# Concession Curve Analysis for Inspire Negotiations

Vivi Nastase
School of Information Technology and Engineering
University of Ottawa, Ottawa, ON
vnastase@site.uottawa.ca

April 10, 2006

**Abstract**

In the course of a negotiation it is often the case that the participants exchange packages of offers, which have, at least in the mind of the negotiators, a certain utility for them. We want to test whether the behaviour of the negotiators is reflected in the topology of the concession curve[1] that plots each offer's utility value in the course of a negotiation. In order to do this, we use data collected with the Inspire electronic negotiation support system, which records utility preference values for all issues under discussion, for each negotiator. We abstract the concession curves using a set of features, such as number of minima and maxima, slope of curve at the beginning and end, and then we use machine learning techniques to test whether we can predict negotiation outcome based on these concessions curve descriptions. We find that there are certain features of this curve, such as the number of minima and maxima, frequency of offers exchanged, that predict with high precision and recall the outcome of negotiations conducted with Inspire.

## 1 Introduction

Negotiation support systems (NSSs) are designed to help the negotiators reach an (optimal) agreement by offering analytical support (SmartSettle[2]),

[1]We use the term *curve* to refer to the function graph of the two-dimensional utility function.

[2]www.smartsettle.com

communication support (Schoop et al., 2003) or both (Kersten and Noronha, 1999). Analytical support comes in the form of numerical evaluation of the offers exchanged, which allows the negotiators to assess their position in the negotiation. Having a record of the numerical values associated with the offers exchanged gives us the opportunity to study how the behaviour of the negotiators is reflected in the evolution of these values during the course of the negotiation, and to study whether we can predict the outcome of the negotiation based on the characteristics and trends observed.

We work with data recorded by the web-based NSS Inspire (Kersten and Noronha, 1999). Inspire has a record of more than 3000 bilateral negotiations conducted since 1996. The negotiation cases available in the system are relatively simple, with a fixed number of issues to discuss (such as *price, delivery time, payment options*), and with a small and fixed number of options (possible values) for each issue. In order to help negotiators assess their position during a negotiation, the Inspire system has a preference rating stage at the beginning of the negotiation process. This stage has three steps, in which the user rates her preferences for the issues under debate, for each option of these issues, and then for several packages which contain one option for each issue under debate. Based on these preferences, the system will give the users an evaluation of their position at the negotiation table for each offer exchanged, in graphical and numerical format.

The sequence of offers exchanged during the negotiation, and the variation in offer utility values, may be very indicative of how the process is going along. This paper presents a study that verifies if this is indeed the case. We represent the sequence of offer utility values through a variety of measures. We use a machine learning system to find if this set of measures (features) can predict the outcome of the negotiation process, and which of the features used are most relevant for this task. We compare the results of predicting negotiation outcomes based on these measures, with results of predicting negotiation outcome using other (different) features based on the same Inspire data.

We discuss related work in section 2, and then we present the data we work with in section 3. The experiments performed, and variations on the data sets used, are discussed in section 4, and the results obtained in section 5. In section 6 we draw conclusions and present some ideas for future work.

## 2 Related Work

Having a numerical value that evaluates the strength of offers exchanged during negotiation procedures, could be very helpful and give important information to negotiators about their position and gains from the ongoing negotiation. Studies of utility values come from multi-attribute utility theory, and experimental evaluation and comparison of methods for eliciting and combining multi and single-attribute utility values (Zanakis et al., 1998), (Fischer et al., 1999), (Kimbrough and Weber, 1994), (Borcherding et al., 1991), (Beroggi, 2000).

Vetschera (2004) presents an analysis of utility values in negotiations conducted with Inspire. The study focuses on cultural and professional information about the negotiators, combined with analysis of single attribute values for the different attributes under debate in Inspire negotiations. The aspects followed are monotonicity of the values assigned, and convexity/concavity of the curves for the values of the individual attributes.

We analyze the concession curve that plots the values of offers over time, and not individual attribute values. Two pairs of sample curves are presented in Figure 1. Each pair of curves corresponds to a pair of negotiators (which negotiate with each other), one that succeeds in reaching an agreement, and one that doesn't.

One of the reasons for analyzing the curve through various measures that summarize and abstract it, is the fact that the actual values on the concession curve are relative, and mean different things for different users. For example, a value of 70 for one user may be a high utility value, while for another it may not be good enough. We replace such absolute values with means and other measures that abstract away from the absolute values, and give a more consistent view of the curve across negotiations. We have designed a set of 4 experiments which will verify that there are characteristics of the change of utility values during the negotiation process that can predict whether the negotiation will be successful or not. Also, we will identify which of these characteristics, or features, bear the most predictive power.

## 3 Data

We extract the data we work with from 3063 negotiations (6126 negotiators) conducted with the negotiation support system Inspire. This data

Agreement reached | No agreement

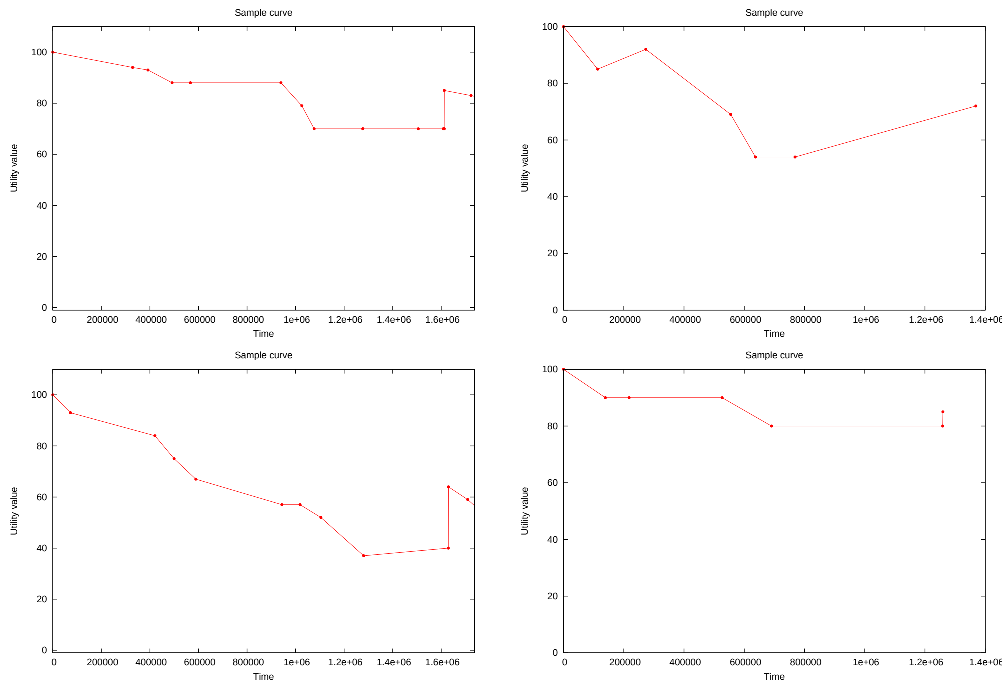


Figure 1: Sample concession curves for two pairs of negotiators

can be grouped into 3 classes, corresponding to 3 possible outcomes of the negotiation process which Inspire records:

- successful (agreement is reached) – 1626 instances,
- one-sided (one negotiator is not responsive) – 668 instances,
- failed (there is no agreement recorded by the system) – 769 instances.

Inspire keeps a transcript of each negotiation, in which are recorded all actions taken by the two users: preference rating steps, messages and offers exchanged, utility values for the offers exchanged, the agreement offer, if any, and post-settlement offers, if any.

We extract from these 3063 transcripts (correponding to 6126 users) the utility values for each offer exchanged and the time when the offer was made, grouped by users. This process will filter out negotiations in which utility values for an offer are not numbers[3]. Users which exchanged only one offer were eliminated from our data as well, since we cannot perform curve analysis on a single point. This leaves us with 4311 instances in the data set, each instance consists of a vector of *¡offer utility values, time point¿* pairs, that capture the offer exchange from one user in the course of one negotiation. For each of the experiments we perform, we extract different features from the utility values information in this data set (which we call *DataSet X*).

## 4 Experiments

We process the vectors of *¡utility value, time point¿* pairs from *DataSet X* in various ways, and we extract measures that describe the curves of the utility values during the negotiation process. After representing the data in terms of these measures (or features), we use machine learning tools to test whether we can predict the outcome of the negotiations based on these features.

The machine learning tool of choice is C5.0 (Quinlan, C50 version 51). This is a tool most frequently used for many classification tasks. We choose it because it has certain characteristics that make it very appropriate for our task:

- it considers each feature separately, according to how well the feature helps distinguish examples from different classes;
- not all features are used in classification;
- the classification model produced (a tree) is easy to understand, and shows us which features have the most predictive power with respect to the negotiation outcome.

In representing our data, we include as features measures of the concession curve that we think are relevant. They may not all have the same relevance, some of them may be totally useless. Because C5.0 is able to choose the best

[3]The Inspire system occasionally assigns a value of "N/A" or other strings for a utility value it cannot compute, and occasionally corrupt values appear as well ("i100" instead of "100").

features at each step that describe the data, and it gives us the possibility to see the features chosen and understand the phenomena in our data, we think this is an appropriate tool to use. Also, in previous work on Inspire data, C5.0 gave the best results from all the tools used. We will try other ML tools in future work. Section 5 presents in more detail how C5.0, and decision tree learning tools in general, works.

### 4.1 Experiment 1

We build *DataSet 1* from the vectors of utility values and time point pairs in *DataSet X*, by extracting the following features to describe the concession curves:

- USER_ID: a unique user identifier;
- CASE_TYPE: the case that the negotiators are debating. Inspire offers a set of cases, each with a small number of issues to debate, and each issue with a small set of possible values;
- START_POINT: the utility value of the first offer for this user;
- END_POINT: the utility value of the last offer for this user;
- DURATION: the duration of the negotiation process, in hours;
- EXCHANGE_FREQUENCY: the average time interval (in hours) between consecutive offer exchanges;
- MAXIMA: the number of maxima in the concession curve;
- MINIMA: the number of minima in the concession curve;
- START_CURVE: the slope of the curve at the beginning of the process (ascending (1), flat (0) or descending (-1));
- END_CURVE: the slope of the curve at the end of the process;
- START_ANGLE: the tangent of the curve at the beginning of the process;
- END_ANGLE: the tangent of the curve at the end of the process;
- OUTCOME: the outcome of the negotiation process: FAILED, ONE SIDED, SUCCESSFUL.

| Data set | successful | one sided | failed |
|---|---|---|---|
| *DataSet 1*<br>*DataSet 3* | 3088 | 68 | 1155 |
| *DataSet 2*<br>*DataSet 4* | 1502 | | 502 |

Table 1: Distribution of negotiation outcomes in the four data sets used in experiments.

### 4.2 Experiment 2

*DataSet 1* contains instances that represent only one user in a negotiation. However, we would like to verify the changes in the utility curve of one user as a response to the changes made by his/her counterpart. For this reason, we build *DataSet 2*, which pairs up users from the same negotiation, while the concession curve corresponding to each user is described in the same terms as in *DataSet 1*. Each feature in *DataSet 1* will appear twice, except for OUTCOME, which is the same for the two users in one negotiation. Because not both users from a negotiation appeared in *DataSet 1* (because of filtering based on corrupted values and single offer exchanges), *DataSet 2* will contain 2004 instances (information on pairs of negotiators).

### 4.3 Experiment 3

From previous experiments with numeric information from Inspire data (Kersten and Noronha, 1999), and from our experiments 1 and 2, we observe that EXCHANGE_FREQUENCY is a very powerful feature, so much so that it takes over the classification process, and does not allow us to study the impact that other features might have in the prediction of negotiation outcomes. We decided therefore to eliminate it, and introduce instead other measures. We build then *DataSet 3*, which includes all features in *DataSet 1* except EXCHANGE_FREQUENCY, plus the following new measures:

- AVERAGE: the average concession[4] made during the course of the negotiation by one negotiator;
- MAX_CONCESSION: the maximum concession;

[4] By *concession* we mean the difference between consecutive offers.

- MAX_CONCESSION_INDEX: the normalized index of the point on the curve where the maximum concession was found (normalized with respect to the total number of offers made);
- MIN_CONCESSION: the minimum absolute concession;
- MIN_CONCESSION_INDEX: the normalized index of the point on the curve where the minimum concession was found (normalized with respect to the total number of offers made);

The size of this set is equal to the size of *DataSet 1* – 4311 instances.

The distribution of outcomes in *DataSet 3* and *DataSet 1* is presented in Table 1.

### 4.4 Experiment 4

Just as for experiment 2, we build from data that describes single user information, a new data set that describes the pair of negotiators in each negotiation. *DataSet 4* consists of instances that describe a pair of users, each user is described in terms of the features from *DataSet 3*, and the OUTCOME appears only once, since it has the same value for the users in the same negotiation. This set, just like *DataSet 2* contains 2004 instances, and has the same distribution of outcomes, presented in Table 1.

## 5 Results

The machine learning tool we use is C5.0 (Quinlan, C50 version 51). C5.0 uses information to build decision trees - at each step C5.0 chooses a feature from the ones that represent the data which produces the most ordered (pure) split of the data set in that node. For a data set $S$, and a feature $F$ with the set of possible values $VF$, the information gain by splitting the set $S$ by feature $F$ is:

$$Gain(S,F) = Entropy(Set) - \sum_{v \in VF} \frac{|S_v|}{|S|} Entropy(S_v)$$

where $|S_v|$ is the cardinality of the subset of instances where the feature $F$ takes value $v$, $|S|$ if the total number of instances in set $S$, and the entropy measures how disordered a set is:

| Exp. | Precision successful | Recall successful | Precision one sided | Recall one sided | Precision failed | Recall failed | Accuracy |
|---|---|---|---|---|---|---|---|
| 1 | 76% | 94% | 42% | 21% | 58% | 22% | 73.76% |
| 2 | 78% | 93% | | | 51% | 23% | 75.19% |
| 3 | 77% | 92% | 43% | 19% | 55% | 27% | 73.85% |
| 4 | 78% | 95% | | | 59% | 20% | 76.44% |

Table 2: Precision and recall results for negotiation outcome classification.

$$Entropy(S) = \sum_{i=1}^{c} p_i log_2 p_i$$

where $p_i$ is the proportion of instances in dataset $S$ that take the $i$-th value of the target attribute, and $c$ is the number of classes in the dataset. High entropy values mean the dataset is very disordered, that there is an approximately equal mixture of classes . Low entropy values mean the dataset is relatively pure, with one predominant class.

The machine learning tool builds a classifier based on the training data. It is then run on the test data, and the performance is measured using precision, recall and accuracy.

For a class $C$, **precision** shows how many examples, out of all those that the classifier assigns to class $C$, are classified correctly. If *TP(C)* is the number of examples that belong to class $C$ and which the classifier handles correctly (true positives), and *FP(C)* is the number of examples that the classifiers assigns, incorrectly, to class $C$ (false positives), precision $P(C)$ of class $C$ is defined as follows:

$$P(C) = \frac{TP(C)}{TP(C) + FP(C)}$$

$TP(C) + FP(C)$ is the total number of examples that the classifier assigns to class $C$.

For a class $C$, **recall** shows how many examples, out of all those that belong to class $C$, are classified correctly. If *TP(C)* is the number of true positives, as defined above, and *FN(C)* is the number of examples that the classifier assigns incorrectly to other classes than class $C$ (false negatives), the recall

$R(C)$ of a class $C$ is defined as follows:

$$R(C) = \frac{TP(C)}{TP(C) + FN(C)}$$

$TP(C) + FN(C)$ is the total number of examples that belong to class $C$ ($TP(C) + FN(C) = |C|$).

The **accuracy** is the number of examples classified correctly (for all classes represented in the dataset), out of the total number of examples in the dataset.

$$Acc = \frac{\sum_{i=1}^{n} TP(C_i)}{\sum_{i=1}^{n} |C_i|}$$

For each of the four data sets (*DataSet 1 - 4*) we perform 5 fold cross-validation experiments using C5.0, where the target attribute is OUTCOME, the outcome of the negotiations. By performing 5 fold cross validation (in which the input data set is randomly split into 5 equal parts, there are 5 rounds of experiments in which alternatively one set is kept for testing while the other four are used for training) we insure that the results reported are not skewed by accidental regularities in the data. The precision and recall results we report are averages over the precision and recall obtained for the 5 folds.

Table 2 shows the results in terms of precision and recall for each of the four experiments.

An interesting observation arrises by comparing the results obtained for data representing single negotiators (experiments 1 and 3), with the results obtained for data representing pairs of negotiators (experiments 2 and 4). The results are quite close (a difference of 1-2%). This means that it is enough for us to look at the behaviour of a negotiator alone (and not within the pair), in order to tell whether the negotiation will be successful or not. Looking at the interaction, in the way in which we have done, does not bring much new information. We plan to explore this issue further, by trying to match negotiator moves in a pair, and verify if this representation will help in predicting negotiation outcomes better.

Analysis of Inspire data has been done before.

Sokolova and Szpakowicz (2005) analyse the textual messages exchanged during the negotiation process. Because of the particularities of textual data, the study focuses on a subset of Inspire data that have the same negotiation topic – sale/purchase of bicycle parts. The experiments reported in (Sokolova and Szpakowicz, 2005) and (Shah et al., 2004) cover a variety of machine learning tools and various learning paradigms. The best results were obtained also with C5.0, when an accuracy of 74.5% was reported. Apart from the similarity in the accuracy obtained, another interesting point comes up from this research: just as in the analysis of numerical data, textual data also allows for the better classification of successful negotiations as opposed to failed negotiations for almost all the tools used. The exception is Naive Bayes, which performs better at identifying failed negotiations. For future work we also plan to apply this method to our data, and verify if the same phenomenon will occur for our numerical data.

Kersten and Zhang (2003) performed analysis on the numerical data (absolute values recorded in transcripts) and negotiator's personal information (extracted from pre and post negotiation questionaires recorded by Inspire) from a subset of 1525 negotiations (the set of negotiations available at the time of this study). They also try out a variety of tools (linear regression, decision tree and rule induction, neural networks). The best accuracy was also obtained for decision tree induction – 75.33%.

Analyzing the decision trees built by C5.0 for each of these experiments, we observe that for experiments 1 and 2 the attribute EXCHANGE_FREQUENCY dominates.

In order to examine the impact of the other features on the outcome, we eliminated this feature and added other concession curve measures.

Analysis of the decision trees gives us some interesting information:

- the number of MAXIMA of a curve is a good indicator of successful outcome;
- when the negotiation starts with the offer with the highest rating it most likely will fail;
- if the average concession is less than 45.17, and the maximum concession is made within the first 2/3 of the negotiation process, a successful outcome is likely.

## 6 Conclusions

The precision and recall results obtained show that concession curve features are very indicative of the negotiation outcome. The fact that successful negotiations have the highest precision and recall values tells us that the features that characterize the change of offer utility values in the course of the negotiation identify best negotiations that succeed. From the set of features used to represent the data, several seem to have good predictive power: frequency of exchange, number of maxima, concession average, the value of the maximum concession made and when this happened.

The analysis presented can evolve in several directions.

We can deepen the analysis of paired negotiators, by verifying whether a concession made by one party is reciprocated by the other. Or in general, we can study how each move of one negotiator is being responded to by the negotiation partner.

Another possible step would be to combine the measures of the utility curves we used with cultural and professional background information for the negotiators that participate in the experiments from which we collected our data. It would be interesting to see whether there are correlations between cultural background and behaviour, as it is captured by the changes in offer utility values.

Also, it would be interesting to combine characteristics of the utility curves with textual information exchange during negotiations. The messages exchanged (together with the offers, or independently) carry interesting information about the behaviour and tactics of the negotiators (Sokolova and Szpakowicz, 2005). We would like to investigate whether the behavioural indicators from the two sources – utility values and textual messages – are consistent with each other.

## References

G.E. Beroggi. 2000. An experimental investigation of preference elicitation methods in policy decision-making. *Journal of Multi-Criteria Decision Analysis*, 1-3(9):76–89.

K. Borcherding, T. Eppel, and D. von Winterfeldt. 1991. Comparison of weighting judgements in multiattribute utility measurement. *Management Science*, 12(37):1603–1619.

G. Fischer, Z. Carmon, D. Ariely, and G. Zauberman. 1999. Goal-based construction of preferences: task goals and the prominence effect. *Management Science*, 8(45):1057–1075.

Gregory Kersten and Sunil Noronha. 1999. Www-based negotiation support: Design, implementation and use. *Decision Support Systems*, 2(25):135–154.

Gregory Kersten and Grant Zhang. 2003. Mining Inspire data for the determinants of successful Internet negotiations. *Central European Journal of Operational Research*, 3(11):297–316.

S. Kimbrough and M. Weber. 1994. An empirical comparison of utility assessment programs. *European Journal of Operational Research*, (75):617–633.

Ross Quinlan. C5.0, version 5.1. C5.0. http://www.rulequest.com.

Mareike Schoop, A. Jertila, and T. List. 2003. Negoisst: A negotiation support system for electronic business-to-business negotiations in e-commerce. *Data and Knowledge Engineering*, 3(47):371–401.

Mohak Shah, Marina Sokolova, and Stan Szpakowicz. 2004. The role of domain-specific knowledge in classifying the language of e-negotiations. In *ICON 2004*, pages 99–108.

Marina Sokolova and Stan Szpakowicz. 2005. Analysis and classification of strategies in electronic negotiations. In *Advances in AI 18, Canadian AI 2005*, pages 145–157.

Rudolf Vetschera. 2004. Preference structures and behavioural consistency in negotiations. http://interneg.concordia.ca/interneg/research/papers/.

S.H. Zanakis, A. Solomon, N. Wishart, and S. Dublish. 1998. Multiattribute decision making: a simulation comparison of select models. *European Journal of Operational Research*, 3(107):507–529.