Gentry Atkinson
CS7312 Spring 2019
Final Project Report

1. Introduction

Amateur, commonly called Ham, Radio has been a popular hobby in the United States since its introduction in the early 20th century. The hobby is comprised of a loosely regulated network of radio operators who talk to one another through voice, text, and image of various wireless platforms. These operators are all licenced by the Federal Communication Commision and typically the lower levels of licensure only permit means of communication which are limited to a distance of roughly 20 to 50 miles depending on various conditions such as terrain and transmitter strength. In order for these individuals to be able to communicate with greater clarity outside of these distances they rely on radio repeaters. A repeater is an automated system which receives and then re-transmits a signal from another station, effectively doubling the distance that particular set can communicate. In fact, repeaters can often more than double the range of a radio in there network by placing their antenna in high, easily visible locations.

However, the ability of amateur radio enthusiasts to use these repeaters is often limited by their poor and scattered documentation and by the wide variety of means through which radios can communicate. During initial research I was able to find a single, aggregated, online source that was freely available to users. Rather, the American Radio Relay League annually publishes a hard-copy collection of repeaters which are submitted and maintained by the thousands of local radio clubs throughout the United States. The focus of this project therefore is to implement an easily searchable collection of radio repeaters which will return to the user a convenient list of radio repeaters in a specific area that match some criteria set by the user.

Furthermore, this project will employ the “A Priori” technique of frequent pattern mining to identify the most available modes of communication in their area in order to aid with the purchase of new equipment. Radio equipment can be quite expensive and it is a frustrating experience to purchase a new radio only to find it poorly supported in your area of operation. This project will endeavor to spare its users from that experience by analysing the most well supported modes of operation in the user’s area.

2. Common Modes of Radio Operation

An piece of radio communication equipment will share an underlying principle in its operation. The equipment will embed a signal representing some useful information in a “carrier frequency”. A carrier frequency is some pure sine wave taken from the electromagnetic spectrum. A good comparison might be a perfect C taken from the spectrum of audible notes. That carrier is then modified, or “modulated”, by alterations of its amplitude, frequency, or

phase. It is outside the scope of this report to explain at any great length what any of those are or how then might be altered to represent some string of information. Rather, it is only necessary to say that there are many analog (such as AM and FM) and digital (such as DSTAR and ECHOLINK) means of modulating a signal. The modes of modulation supported in this project are: AM, FM, DSTAR, FUSION, DMR/MARC, DMR/BM, and ECHOLINK.

Carrier frequencies are customarily expressed in megahertz. They are also grouped together into "bands" for both convenience and legal reasons. A particular band will have various privileges assigned to it by the FCC, portions of the electromagnetic spectrum might only allow voice communications while others will only permit the transmission of TV images. Particular pieces of hardware are also built to only operate on certain bands. The restrictions of a piece of hardware are dictated both by the legal restrictions previously mentioned and by the expense of adding the circuits to the hardware which would permit it to operate on different bands.

Bands are commonly referred to by the wavelength that most closely fits the range of frequencies represented. So any frequency which lies between 144 mHz and 148 mHz will fall into the 2 Meter band of the spectrum. Likewise any frequency which lies between 420 and 450 MHz is lumped into the 70 Centimeter band. There is a granularity to bands as well, meaning that a band might have sub- and super-bands. So the 2 Meter band also falls into the Very High Frequency band of frequencies which covers all frequencies from 30 Mhz to 300 Mhz. For simplicity's sake this project will focus on bands which share a unified set of privileges as defined by the FCC. In this example all frequencies in the 2 Meter band all share the same privileges (voice or data, no greater than 200 Watts of transmission power) while all frequencies in the VHF band do not. So 2 Meter has been included in the project while VHF has not. The bands supported by this project are: 6 Meter, 2 Meter, 1.25 Meter, 70 Centimeter, 33 Centimeter, and 23 Centimeter.

Finally, many radios include a feature called Squelch which is meant to help silence background noise in radio communication. This is done by embedding some information in addition to the communication signal in the carrier frequency. The most common means of doing this are either a tone in the audible range, or a digital carrier code. Squelch is not necessary and not be used by all repeaters, but using a squelch setting that matches the repeaters is still necessary for communication and so its been included as a searchable feature in this project. The squelch methods supported by this project are: Tone and Digital.

3) Tokenization:

The first step in both the directory searching and pattern mining applications of this project is tokenization. A candidate pull is selected from the full directory based of the locations defined by the user. The system allows the user to input as many city names as necessary because radio waves easily travel far enough to allow communication between several cities.

Both the candidate pool and and the user query are then compared to a series of regular expressions to produce an array of tokens which describe a station.

Regular expressions were chosen because they allow great flexibility in the type of input that the system will take. So the strings “145.8 Mhz” and “2 meter” will both push an instance of the “2M” token into the array of tokens. Flexibility is necessary both because it adds to the user’s enjoyment of the system and because there is not a well unified standard for formatting data within the ARRL’s directory. It is important to remember that all of the information is the directory is submitted by local radio clubs with very little revision, so an in-flexible system would have had difficulty coping with structure of the data.

4) Directory Search

After the document pool and the user query have been tokenized the documents are sorted according to the Jaccard similarity. Jaccard was chosen because this particular information needs lends itself better to Boolean Searching rather than Vector Space, HITS, or other newer algorithm. Essentially, the user base of this project are expected to be knowledgeable in the subject matter and will desire results that match specific needs, e.g. a particular band. This search space was found to be more similar to library or legal searches than it is to web searches and so an appropriate similarity measure was chosen with that in mind.

From the sorted candidate pool, up to 10 documents are returned to the user. This was done to facilitate the read-ability of the resulting web-page. Users who have not fulfilled their information need with the first 10 results will have to restructure their query and re-search the database. Some locations will not 10 repeaters available and so will return the entire candidate pool, sorted best-first for the user to peruse.

5) A-Priori Recommender

After the search results are returned to the user the system will process the tokens generated by the candidate pool to generate the set of all tokens returned for the user-specified location. Candidates for the A-Priori Process are generated by self-combination. Pruning was considered for this project but rejected on the grounds that with the small sets generated by this particular data set, it would cause more work for the system than it would save. The candidate list is then compared to the data to calculate their support. Candidates with support less than the minimum support value (defined as 0.1 for this project) are eliminated.

This process iterates until no candidates pass the minimum support value. The first of the previous candidate set is then returned to the user as a recommendation. This recommendation is meant to aid users in the selection of equipment that is appropriate for their area of operation. As such this process is not dependent on the user’s entered query. The only consideration for this section is the frequency with which combinations of tokens appear in a particular search area.

6) Results

To test the efficacy of the system several artificial queries were generated and then performed by the system so that the validity of the results could be measured. The queries are:

1. Austin, Allen, DTAR, digital squelch
2. Arlington, fm, 445 MHz
3. Beaumont, Blanco, 6m

First Query:
7 relevant results exist in the directory. 10 results are returned with 7 being relevant. This gives us:
Precision: 0.7
Recall: 1.0

Second Query:
5 relevant results exist in the directory. 8 results are returned with 4 being relevant. This gives us:
Precision: 0.5
Recall: 0.8

Third Query:
2 relevant results exist in the directory. 10 results are returned by the search with 2 being relevant. This gives us:
Precision: 0.2
Recall: 1.0

Overall:
Precision: 28 documents returned with 13 being relevant = 0.46
Recall: 14 extant relevant documents with 13 being returned = 0.93
F-Measure: 0.62

Overall it is reasonable for this project to prioritize recall of precision. It is more important for users to find the complete set of repeaters that match their query than it is for all of the results to be relevant. The additional results are provided to the user for his or her convenience and do not detract from the usefulness of the other results. Furthermore, 10 results are not difficult to read in a well formatted table so it is no an unreasonable burden for the user to peruse all 10. With this in mind the system achieved an overall recall of 93% which is very good and more than sufficient to satisfy a hobbyist.

The three test queries returned the following equipment recommendations:

1. FM, 70cm, tone squelch
2. FM, tone squelch
3. FM, 6m, tone squelch

7) Conclusions

Amateur radio has attracted a dedicate base of hobbyists for more than 100 years. In many senses this group has been quick to adopt new technologies in the form of communication protocols, antennas, and digital processing. But in terms of ease-of-use or quality-of-life technologies the hobby of lags behind. Rather, the difficulty of practicing the hobby is often held as a badge of honor. It is the aim of this project to make the hobby more accessible to newcomers and more easily practicable by its veterans.

By making repeaters directories more easily accessible, users will have an easier time engaging their local networks. By receiving data-based recommendations regarding equipment purchase, users can be more confident that they will receive a value for their purchases that is equal to its cost.

This project has developed a system which consistently gives high-recall search results for a specified user area. It has also developed a recommender which gives its users reliable insight into the type of equipment that is most commonly being used in their area of operation. Wide adaptation of this, or of a similar system, will help sustain the practice of Ham Radio going forward into an increasingly wireless future.

To use this system live, please visit: https://userweb.cs.txstate.edu/~gma23/directoryHome.html
(Please keep in mind that not all cities have been entered in the data set yet.)
All code is available at: https://github.com/gentry-atkinson/cs7312_assignments