A Recommendation Engine for Subsistence Farmers

Daniel P. Stormont and Samantha L. Snabes
The Pineapple Project, Root2Market, Mountain View, U.S.A.

Keywords: Expert Systems, Planning and Scheduling.

Abstract: Many of the world's agricultural workers, especially in the equatorial regions, are subsistence farmers. This paper describes the design of a Recommendation Engine that is the core of a crowd-sourced tool to assist subsistence farmers in better utilizing their land. This tool is being developed by a worldwide team of developers through hackathons and volunteer work. The paper provides an introduction to the project, an overview of the system design, current and future designs planned for the Recommendation Engine, and the next steps for the project.

1 INTRODUCTION

According to the World Bank, 75% of the world's poor live in rural areas where the majority of them are subsistence farmers (WorldBank, 2007). Subsistence farming is defined as "a system of farming that provides all or almost all the goods required by the farm family usually without any significant surplus for sale" (Dictionary, 2012). As also noted in (WorldBank, 2007), one of the "pathways out of poverty" in the World Bank agriculture-for-development agenda is to "improve livelihoods in subsistence agriculture and low-skill rural occupations." An essential element of improving livelihoods is providing farmers with better information: about crops, growing conditions, pests, growing tips, produce handling suggestions, and market information. Providing that information is the objective of the Pineapple Project.

This project was incubated in a number of global hackathons focused on using technology to assist those in need, especially in developing nations. A multi-national team was formed and key members of the team have continued on with the project to provide continuity between hackathons. Presently, a number of partnerships have been formed with other ongoing research projects to be able to make continuous progress, although there is still an intent to utilize the energy and new ideas resulting from participating in hackathons in the future. The project was launched at the National Aeronautics and Space Administration (NASA)-sponsored International Space Apps Challenge (ISAC) in April 2012 (ISAC, 2012). Since its launch, the project has been advanced in two Random Hacks of Kindness (RHoK) global events (RHoK, 2012) and at the United States Agency for International Development (USAID) Hacking4Hunger event (USAID, 2012). The Pineapple Project was an award winner in all four hackathons.

The remainder of this paper will present the design of the system, as it is currently envisioned, and then focus on the design of the core of the system: the Recommendation Engine. This engine combines geographic location, growing conditions data, and crop data to assist farmers with planting decisions. The paper will conclude with a description of the next steps for the project.

2 SYSTEM DESIGN

The crop recommendation system is being built in phases by team members around the world, so the design is distributed, extensible, and highly modular. Figure 1 shows a block diagram of the system. There are a number of services and data sources that interact with the Recommendation Engine – the common component in the system.

2.1 Crop Database

The crop database is presently limited to a small subset of tropical fruits common to the equatorial region of the world. The ideal growing condition parameters for these fruits are stored in the database, with the type of fruit as the key value. This data is sufficient for the planned pilot testing of the system, but it will need to

P. Stormont D. and L. Snabes S..
A Recommendation Engine for Subsistence Farmers.
DOI: 10.5220/0004331805780581
In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence* (ICAART-2013), pages 578-581
ISBN: 978-989-8565-39-6

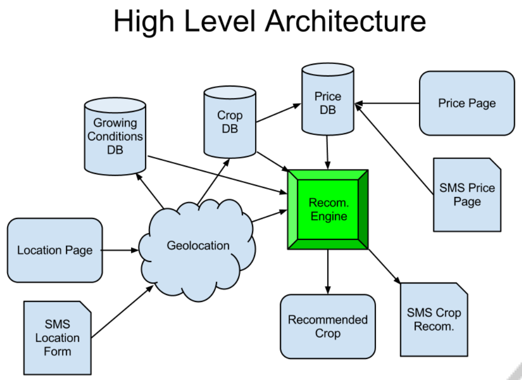

Figure 1: The crop recommendation system block diagram.

be expanded for future versions of the system, either manually or by incorporating additional data sources.

2.2 Growing Conditions Database

The growing conditions database contains historical climatological data, indexed to a grid associated with latitudes and longitudes on Earth. For the pilot testing, the system is making use of the Prediction Of Worldwide Energy Resource (POWER) data set provided by NASA (Hoell et al., 2007). This data set provides daily integrated surface radiation, daily averaged dew point temperature, daily maximum and minimum temperatures, and daily precipitation on a one-degree latitude/longitude grid. The database also includes soil conditions data from the Food and Agriculture Organization of the United Nations (FAO, 2012). All of the data is stored in a NoSQL database for fast retrieval and because of the sparse nature of the data sets (Strauch, 2012). (Much of the data is omitted since the only data of interest to the project is the data that covers land masses.)

In future iterations of the system, the growing conditions data will be expanded with other sources–including crowd-sourced data–and additional parameters will be added, such as polyculture, diseases, and pests.

2.3 Price Database

This database is not part of the pilot project, which is focused on growing conditions only. There are a number of projects that are looking at the issue of crowd-sourced pricing information and producer-driven markets at the national and regional levels, such as the OpenMarkets Foundation (OpenMarketsFoundation, 2011), but none of them are to a point that their systems can be integrated with the crop recommendation system. Future iterations of the system will consider methods for providing price and demand data as a source to the Recommendation Engine.

2.4 Geolocation

Geolocation is critical for filtering the data provided to the Recommendation Engine. The geolocation in the current system will take advantage of the Global Positioning Satellite (GPS) data provided by a device that is so equipped or by doing a lookup on a location provided through the interface. If the user knows the latitude and longitude of the location they want a recommendation for, that can be entered into the system. Otherwise a lookup of a provided location, such as a municipality, will be used to query a geocoding service. At present, the system is using the Google Geocoding API (Google, 2012), but this will need to be supplemented with other services for areas that are not well represented in the Google data.

2.5 User Interface

Three user interfaces have been prototyped during the hackathons: a web interface (actually, several web interfaces), a smart phone app, and a Short Message Service (SMS) interface for standard cellphones.

2.5.1 Web Interface

The web interface is the one used most by the collaborating team members in developing the crop recommendation system. It provides facilities for textual, graphical, and pointing device interaction and displays; thus it is more visually appealing and it is easier to assess the quality of the data and the results. However, for the pilot program, it is not likely that a subsistence farmer in a rural area of a developing nation will have access to a computer and a high-speed data connection to utilize this interface.

2.5.2 Smart Phone

Some of the international participants in the ISAC prototyped an iPhone app for accessing basic crop data. It is functional and could be modified to access a mobile version of the website, but smart phone adoption has not reached the level of penetration in developing nations that would allow the project to focus on this as a primary interface for the system...yet. Smart phone prices are coming down and adoption is accelerating, even in less affluent areas of the world, so this will likely be a secondary option for accessing the system.

2.5.3 Cellphone Text

Simple cellphones (or “flip phones”) are ubiquitous throughout the world and even the poorest nations typically have at least one cellphone provider for voice and SMS services. Thus, the project has focused on the SMS interface as the primary interface for the pilot test of the system. The SMS interface is simplistic, but the system can still gather enough information in one or two text messages to make a recommendation. All of the processing is done on the server side by the Recommendation Engine through its interfaces to other services and data sources.

3 RECOMMENDATION ENGINE DESIGN

Central to everything that the crop recommendation system does is the Recommendation Engine. This engine is responsible for taking the inputs from the user interface and the geolocation/geocoding service and filtering the data in the database to come up with a recommendation or recommendations to the farmer that will be appropriate to the location of the planting and the time of the year.

3.1 Interface Design

The distributed nature of the teams working on this project and the desire to keep the system modular and extensible to allow for replacing elements of the system when better data sources or sources or services that are more appropriate to a given location are identified drives the need for well defined interfaces between all of the system components. The interfaces are defined with APIs for services (at present these are usually APIs defined by external organizations, such as the Google Geocoding API (Google, 2012)) and with a JavaScript Object Notation (JSON) format (JSON, 2012) for data interfaces. An example of a JSON data interface definition for the elevation associated with a geolocation is as follows:

```
{  "Elevation" :
   {
      "MinEl" : null,
      "AvgEl" : 456,
      "MaxEl" : null
   }
}
```

JSON has the advantage of being readily understood by humans reviewing the data and also being fairly easy to parse and to convert from other formats. The recommendation system uses Python scripts on the server extensively for data conversion and parsing. Converting JSON to Python is a simple direct conversion from JavaScript objects and arrays to Python dictionaries and lists. Conversion from extensible Modeling Language (XML) to JSON is also straightforward with Python libraries.

All of the interfaces used by the Recommendation Engine at present are defined in a JavaScript Interface Format specification that has been made available to the developers working on the project.

3.2 Current RE Design

The current design of the Recommendation Engine makes use of logic programming to determine the crop or crops that meet the growing criteria. This is a method expert systems have used since the early years of expert systems programmed in Prolog. In this case, the expert system is programmed in server-side Python, but the logic is the same.

To begin the process, the location of the area to be planted is determined either through the geolocation service on the device, a geocoding service, or by direct entry of a latitude and longitude. This location is used as a hash into the growing conditions database. The retrieved growing conditions contain the temperature, sunlight, and precipitation parameters for the full year. The current date or a user-entered planting date is used to limit the growing conditions for planting to the data for a single month.

With the growing conditions data retrieved, the expert system in the Recommendation Engine selects crops from the crop database that satisfy the growing conditions. For example, the Recommendation Engine will determine if either of the temperature extremes for the month (minimum or maximum temperature) would violate the temperature range for a crop in the database. If so, that crop is removed from consideration. The same is done for the sunlight and precipitation data. The crops that meet the selection criteria form the initial recommendation. If the initial recommendation list is empty, the user is informed that no crops meet the criteria for that area and time of year and the process terminates. Otherwise, the growth rate of the crops in the recommendation list is used to retrieve additional months of data from the growing conditions data to make sure none of the constraints on the selected crops are violated at some point during the growth period. The crop recommendations that survive this second check are presented to the user.

The advantage of the logical programming approach is that the decision process can be logged and used for troubleshooting program and/or data errors.

It can also serve an educational purpose by providing an answer to the question of why a particular crop in the crop database was not selected by the Recommendation Engine. However, while the logic programming approach is feasible for the small size of the crop database and the limited parameters being considered for the pilot testing, it will not scale well to some of the desired future capabilities for the system. That will require a different approach.

3.3 Proposed Future Design

To truly be a useful tool for subsistence farmers, the crop recommendation system needs to have more than just a small subset of tropical fruits in it and it needs to take many more factors into consideration. In the crop database, any crop that could be grown in a region should be considered by the Recommendation Engine. The price database also needs to be implemented and incorporated into the recommendation. The farmer should be able to get an estimate of the market value of the crop when it is harvested, as well as the best place to take the crop to market. The growing conditions database needs to be expanded to include parameters like polyculture (simultaneously planting multiple crops) or crop rotation when making a recommendation. And tips on dealing with pests and diseases, as well as methods for harvesting, storing, and transporting the produce to market should be made available.

These enhancements to the functioning of the Recommendation Engine will require a redesign, since logical programming doesn't scale well when multiple factors have to be examined. The planned approach to the redesign will be to replace the expert system in the Recommendation Engine with a neural network that takes all of the growing conditions, pest, crop, and price data as input and outputs crop recommendations and tips to the grower. The disadvantage is that the decision process is no longer observable; but the neural network should be faster and more flexible.

4 NEXT STEPS

The immediate objective for the project team is to get an SMS-based version of the system into the field for testing with subsistence farmers and Non-Governmental Organization (NGO) personnel on the ground in one of the equatorial regions.

In the longer term, it will be necessary to continue expanding the capabilities and scope of the crop recommendation system, in order to increase its value to the farmers that use it and their families and communities. The ultimate goal is to reduce hunger and poverty in the developing world. This will require advancing the design and implementation of the system and it will require close monitoring of trends, like increasing use of smart phones, that could make the system more useful to its users.

5 CONCLUSIONS

This project is an ambitious undertaking with the potential to really make a difference in people's lives. It has inspired enthusiasm from sponsors, hackathon judges, and especially from the participants in the project. Within the year we hope to take a pilot version of the system into the field and begin to translate this passion for helping others into tangible results.

REFERENCES

Dictionary (2012). *subsistence farming*. Merriam-Webster.com, http://www.merriam-webster.com.

FAO (2012). *pH - TOPSOIL*. Food and Agriculture Organization of the United Nations, http://www.fao.org/geonetwork.

Google (2012). *The Google Geocoding API*. Google, https://developers.google.com.

Hoell, J. M., Stackhouse, P. W., Westberg, D. J., Chandler, W. S., Whitlock, C. H., and Zhang, T. (2007). Climatological to near real time global meteorological data for agriculture, range, and forestry applications. In *Fall American Geophysical Union Meeting*, San Francisco, CA.

ISAC (2012). *The Pineapple Project*. International Space Apps Challenge, http://2012.spaceappschallenge.org.

JSON (2012). *Introducing JSON*. JSON.org, http://json.org.

OpenMarketsFoundation (2011). *Evaluating the impact of an SMS-based Market Information System on smallholder farmers in the Volta Region*. GNAFF.

RHoK (2012). *The Pineapple Project*. Random Hacks of Kindness, http://www.rhok.org.

Strauch, C. (2012). Nosql databases. Technical report, Hochschule der Medien, Stuttgart.

USAID (2012). *The PineApple Project*. United States Agency for International Development, http://idea.usaid.gov.

WorldBank (2007). *World Development Report 2008: Agriculture for Development*. The World Bank, Washington, DC.