Processing Information Intent via Weak Labeling

Anthony Tomasic
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA
tomasic+@cs.cmu.edu

Isaac Simmons
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA
ids@cs.cmu.edu

John Zimmerman
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA
johnz@cs.cmu.edu

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces –*interaction styles, natural language.*

General Terms
Algorithms, Experimentation, Human Factors

Keywords
Information Intent, Weak Labeling, Domestication

1. INTRODUCTION
In [4] we describe the Virtual Information Officer (VIO), a system designed to determine user intent from natural language messages and assist with task completion. VIO assists users with requests that specify a single form filling task and include all the information needed to execute it. There are four main functions that VIO performs in order to assist the user: form ranking, form field analysis, entity resolution and user interface enhancement (Figure 1). Additionally, VIO logs values when the user submits the completed task. These activity logs are used to generate weak labels that drive the learning of the system. Form ranking is handled as classification problem between all possible form targets to allow the system to suggest the most likely forms. Form field analysis is handled as a natural language text extraction problem with one trained model per field that appears on any form. Entity resolution is based on a learning reference resolution model that finds entities most likely to be mentioned in the message. The user interface enhancements take the results of the machine learning components and use them to automatically fill the target form with suggestions. Augmenting existing forms with suggestions can result in a confusing and error prone experience. In [3] we describe the extensive design of a suggestion interface attuned to a mixed-initiative dialog between the user and VIO.

Traditional natural language processing techniques can determine user intent for a well understood domain. However, there are two major barriers to deploying such a system: large amounts of domain specific engineering and hand labeling training data for machine learning components. VIO is explicitly constructed to test the hypothesis that NLP can be successfully used without extensive domain engineering and without hand labeling of corpora. VIO uses meta-data descriptions of the forms along with logs generated by users completing tasks. The system has no domain specific engineering and uses only knowledge that is contained in the meta-data, a dictionary of common names, or in the existing state of the database. The system starts out "untrained" and learns as it observes users completing tasks.

In order to evaluate our hypothesis, we implemented an end to end prototype of VIO. We then measured performance of the individual components of the system. Using a collection of real requests we illustrated that acceptable performance levels arise with a relatively small number of training messages.

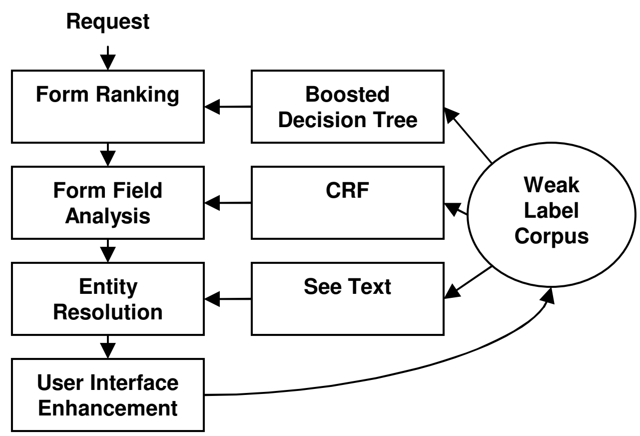

Figure 1: Functional Architecture of VIO

2. EXAMPLE
The system uses the results of the form ranking, form field analysis, and entity resolution steps to generate suggestions and assist the user with filling out the form as quickly as possible. For example, one of the anonymized test messages related to a set of website content update forms requested the following:

> Hi Blake,
> My new title is Research Scientist, 4200 Lorman Park, Rm 390.
> Thanks!
>
> David McCullar
> Atlantic Northwest Trailer, Inc

The form ranking and entity resolution steps attempt to determine what form action is best suited to the task and which record to operate on. The ranked list of form suggestions is embedded directly into the email message with links to the appropriate forms. In our test set, the agent suggested only one form: Modify Person. When the user visits the form, the system uses the results of entity resolution to automatically select the top ranked record. In the example above, clicking on the modify person link brings the user directly to the page to modify David McCullar's record. When successful, these two steps eliminate the search tasks of selecting the correct form and record. In the case of an error, the user is able to make corrections.

CIKM'06, November 5–11, 2006, Arlington, Virginia, USA.
ACM 1-59593-433-2/06/0011.

Form field analysis is used to assist the user when the form is displayed by filling new extracted values into the appropriate fields. In the McCullar example, the annotators successfully extracted his title, first name, last name, and organization. His record already had most of the information, and the system automatically suggested the new title. This leaves the task of updating the street address and room number to the user.

In this example, the system correctly identified both the form and the record thus greatly reducing navigation time. Of the three fields to be updated, one was automatically filled. In the absence of suggestions, the interface works as a traditional form system allowing the user to naturally complete the task.

Once the user finishes and presses commit, the system writes several values to its set of weak labels. It records the form that was used, the record that was updated, and the original and final values of every field on the form.

3. MACHINE LEARNING

Form ranking is driven by a *k*-way boosted decision tree classifier with a simple bag of words feature set to generate the ranked list. One classification model is trained per form and all the confidence weights are ranked and subjected to a threshold.

VIO generates a set of anchored labels from the field information in the weak labels using the domestication algorithm. This algorithm attempts to locate similar strings in the source document and label them. Those labels are used to train conditional random field annotators that generate the extractions for the form field analysis step.

VIO looks up each of the extractions in a dictionary containing existing database values. A naive bayes classifier learns the relative importance of each field and each match contributes to a per-entity score in our reference resolution algorithm. The entities are then sorted and subjected to a threshold to produce the ranked list of suggestions for the entity resolution step.

4. EXPERIMENTAL EVALUATION

In order to evaluate the system on a realistic workload, we acquired a corpus of e-mail from a departmental webmaster. These messages were anonymized and split into a training set of 195 messages and a test set of 39.

The requests in each of the training messages were completed by selecting and completing the appropriate form from a set of website update forms. During this training period, we used an untrained VIO that made no suggestions and simply recorded weak labels. The system was then allowed to train its learned models and we measured its performance in assisting with the remaining messages in the test set.

The mean reciprocal rank of the correct form in the form ranking suggestions was 0.94. It was the first or second ranked suggestion for all but one of the test messages.

Label domestication was evaluated using hand generated labels as the standard. The algorithm had fairly uniform performance with an average F1 value of 0.95. Repeated values and embedded formatting requirements of the form caused the majority of errors.

Extraction performance for each field corresponded strongly to the number of domesticated labels in the training set. F1 performance values were around 0.86 for common fields with mixed results for the less common ones. For all recall values, the extractors had high precision. As such, the system rarely annoys users with incorrect suggestions. Extractions were also used to generate the suggestions for entity resolution and the mean reciprocal rank of the correct item in that list was 0.85.

The system reliably suggests the correct form as well as picking top ranking entity instance for modification. Extractions perform well for frequently used fields, but poorly on long values and uncommonly used fields. Since form field analysis performance is related to the frequency that a field appears, we get our best performance where automation offers the largest time savings.

5. CONCLUSION

In this paper we proposed and analyzed a system that determines information intent of a user and assists with task completing by using machine learning algorithms trained on weak labels. Weak labels are generated by silent observation of the user's interaction with the system. Our system tests the hypothesis that a practical NLP system can be built without extensive domain engineering or hand-labeling of corpora.

To better understand the performance, advantages and disadvantages of our solution, we built an end-to-end prototype, the Virtual Information Officer and ran tests to determine its performance. Our experiments showed that our algorithms based on weak labels performed well [4].

VIO solves three problems: form ranking, form field analysis, and entity resolution. Our analysis shows that information intent can be successfully analyzed by solving these problems using only weak labels for training. In [3] we have shown that the system decreases the amount of time needed to complete a task with no significant impact on error rate in user studies. The domain independent nature of VIO implies a broader applicability – users can be assisted in completing information intent requests that translate into form completion tasks.

6. ACKNOWLEDGEMENTS

Thanks to Robert McGuire and Kelvin Goh for their work on the VIO. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHC030029.

7. REFERENCES

[1] William W. Cohen, Einat Minkov, Anthony Tomasic, *Learning to Understand Web Site Update Requests*, in IJCAI, 2005, pp 1028-1033.

[2] Anthony Tomasic, John Zimmerman, Isaac Simmons, *Linking Messages and Form Requests*, in Intelligent User Interfaces (IUI), 2006, pp 78-85.

[3] John Zimmerman, et. al., *Mixed-Initiative Form Completion for Information Intent*, in preparation, 2006.

[4] Anthony Tomasic, Isaac Simmons, John Zimmerman, *Experimental Evaluation of Processing Information Intent via Weak Labeling*. Technical Report, Department of Computer Science, Carnegie Mellon University, pending.