Journal of Universal Computer Science, vol. 12, no. 2 (2006), 127-139
submitted: 10/2/04, accepted: 3/7/05, appeared: 28/2/06 © J.UCS

Multiple Explanations Driven Naïve Bayes Classifier

Ahmad Almonayyes
(Dept. of Mathematics & Computer Science
Kuwait University
P.O.Box: 5969, Safat, Kuwait 13060
sami@mcs.sci.kuniv.edu.kw)

Abstract: Exploratory data analysis over foreign language text presents virtually untapped opportunity. This work incorporates Naïve Bayes classifier with Case-Based Reasoning in order to classify and analyze Arabic texts related to fanaticism. The Arabic vocabularies are converted to equivalent English words using conceptual hierarchy structure. The understanding process operates at two phases. At the first phase, a discrimination network of multiple questions is used to retrieve explanatory knowledge structures each of which gives an interpretation of a text according to a particular aspect of fanaticism. Explanation structures organize past documents of fanatic content. Similar documents are retrieved to generate additional valuable information about the new document. In the second phase, the document classification process based on Naïve Bayes is used to classify documents into their fanatic class. The results show that the classification accuracy is improved by incorporating the explanation patterns with the Naïve Bayes classifier.

Key words: Case-Based Reasoning; Data mining; Explanation Patterns; Naïve Bayes; Text classification
Category: I.1.2, I.1.7, I.2.1, I.2.6

1 Introduction

The knowledge of foreign language plays a big role in intelligence and counter-terrorism. The intelligence community relies heavily on language to create finished intelligence products for decision makers. The information is gathered from intelligence reports, embassy reporting, media news, internet which is now increasingly in non-English languages, or other resources. Of course, the finished product is in English, but the input may come from several different foreign languages and need to be evaluated by a range of people with the ability to translate and interpret the data in its original language within its particular context. A lack of language skills can limit intelligence analyst insight into foreign culture, constraining their ability to understand and anticipate deterioration in a particular situation, and hence, endangering national security readiness to confront a potential danger. For example, it has become clearer than ever that those events in the Middle East affect our daily lives. The world today faces a critical shortage of linguistically competent professionals to assist intelligence analysts in classifying Arabic-written documents (e.g. emails) which may contain information that would be harmful to the world stability.

Arabic is considered a difficult language to learn due to the fact that it has many forms, the modern standard (the written language), and Arabic dialect (the spoken

form in one country or region). Therefore, while most text mining research concentrate on processing English documents only, mining from documents written in other languages allow access to previously unexploited information and offers a new host of opportunities.

Data can be found in many different forms. Some formats are more appropriate for automatic data analysis and easier to handle than others. The usual data analysis methods assume that the data is well- defined in a number of fields with a predefined range of possible values. The question is what can be done if the data is stored purely in textual form, consisting of no records and no variables. Several document categorization techniques were developed to classify documents into pre-defined categories based on the vector-based model. The dimensions of the vector space are formed by the important words given in the documents. The documents that have already been categorized, according to the distances between the vectors, are used to generate model for assigning content categories to new documents. [Mitchell 1997] describes techniques to integrate machine learning and data mining for data analysis with varying knowledge representations and large amounts of data. [Cohen 1996] discusses rule-based learning classifier RIPPER in the context of mail filtering. RIPPER forms sets of simple rules for data described by sets of attribute-value pairs. Each rule tests a conjunction of conditions on attribute values. Rules are returned as an ordered list, and the first successful rule provides the prediction for the class label of a new example. The system uses large batches of training data to learn the rules in a greedy fashion. The classifier must constantly be kept up-to-date and training and classification are highly intertwined since new rules are formed when a sufficient amount of data has been covered.

Another classification algorithm that provides efficient training and quick classification is Naïve Bayes [Hastie et. al. 2001] [Lewis 1994] [McCallum et. al 1998] [Mitchell 1997]. In this algorithm, adding a document to a trained model requires the recording of word occurrence statistics for that document, no rule need to be learned and no weights need to be optimized. Training consists of updating word counts and classification consists of normalized sum of counts corresponding to the words in question. Hence, training and classification are both simple and efficient and can be integrated into the learning model.

Another classification approach uses background knowledge as indices into the set of labeled training examples [Zelikovitz et. al 2000]. If a piece of knowledge is close to both a training example and a test example, then the training example is considered close to the test example, even if they do not share any words. In this way, the background provides a mechanism by which the labeled examples are chosen to be used for classification of a new test example. However, these approaches neglect the explanations of why particular categories have been formed and how the different categories are related to each other. Some aspects of text mining involve natural language processing [Jackson et. al. 2002] [Manning et. al 2001] where the model of reasoning about a new text document is based on linguistic and grammatical properties of the text, as well as extracting information and knowledge from large amount of text documents.

In this paper, we focus on processing Arabic-written documents (standard and Arabian Gulf dialects) in order to classify, extract, and analyze information about fanaticism. The system incorporates Naïve Bayes classifier with Case-Based

Reasoning (CBR) [Kolodner 1993] to classify and analyze texts of fanatic content. The Bayes classifier uses the training set to create a probabilistic model based on a knowledge structure called an eXplanation Pattern (XP) [Mitchell 1997] which explains events (i.e. facts) in the document. An XP is a directed, cyclic graph of concepts which represents a specific aspect of fanaticism, and indicates a particular fanatic category. Similar cases of fanaticism are indexed under each abstract XP which explains them. The CBR model analyzes large collection of unstructured textual documents for the purpose of extracting interesting patterns of knowledge. The strength of this approach lies in the fact that background knowledge (i.e. XPs) is used to place texts that do not share any syntactic words in close proximity to each other based on the causal structures which link those words together. Furthermore, the relevance of features (i.e. facts presented in the causal structure of an XP), is used in the computation of posterior probability for each class.

This paper is organized as follows. Section 2 gives the experimental methodology used to collect data and represent Arabic concepts. An overview of Case-Based reasoning, and the case classification algorithm are all presented in section 3. The performance results including a sample run of the system are given in section 4. Finally, section 5 presents our conclusions, limitations, and future work.

2 Experimental Methodology

In our domain, text documents consist of a set of independent, but semantics-oriented Arabic vocabularies which are automatically converted to their equivalent English words. The semantic links between words are used to relate them conceptually. The process of conversion is done in straightforward fashion by searching database of Arabic words to find their corresponding meanings in English. A conceptual hierarchy structure [Kolodner 1996] is used to connect words which are compatible, that is, it puts them into equivalent semantic hierarchy. The idea here is that if two words express similar concepts, they will be organized under the same conceptual category. Hence, two lexically dissimilar words (e.g. two words from different Arabic dialects but having similar meanings) considered conceptually similar if both are specialized concepts of a more general one. If a word is not found in the conceptual hierarchy, it will be treated as a surface concept. For example, despite the fact that the word *Shiite* is lexically different from the word *Jaafri*, they are conceptually similar because both represent the same *Shiite* doctrine, and both correspond to the same religion. It is also proposed that the frequency of concept occurrence in the explanation structures provides a useful measurement of the concept significance. Thus, the process we use for extracting features and reducing the feature set to a manageable size is based in three steps: 1) gather training sets of fanatic and non-fanatic texts, remove stop-words, and stem the words[1] 2) generate vocabularies of concepts occurring in the texts by using the conceptual hierarchy structure, and 3) count the frequencies of each concept in the training data.

The data collection process is started by distributing a list of 10 questions to 1041 university students whom were asked, anonymously, to write their views about

[1] Stemming converts each word to its neutral form. This process incorporates a great deal of linguistic knowledge in Arabic and it is beyond the scope of this paper.

various events involving fanaticism and terrorism (e.g. Sep. 11 terrorist attack, USA policy in Afghanistan, etc.). Each document in the corpus[2] averaged roughly 100-400 words. The documents in the training set have been analyzed by an expert in radicalism to extract crucial features used to express views of fanatic trend. 320 concepts were identified. Domain experts were asked to categorize documents into three classes, Not-Fanatic (NF), Code-Attitude Fanaticism (CAF), and Code-Red Fanaticism (CRF). The CAF indicates signs of hostility towards the adversaries, but with no violent tendencies. CRF indicates a willingness to use violent actions against adversaries.

3 Case-Based Reasoning

CBR is an Artificial intelligence (AI) approach that can allow the agent to learn from past successes. It is a method that finds the solution to the new problem by analysing previously solved problems, called cases, or adapting old solutions to meet new demands [Almonayyes 2001] [Kolodner 1993]. It suggests a model of reasoning that incorporates problem solving, understanding and learning and integrates all with a case base library. CBR has the capability of using previous solutions to suggest solutions to new problems instead of solving the problem from scratch. CBR accesses a dynamic memory to store and retrieve previous experiences. CBR also recognizes the similarity between cases so that only those potentially applicable to the current problem are recalled. As well, it chooses the most appropriate case from a set of potentially applicable ones. Then, CBR has the ability to transfer the appropriate information from one case to another and then receives a feedback from the real environment to evaluate its decisions and learn from them.

This work implements a computational approach for generating a list of explanatory hypotheses that account for a *fanatic aspect* of the situation. The list is called Explanatory Hypothesis Set (EHS) [Almonayyes 1997]. Each explanatory hypothesis in the set represents a causal explanation structure which pertains for a particular source of fanaticism. Two-step process is used to generate the EHS. The first step is to use the results of a hardwired inference process which is based on a discrimination network of multiple questions. The network is used to guide selection of important features, before an explanation is available. The knowledge structure associated with each question highlights certain aspects of the situation. The intuition here is that cases which are similar to one another are clustered together below the XP which describes a particular viewpoint of fanaticism. In a discrimination network tree each internal node represents a question associated with multiple branches of the tree (i.e. different answers to the question). The program traverses the network by answering the question at each node. When the system reaches an internal node where there is a lack of information in the new case (i.e. email) to answer the current question, then all XPs below this node are retrieved for potentially describing the case. The point here is to be able to improve the information-gathering capabilities by providing the intelligence operators with the much-needed sensitive information on terrorism while disregarding irrelevant ones. On the other hand, if the system reaches

[2] Due to privacy concerns regarding emails, the data sets are constructed to mimic the structure and properties of personal email.

a leaf node (i.e. an explanation pattern) by answering all the questions in an internal node's siblings, then that only XP is returned as the output of the traversal process.

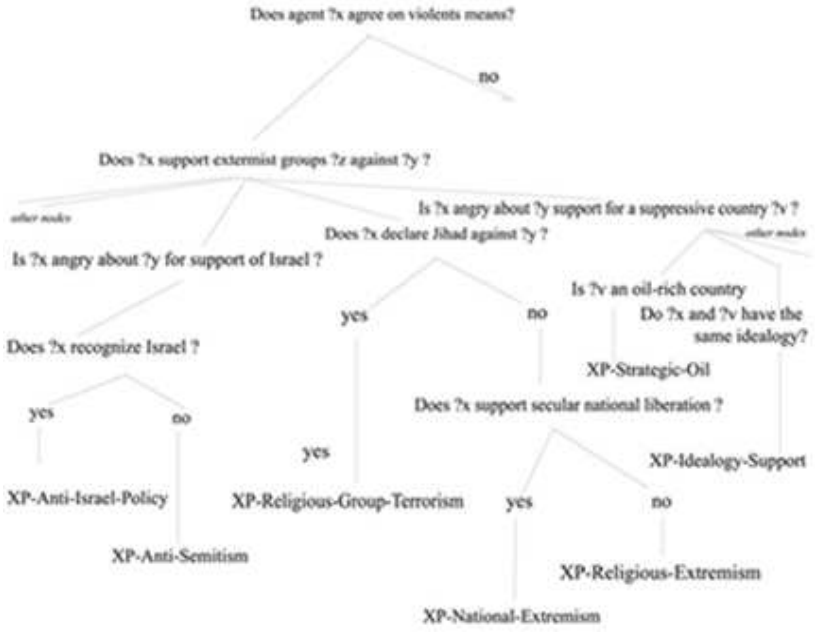

Figure 1: A Portion of the Discrimination Tree Structure

In the second step, the system retrieves the abstract pre-stored explanatory structures (i.e. XPs) indexed under the terminal nodes reached by the traversal process given in the previous step. Each old case could be associated with several explanations which relate particular features or events in the case. The idea is that XP search can look for other cases that share not only the same main features, but also the same causal configuration according to several fanatic points of view. This is a useful learning strategy since it is easier to explain a document as an instance of a previous analysis. Moreover, the retrieved XPs are used by the Naïve Bayes classifier to classify the document into a particular fanatic category which is associated with multiple XPs each of which gives an interpretation of a text according to a particular fanatic aspect (e.g. suicidal, killer, zealous, etc.). This will help an intelligence analyst to pinpoint the important documents. Figure 1 shows a portion of the discrimination tree structure. Finally, a matching and ranking algorithm is applied to select the most similar old cases which are used to generate additional valuable information about the new case (e.g. possible identification of the email sender). The point is that these cases can provide important information about a new document which might have been overlooked by an intelligence analyst. In other words, by having a reference to the most relevant past instances, an intelligence analyst can come to a profound understanding of a new document and possibly avoid previous case-specific miss-judgments.

3.1 Case Classification

Explaining a fanatic situation requires selecting the important features of fanaticism. To retrieve a relevant XP or build a relevant explanation, we need to concentrate on causally-relevant features. Therefore, in order to make sense of an event, an explanation needs to show that the event is actually reasonable, given other reasonable information. This is can be done by showing how the fanatic situation was caused by factors that are already known, or are thought likely. However, evaluating plausibility of facts is potentially an explosive inference problem. One way to circumvent this problem is to give general suggestions about the types of factors that are likely to be relevant to explaining a class of concepts. This strategy helps the understanding model to direct search by pointing to attributes that are often worth considering. For example, a zealous suicide bomber is driven by the belief that some enemy is threatening the existence of Islam, and his mission in life is to protect Islam at any cost. His belief can be inferred from general set of facts even though some information about the perpetrator is missing. The point is that events, in real-world situations, are so complicated that we often lack the detailed models necessary to form a complete predictive explanation, and need to learn from weaker accounts. Also, incomplete explanations may give enough information for understanding a new situation, making the search for all possible facts unnecessary. In this work, we suggest using the Naïve Bayes learner to compensate for the lack of knowledge needed to analyze documents. The basic idea is that the presence of a concept in an explanation structure (i.e. XP) will tend to higher the probability of finding the concept in the category of fanaticism, and vise versa. The Naïve Bayes classifier is used to estimate the probability of each category (i.e. NF, CAF, and CRF) for a given document. The classification is based on the prior probability of a category occurring, and the conditional probabilities of the concepts occurring in a document, assuming independence of concepts. Therefore, the probability of a concept O given class v_j is estimated from the data as follows:

$$P(O|v_j)=\frac{1+\sum_{i=1}^{|T|} n_{io}P(v_j|t_i)+\sum_{k=1}^{|XP|} n_{ko}P(v_j|xp_k)}{|V|+\sum_{x=1}^{|V|}\sum_{i=1}^{|T|} n_{ix}P(v_j|t_i)+\sum_{x=1}^{|V|}\sum_{k=1}^{|XP|} n_{kx}P(v_j|xp_k)}$$

```
(define-case document21
 ({description}
  (features (believe usa policy threaten Islam alqhaeda destroy preserve
   Jihad holy task demand sacrifice money soul will use suicide mean protect))
  (agents
      (agent1 ($fanatic) agent2 (alqhaeda) agent3 (usa)))
  (events
          (state-1 (*policy*-object
            (agent3 value(pro (P-orientation(western-block)))))
               threaten  (G-*preservation*-state(agent2  object(Islam) value(yes))))
          (state-2 (G-*preservation*-state
                 (agent2  object(Islam) value(yes)))
             initiate  (*destruction*-state
                   (agent3 object(*national-security*) value(yes))))
          (state-3 (G-*Jihad*-state
                (agent2  value (yes)))
             initiate  (*sacrifice*-state (agent1 object(money) value(yes))))
           (state-4 (G-*Jihad*-state
                (agent2  value (yes)))
             initiate   (*sacrifice*-state
                   (agent1 object(soul) value(yes))))
           (state-5  (G-*preservation*-state
                   (agent1 object(*islam*) value(yes)))
             initiate  (*suicide*state
                      (agent2 object(agent3) value(yes)))))))
```

Figure 2: Conceptual Representation of a Case

Where $|T|$ denotes the number of documents in the training data, n_{io} is the number of times the concept *O* (i.e. feature) occurs in the text t_i, and $P(v_j|t_i)$ is the conditional property of classification v_j given the text t_i. The term n_{ix} represents the frequency with which the concept *O* occurs in the training data, n_{ko} is the number of times *O* occurs in the explanation patterns. Given the above equation, we can now write the naïve Bayes classifier as

$$\arg\max_{v_j \in \{NF, CRF, CAF\}} P(v_j|t_i) = P(v_j)\ P(O_1, O_2, ..., O_n / v_j)$$

$$= P(v_j) \prod_i P(O_i / v_j)$$

The positive advantage of this approach is that the accuracy of classification is increased on the basis of the statistical weight given to the features in the XPs. This is in turn has the valuable effect of recalling the most appropriate cases in the memory. Also, by using several XPs, the new document is examined from several viewpoints each of which represents an aspect of fanaticism that should be considered by the intelligence analyst.

processing case document21….

applying case-based explanation process to document21
…binding $fanatic to ?x …answering question DOES AGENT $fanatic SUPPORT VIOLANCE?...
binding relevant facts in document21…
…inferring knowledge using domain rules...
inferring CD#11*VIOLANCE*-STATE…
processing phase 1…
checking relevant questions…
patterns found…
binding alqhaeda to ?z…binding usa to ?y…answering question DOES $fanatic SUPPORT EXTREMIST GROUPS alqhaeda AGAINST usa?...
… inferring knowledge using domain rules...
inferring …CD#30*DESTRUCTION*-STATE…CD#21 *JIHAD*-STATE… CD#11 *VIOLANCE*-STATE …CD#30 *SUICIDE*-STATE… CD#9 G-*PRESERVATION*-STATE…
checking relevant questions…
patterns found…
DOES $fanatic declare JIHAD AGAINST usa?
… inferring knowledge using domain rules...
inferring CD#21 *JIHAD*-STATE…CD#19 *SACRIFICE*-STATE… CD#9 G-*PRESERVATION*-STATE…CD#41 *G-LIBERATION*-STATE…CD#11 *VIOLANCE*-STATE…
processing phase 2…
checking relevant explanation patterns...
retrieving XP#9 XP-religious-group-terrorism…
applying naïve bayes classifier…
classifying document21 as CODE-RED FANATICISM
retrieving the first 5 similar cases from memory…
cases found…case#220…case#98…case#74…case#231…case#101

Figure 3: A Sample Run of the System

4 System Performance

In this work, a frame-like structure is used to represent facts in documents. A frame representation consists of a head name, slots, and slot fillers. In general, actions, goals, and the state of the world are represented in a canonical manner in terms of high-level facts such as *violate* and *G-*defeat*-state*. Moreover, the relationship between actions and goals of actors are represented according to a set of causal and motivational links [Kolodner 1993] [Schank et. al 1990] such as *motivate* which denotes a state or event that motivates a goal, *threaten* which denotes an anticipated threatening situation to a desired goal state, permanently or temporarily, and *initiate* which denotes a goal or event that initiates a situation. To illustrate how our process works, suppose the system is introduced to classify and explain the following translated version of an Arabic fanatic document:

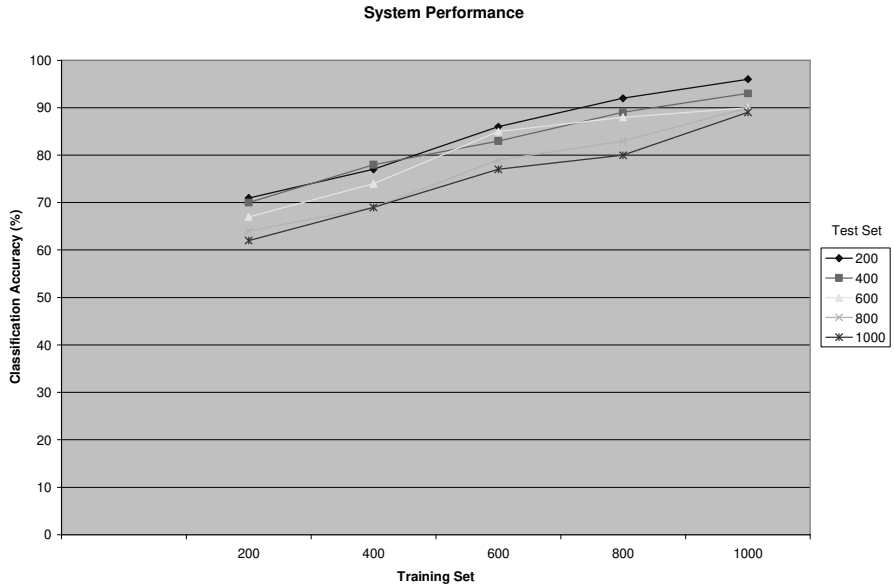

Figure 4: Combining XPs and Naïve Bayes

"believe that USA policy threatens Islam, therefore, Al-Qhaeda must destroy USA in order to preserve Islam. Jihad is a holy task that demands a sacrifice of money and soul. I'm willing to use suicidal means to protect Islam."

The case is conceptually represented as shown in Figure 2. The first step is to apply a built-in inference process that is based on the discrimination network organization. The traversal process of the discrimination net is determined by what questions are asked and which answers are inferred from the set of facts about the case. The traversal process continues until the system reaches a node where an appropriate XP that explains the new document is found. The set of facts inferred is used to bind the variables given in the premises of the retrieved XP. In other words, the branch in the network to traverse is determined by the facts contained in the document, and the general background knowledge of the world (e.g. international context, terrorist groups, Islamic laws, etc.). A sample run of the program is shown in Figure 3**.** In this example, the system retrieves one XP *XP-religious-group-terrorism* which explains a fanaticism viewpoint of the document as:

"a radical Islamist who believes that the practice of Islam has become corrupted and must be reformed. USA represents a threat to Islamic laws and must be defeated through holy war (i.e. Jihad)."

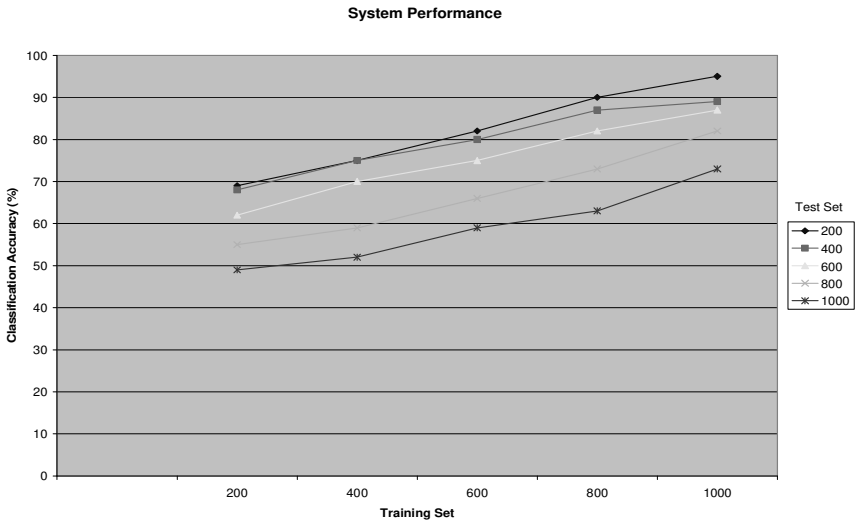

Figure 5: Using Naïve Bayes without XPs

It is worth pointing out that the XP is retrieved on the basis of the information available in the new document. However, if additional information had been given, the system would have investigated other questions and probably retrieving other relevant XPs. Moreover, the system retrieves the best 5 similar cases to the current document. The cases can be used to provide additional valuable information about the new case (e.g. expectations about forthcoming events based on past events mentioned in similar emails, a profile of a fanatic person, analysis of a fanatic mind, etc.).

Our system was developed using Visual LISP for platform Windows NT. LISP is a powerful descriptive language for expressing the events in the world. We report the classification accuracy for using Naïve Bayes classifier with and without explanation patterns. Multiple runs were executed as we vary the number of training and test sets. By looking at Figure 4 and Figure 5, it is obvious that by incorporating the background knowledge into our model, the classification task has produced betterresults for training and test sets of all sizes. For example, when two runs were made for 600 training examples, and 1000 test examples (shown in the legend), the classification accuracy of our scheme was %77 as compared to %55 without using the XPs. In general, the classification accuracy has improved by up to %22 when using XPs. The results verify the fact that the efficacy of background knowledge makes up for the limited training data. Also, it is obvious that the size of training examples represents an essential part of the construction of the model. The figures show that robust classification can be achieved by increasing the size of training set. For example, in Figure 4., the system scored a classification accuracy of %96 when the number of training examples was 1000, and the test set was 200. However, this percentage tends to decline as the size of test data increases and the training data decreases. Finally, since analyzing a case about fanaticism is a highly subjective matter, it was crucial for the analyses to be consistent throughout the data set. Nonetheless, because of the different subjective interpretations specified by domain

experts, a few discrepancies in the analyses can cause misclassifications of the test set.

5 Conclusions and Future work

Some of most sensitive information about terrorism comes from "open sources" such as media broadcasts, newspapers, emails, etc. After Sep. 11, the world community decided to take the initiatives in confronting the terrorist groups who usually do not have any concerns of human life. The main challenge which faces intelligence experts is enhancing information-gathering capabilities, so they could track and eradicate terrorists. This work is characterized by processing Arabic documents in order to extract useful information about fanatic content. The aim here is to take the CBR one step towards the application level in an ill-structured domain such as fanaticism. To achieve this, challenging issues related to integrating memory indexing, retrieval, and classification have been investigated throughout this work. The emphasis on these issues stems from the fact that using old fanatic cases to guide intelligence analysis substantially depends on the efficient retrieval of useful cases.

This work is a step forward in attempt to design a decision-support system in the domain of fanaticism in order to provide the intelligence operators with new ways of reasoning. The analyses are used as a platform for retrieving and classifying old fanatic cases. The retrieved cases are can be used to provide the intelligence operator with predictions of possible events, a possible identification of a terrorist, a location of a terrorist group, etc. Moreover, they can be used to pinpoint the major factors enticing terrorist activities for which an intelligence analyst may have overlooked.

In this work, a hardwired inference process based on a discrimination network of multiple-questions has been devised to select those relevant facts from possibly large amount of information about *fanaticism*. It is also used to direct the system towards investigating several aspects of fanaticism may otherwise be overlooked by intelligence analysts. Our cases are indexed on the basis of several explanatory knowledge structures (i.e. XPs) each of which explains different attitude of fanaticism. The CBR model is used to analyze the documents from several aspects in order to extract interesting patterns of knowledge. Finally, the classification process is improved by computing the probabilities of relevant features given in the causal structures (i.e. XPs). The output of the classifier pinpoints the most crucial fanatic documents (i.e. Code-Red Fanaticism) which may contain information that would be critical to the intelligence community. The results have shown the utilization of background knowledge yields better classification rates.

This work is a first attempt to implement a problem understanding model in the domain of fanaticism. There are several challenging issues need to be addressed. For example, the inconsistency of user's input specifications with domain terms and rules which are used to represent knowledge should be investigated further. The scenarios of fanaticism implemented are fairly limited and further work should be done to handle the cognitive and emotional propositions involved in analyzing a fanatic situation.

References

[Aamodt et. al. 1994] Aamodt, A. and Plaza E.: "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System"; *AI Communications, The European Journal of Artificial Intelligence*, (1994), pp.39-59.

[Almonayyes 1997], Almonayyes A.: "A multi-level Indexing Scheme for Retrieving Cases of Multiple Points of View"; *Fifth German Workshop on Case-Based Reasoning - Foundations, Systems, and Applications*, (1997), Bonn, Germany.

[Almonayyes 2001] Almonayyes A. and Hassanein H.: "Application of Case-Based Reasoning for Call Admission Control in ATM Networks"; *The Journal of Experimental and Theoretical Artificial Intelligence*, (2001).

[Cohen 1996] Cohen, W.; "Learning Trees and Rules with set-valued features"; Proceedings of the National Conference on Artificial Intelligence; AAAI Press, (1996), pp249-270.

[Hastie et. al. 2001] Hastie, T., Tibshirani, R. and Friedman, J. H.: "The elements of Statistical Learning"; Data Mining, Inference, and Prediction; *Springer Series in Statistics*, Springer Verlag (2001).

[Jackson et. al. 2002] Jackson, P. and Moulinier, I.: "Natural Language Processing of Online Applications"; Text Retrieval, Extraction, and Categorization; *Natural Language Processing*, 5, John Benjamins (2002).

[Kolonder 1993] Kolonder J. L, "Case Based Reasoning, San Mateo"; Morgan Kufman, (1993).

[Kolodner 1996] Kolodner, J.: "Making the Implicit Explicit"; Clarifying the Principles of Case-Based Reasoning, Lessons and Future Directions; *MIT Press,* (1996), pp349-370.

[Lewis 1994] Lewis D.: "A comparison of two learning algorithms for text categorization"; Third Annual Symposium on Document Analysis and Information Retrieval; (1994), pp81-93, Las Vegas, Nevada, USA.

[Manning et. al 2001] Manning, C. D. and Schutze, H.: "Foundations of Statistical Natural Language Processing ", (2001) *MIT Press*.

[McCallum et. al 1998] McCallum A. and Nigam K.: "A Comparison of Event Models for Naïve Bayes text Classification"; AAAI-98 Workshop on Learning for Text Categorization, (1998).

[Mitchell 1997] Mitchell T.: "Machine Learning"; *McGraw Hill*, (1997).

[Schank et. al 1990] Schank, R., and Osgood, R. 1990: "A Content Theory of Memory Indexing"; Northwestern University, Institute of Learning Sciences, Technical Report no. 2, (1990).

[Zelikovitz et. al 2000] Zelikovitz K. and Hirsh H.: "Improving Short Text classification Using Unlabeled Background Knowledge to Assess Document Similiraty"; *Proceedings of the Seventeenth International Conference on Machine Learning;* (2000), Pp. 1183-1190, San Francisco: Morgan Kaufman Publishers.