services that correspond to relational algebra operations (selection, projection, join and union), as well as some domain-dependent ones, such as product classification conversions. For example, assume that service S1 requires as input employment data according to the NAICS industry classification, but the only available source S2 for employment data is classified by SIC codes. However, the system knows of a source S3 that contains a conversion table from SIC to NAICS industry codes. Then, the system will automatically insert a conversion service S4 that adapts the data produced by S2 to the data required by S1. Figure 2 illustrates the insertion of an adaptor service.

A more detailed description of the techniques for automatic workflow generation appears as a research paper in this conference [1].

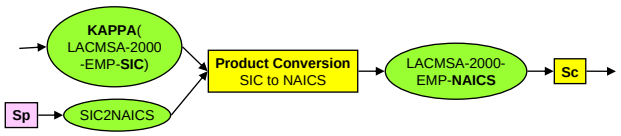


Figure 2: Adaptor Service: Product Conversion

3. MODEL VALIDATION

Our major accomplishment was to use the Argos automated workflow to generate inputs to our intra-metropolitan freight flow model. We applied the model using 2001 data and compared our estimated network traffic volumes to actual screenline counts. Results were good; simple correlation of estimated and actual counts across 18 screenlines is 0.80. A detailed description of the sources and ideas behind the estimation of freight flows for metropolitan highway networks using secondary data sources appears in [2].

Figure 3 shows the calculated truck traffic volumes in a portion of the Los Angeles highway network. We demonstrated our model to the modeling team at Southern California Association of Governments.

Figure 3: Estimated Truck Volume in the Los Angeles Highway Network

4. MODEL EXTENSIONS

A second activity in transportation planning was an analysis of the relationship between residential land values and economic accessibility. Development of the computational workflow generated various measures of economic activity that could be used to develop measures of industry sector accessibility. Urban economic theory posits relationships between land values and accessibility. We developed eight measures of "industry sector accessibility," and two measures of "freight accessibility."

Using multi-level modeling we were able to test hypotheses on the relationship between residential housing prices and access to various economic activity sectors. We find that most of the variation of residential land values in the greater Los Angeles area is accounted for by distance to the coast, a proxy for amenities in terms of weather, air quality and access to coastal recreation opportunities. Accessibility impacts are significant and differ by sector. Heavy manufacturing and resource extraction have a negative effect on housing price, while retail services, finance, health care and entertainment have a positive effect. Results are summarized in Table 1. We conclude that simple measures of job accessibility are inadequate for explaining house price effects.

Table 1: Impact of Accessibility Measures on Residential Prices

Level 1	
Intercept	–
Lot size	+
Dwelling unit size	+
Distance to coast	–
Level 2	
Job factor 1 resource/extraction/manufacturing	n/s
Job factor 2 machinery/heavy manufacturing	–
Job factor 3 retail/services/entertainment	+
Job factor 4 petroleum/mineral extraction	–
Job factor 5 broadcasting	+
Job factor 6 construction	+
Job factor 7 utilities	+
Job factor 8 prof services/education/transportation	–
Freight factor 1 resource extraction/manufacturing	–
Freight factor 2 electronics/equipment	+

Perhaps the most important aspect of the Argos project is that is has demonstrated that collaboration between computer science and social science can lead to significant contributions in both fields, while also contributing to the state of practice in the domain field.

5. ACKNOWLEDGMENTS

We would like to thank the current and former members of the Argos group for their contributions, including our social scientists: Qisheng Pan, LanLan Wang, and JiYoung Park; computer scientists: Dipsy Kapor, Stefan Decker, Mountu Jinwala, Andreas Harth, Naqeeb Abbasi, Matthew Weathers and Karanbir Jassar; as well as our government partner agencies.

This material is based upon work supported by the National Science Foundation under Award No. EIA-0306905. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

6. REFERENCES

[1] Ambite, J. L., and Kapoor, D. Automatic generation of data processing workflows for transportation modeling. In *Proceedings of the 8th Annual International Conference on Digital Government Research (dg.o2007)*, Philadelphia, PA, USA, 2007.

[2] Giuliano, G., Gordon, P., Pan, Q., Park, J., and Wang, L. Estimating freight flows for metropolitan highway networks using secondary data sources. In *Proceedings of Transportation Research Board Commodity Flow Survey Conference*, Transportation Research Circular, E-C088, pp 154–158, July 2005.