ample, if *Crystal* identified 9 messages predicting for Conservative Party, 3 messages for NDP, and 1 message for Liberal among 13 messages in the riding "Blackstrap", the system will predict that the Conservative Party would win in "Blackstrap".

Table 4 shows the system performance with Acc_{riding}. Note that people who write messages on a particular web site are not a random sample for prediction. So we introduce a measure of confidence (*ConfidenceScore*) of each system and use the prediction results when the *ConfidenceScore* is higher than a threshold. Otherwise, we use a default party (i.e., the incumbent party) as the winning party. *ConfidenceScore* of a riding *R* is calculated as follows:

$$ConfidenceScore = count_{message}(P_{first}) - count_{message}(P_{second})$$

where $count_{message}(P_x)$ is the number of messages that predict a party P_x to win, P_{first} is the party that the most number of messages predict to win, and P_{second} is the party that the second most number of messages predict to win.

We used 62 ridings to tune the *ConfidenceScore* parameter arriving at the value of 4. As shown in Table 4, the system which just considers the incumbent party (INC) performed fairly well (78.03% accuracy) because incumbents are often re-elected in Canadian elections. The upper bound of this prediction task is 88.85% accuracy which is the prediction result using numerical values of a prediction survey. FRQ and MJR performed 63.14% and 36.63% respectively. Similarly to Evaluation1, JDG which only uses judgment word features performed worse than both *Crystal* and NGR. Also, *Crystal* with our feature generalization algorithm performed better than NGR with non-generalized n-gram features. The accuracy of *Crystal* (81.68%) is comparable to the upper bound 88.85%.

6 Discussion

In this section, we discuss possible extensions and improvements of this work.

Our experiment focuses on investigating aspects of predictive opinions by learning lexical patterns and comparing them with judgment opinions. However, this work can be extended to investigating how those two types of opinions are related to each other and whether lexical features of one (e.g., judgment opinion) can help identify the other (e.g., predictive opinion). Combining two types of opinion features and testing on each domain can examine this issue.

In our experiment, we used General Inquirer words as judgment opinion indicators for JDG baseline system. It might be interesting to employ different resources for judgment words such as the polarity lexicon by Wilson et al. (2005) and the recently released SentiWordNet[12].

Our work is an initial step towards analyzing a new type of opinion. In the future, we plan to incorporate more features such as priors like incumbent party in addition to the lexical features to improve the system performance.

7 Conclusions

In this paper, we proposed a framework for working with *predictive opinion*. Previously, researchers in opinion analysis mostly focused on judgment opinions which express positive or negative sentiment about a topic, as in product reviews and policy discussions. Unlike judgment opinions, predictive opinions express a person's opinion about the future of a topic or event such as the housing market, a popular sports match, and election results, based on his or her belief and knowledge. Among these many kinds of predictive opinions, we focused on election prediction.

We collected past election prediction data from an election prediction project site and automatically built a gold standard. Using this data, we modeled the election prediction task using a supervised learning approach, SVM. We proposed a novel technique which generalized n-gram feature patterns. Experimental results showed that this approach outperforms several baselines as well as a non-generalized n-gram approach. This is significant because an n-gram model without generalization is often extremely competitive in many text classification tasks.

This work adopts NLP techniques for predictive opinions and it sets the foundation for exploring a whole new subclass of the opinion analysis problems. Potential applications of this work are systems that analyze various kinds of election predictions by monitoring texts in discussion boards and personal blogs. In the future, we would like to

[12] http://sentiwordnet.isti.cnr.it/