Knowledge Representation and Logic for Beliefs

Abstract

This paper presents Belief Augmented Frames, or BAFs. A BAF represents a concept or item in the world, and slot-value pairs represent relations between BAFs. Each BAF is assigned two belief masses. The Supporting Mass represents the degree in which the evidence supports the existence of the concept or object represented by the BAF. The Refuting Mass represents the degree in which the evidence refutes the existence of the concept or object. The novelty of BAFs comes from the independence between the Supporting and Refuting masses. A logic system called BAF-Logic based on fuzzy-logic style min-max functions and Predicate Logic is also introduced to perform reasoning on BAF and their relations. An example application of BAFs to the text classification problem is given.

Introduction

Traditionally uncertainty is represented by probabilities, where an event or fact has a probability of *P* of being happening, or of being true, and a probability of *1-P* of not happening, or not being true. This approach has its weaknesses. Probabilities are unable to model ignorance, and this leads to difficulties. (Shortliffe and Buchanan 1985) give an example where doctors assigning a probability of *P* to a patient having a disease, are reluctant to assign a probability of *1-P* to a patient not having the disease.

Belief measures address this, where degrees of belief are modeled by a range of values rather than a single point. (Dubois, Prade and Smets 1996) presents an excellent argument why beliefs should not be represented by just a single point-probability.

This paper presents the concept of Belief Augmented Frames, or BAFs. Frames are a powerful method of representing knowledge in AI, providing structure and operations that allow us to model an agent's world effectively. In this paper we enhance AI frames with belief measures. This introduces uncertainty into the frame slot-value pairs (and consequently, on the relationships between frames), and allows us to model ignorance. We then present a reasoning system that performs non-monotonic reasoning on relationships between frames using the belief measures, and an example application of BAFs to the text classification problem.

Related Work

In 1967 Dempster modeled uncertainty with a range of probabilities (Dempster 1967) rather than a single number. Shafer extended this in 1976 (Shafer 1976)], producing what is known today as Dempster-Schafer Theory (DST). In DST the environment is assumed to be a fixed set of mutually exclusive elements, symbolized by Θ. Dempster's Rule of Combination is applied to combine new evidence scores into existing evidence. Belief scores are computed from these evidence scores, together with plausibility and ignorance scores. Smets introduces the Transferable Belief Model (TBM) in (Smets 2000). The TBM may be viewed as a generalization of DST. Where DST takes a closed-world assumption, TBM assumes an open world, and as such the belief mass of the empty set may be non-zero.

Much work has been done on formulating new belief reasoning formalisms. (Boeva, Tsiporkova and De Baets 1999) extends classical modal logic with plausibility and belief measures. Modal logic is an extension of Propositional Logic, and consists of a set of *possible* worlds, a binary relation between worlds called an *accessibility function*, and an *assignment function* that assigns truth values to atomic propositions about each possible world. Boeva et al treat the accessibility function as a multi-valued function, thus inducing plausibility and belief measures on this function on each of the possible worlds. The inverse of the assignment function is also viewed as the second multi-valued function, inducing plausibility and belief measures on the propositions of each possible world.

(Koller and Halpern 1992) proposes two new types of entailments to reason with imprecise information. A *cautious entailment* allows only completely justified conclusions. For example, if we know that "John is 1.88 meters tall", and later obtain a contradictory piece of evidence that "John is about 1.90 meters tall", then a cautious entailment allows us to conclude that John is any height between 1.88 and 1.90 meters tall. I.e. a cautious entailment allows us to conclude any value between two contradictory values. A *bold entailment* on the other hand allows us to conclude that "John is approximately *h* meters tall" for any *h* between 1.88 and 1.90 meters tall. Thus we might "guestimate" that John is 1.92 meters tall. The authors present theorems to investigate the properties of their logic system.

(Parsons and Kubat 1994) propose a symbolic reasoning system based on rough sets. Details of rough sets may be found in (Pawlak 1984) and (Pawlak et al 1988). Briefly, the authors define logical relations in terms of operations on rough sets. Objects of interest are organized into rough sets, and the logical relations are rendered to rough set operations that manipulate members of the set to perform reasoning. A proposition *p* is determined to be *true, roughly true, unknown, roughly false and false* based on the set membership after the set operations corresponding to the logical operations in *p* are performed on the rough set. A *min-max* approach similar to that used in fuzzy logic is used to combine these "rough values" to produce the final outcome.

Finally, (Haenni, Kohlas and Lehmann 1999) proposes a framework for unifying Dempster-Schafer type reasoning systems (including the Transferable Belief Model) and Probabilistic Argumentation Systems. They argue that PAS provides for a powerful modeling language that will work on top of DST, and that DST forms an efficient computational tool for PAS, and provide rules for "interfacing" the two systems.

Belief Augmented Frames

For convenience, the Belief Augmented Frames are assumed to be within an agent called "You", and will be described with reference to this agent.

Definitions

Definition 1 A Belief Augmented Frame Knowledge Base (BAF-KB, or simply KB) is defined to be a set of concepts C. Informally, a concept $c_i \in C$ corresponds to an idea or a concrete object in the world. For example, "train", "orange", "car" and "sneeze" are all valid concepts in the BAF-KB. Since all objects and concepts are abstracted into ideas in Your "mind", this work will not differentiate between a tangible object (e.g. a car) versus an abstract idea (e.g. the color blue). The words "object" and "concept" will be used interchangeably.

Definition 2 A Supporting Belief Mass (or just simply "Supporting Mass") φ^T measures how much we believe in the existence of a concept or a relation between concepts is true. A Refuting Belief Mass ("Refuting Mass") φ^F measures how much believe that a concept does not exist, or a relation between two concepts is untrue. In general, $0 \le \varphi^T, \varphi^F \le 1$, and $\varphi^T + \varphi^F \ne 1$. The last condition is in fact the reason why we have both a supporting and a refuting belief mass; this allows us to eliminate the constraint that $\varphi^F = 1 - \varphi^T$. The Supporting and Refuting Belief Masses for the existence of a concept c_i is denoted as φ^T_i and φ^F_i respectively, and for the *kth* relation between concept c_i and c_j they are denoted as φ^T_{ijk} and φ^F_{ijk} respectively.

Note that by this definition, it is possible that $\varphi^T_i + \varphi^F_i > 1$ *and* $\varphi^T_{ijk} + \varphi^F_{ijk} > 1$. The Utility Function U_i and U_{ijk} may be used to re-map the combined masses to the range $[0, 1]$ if this is desired.

Definition 3 A concept $c_i \in C$ is defined as a 4-tuple $(cl_i, \varphi^T_i, \varphi^F_i, AV_i)$, where cl_i is the name of the concept, φ^T_i is our supporting belief mass that this concept exist, φ^F_i is our refuting belief mass. AV_i is a set of relations relating c_i with some $c_j \in C$. Note that there is no restriction that $i \ne j$, so a concept may be related with itself.

Definition 4 A relation $av_{ijk} \in AV_i$ is the k^{th} relation between a concept c_i to a concept c_j. A relation av_{ijk} consists of a 4-tuple $(al_{ijk}, cd_j, \varphi^T_{ijk}, \varphi^F_{ijk})$, where al_{ijk} is the name of the k^{th} ($k \ge 1$) relation between c_i and c_j, cd_j is the label for c_j, φ^T_{ijk} is our supporting belief mass that the k^{th} relationship between c_i and c_j is true, while φ^F_{ijk} is our refuting belief mass.

Definition 5 The Degree of Inclination DI_i for the existence of a concept c_i and DI_{ijk} for the *kth* relation between concepts c_i and c_j is defined as the difference between the supporting and refuting belief masses:

$$DI_i = \varphi^T_i - \varphi^F_i \quad (1a)$$

$$DI_{ijk} = \varphi^T_{ijk} - \varphi^F_{ijk} \quad (1b)$$

For convenience we use the notation DI when it is immaterial whether we are referring to DI_i or DI_{ijk}. DI measures the truth or falsehood of a statement, as is bounded by $[-1, 1]$.

DI gives us a convenient way to detect conflicting facts. Suppose we have a fact P (a *fact* might refer to the existence of a concept, or the existence of a relation between concepts) with degree of inclination DI_P. Suppose we re-evaluate P and obtain $DI_{P'}$. The facts are contradictory if $DI_P \, . \, DI_{P'} < 0$, since they give opposing truth values after re-evaluation of P.

Definition 6 The Utility Function U_i and U_{ijk} is defined as:

$$U_i = \frac{1 + DI_i}{2} \quad (2a)$$

$$U_{ijk} = \frac{1 + DI_{ijk}}{2} \quad (2b)$$

For notational convenience we will use U to refer to either U_i or U_{ijk}. U shifts the range of DI from $[-1, 1]$ to $[0, 1]$ to allow φ^T and φ^F to be used as a utility function (hence its name) for decision making.

Definition 7 The Evidential Conflict EC_i or EC_{ijk} is defined as:

$$EC_i = 1 - |DI_i| \quad (3a)$$

$$EC_{ijk} = 1 - |DI_{ijk}| \quad (3b)$$

The term EC is used to refer to either EC_i or EC_{ijk}, when the context is unimportant. If EC is large, this means that φ^T and φ^F are very close in value. I.e. the evidence provided is conflicting and equally supports or refutes a fact. EC therefore measures the amount of conflict in the supporting and refuting evidence.

EC_i is the evidential conflict in the existence of a concept or object i, and EC_{ijk} is the corresponding measure for the *kth* relation between a concept i and another concept j.

Definition 8 The plausibility of the existence of a concept or a relation given the refuting belief mass is given by:

$$Pl_i = 1 - \varphi^F_i \quad (4a)$$

$$Pl_{ijk} = 1 - \varphi^F_{ijk} \quad (4b)$$

Definition 9 The Evidential Interval EI_i and EI_{ijk} are given by:

$$EI_i = [\varphi^T_i, Pl_i] \quad (5a)$$

$$EI_{ij} = [\varphi^T_{ijk}, Pl_{ijk}] \quad (5b)$$

Definition 10 The Ignorance Ig_i and Ig_{ijk} are given by:

$$Ig_i = Pl_i - \varphi^T{}_i \quad (6a)$$
$$Ig_{ijk} = Pl_{ijk} - \varphi^T{}_{ijk} \quad (6b)$$

Together Definitions 7 to 10 measure the quality of the evidence supporting and refuting the existence of a concept or of a relation. Table 1 gives the interpretation of the evidential interval *EI*:

Evidential Interval	**Interpretation**
[0, 0]	The evidence provided completely refutes the fact.
[1, 1]	The evidence provided completely supports the fact.
$[\varphi^T, Pl]$ $0 < \varphi^T, Pl < 1$ $Pl \geq \varphi^T$	The evidence both supports and refutes the fact.
$[\varphi^T, Pl]$ $0 < \varphi^T, Pl < 1$ $Pl < \varphi^T$	The evidence supporting the fact exceeds the plausibility of the fact. I.e. the evidence is contradictory.

Table 1 Interpretation for Evidential Interval *EI*

Definition 11 If You are unaware of the existence of a concept c_i or a relation av_{ijk}, or if You have no information on the reliability of this information, You will assign a default support belief mass of $\varphi_D{}^T$ and $\varphi_D{}^F$ to this information. In general $\varphi_D{}^T$ and $\varphi_D{}^F$ should be small to reflect Your lack of confidence in the relation or the existence of the concept, and $\varphi_D{}^T \approx \varphi_D{}^F$ to give us a small DI_D, reflecting Your ignorance.

Your ignorance in BAF can come in one of two ways; either both the supporting and refuting belief masses are equally strong resulting in $DI = 0$, or You are genuinely unaware of this concept or object. The default belief masses are useful in modeling the latter and makes allowances for an open-world assumption. I.e. whenever You encounter an object, concept or relation that you are unaware of, assign it a supporting belief mass of $\varphi^T{}_D$ and a refuting belief mass of $\varphi^F{}_D$.

Combining Belief Masses

We now define how both supporting and refuting belief masses may be combined, and in the following section we will formally define the rules of combination.

To simplify notation, we will use single-letter propositional symbols like *P* and *Q* to represent the fact that a concept c_i exists, or a relation av_{ijk} exists between concepts c_i and c_j. We write $(P, \varphi^T{}_P, \varphi^F{}_P)$ to represent a proposition *P* with supporting belief mass $\varphi^T{}_P$ and refuting belief mass $\varphi^F{}_P$.

Note that while we use propositional logic style symbols like *P* and *Q* to express symbols, BAF-Logic (as defined later) is a first-order logic system. Clauses are defined over entire classes of objects instead of for individual objects.

Definition 12 Given a proposition $(P, \varphi^T{}_P, \varphi^F{}_P)$ and given a proposition $(Q, \varphi^T{}_Q, \varphi^F{}_Q)$, we define:

$$\varphi^T{}_{P\wedge Q} = min(\varphi^T{}_P, \varphi^T{}_Q) \quad (7)$$

Intuitively, this states that since both *P* and *Q* must be true for $P \wedge Q$ to be true, Your knowledge of $P \wedge Q$ being true will only be as good as your most unreliable piece of evidence supporting $P \wedge Q$.

Definition 13 Continuing with propositions *P* and *Q* above, we define:

$$\varphi^T{}_{P\vee Q} = max(\varphi^T{}_P, \varphi^T{}_Q) \quad (8)$$

Again this states that $P \vee Q$ is true when either *P* is true or *Q* is true, You are willing to invest as much confidence in $P \vee Q$ as the strongest piece of evidence supporting $P \vee Q$.

Having defined both $\varphi^T{}_{P\vee Q}$ and $\varphi^T{}_{P\wedge Q}$, deriving the corresponding $\varphi^F{}_{P\vee Q}$ and $\varphi^F{}_{P\wedge Q}$ involves using De-Morgan's theorem.

$$\neg (P \wedge Q) = \neg P \vee \neg Q \Rightarrow \varphi^F{}_{P\wedge Q} = max(\varphi^F{}_P, \varphi^F{}_Q) \quad (9a)$$
$$\neg (P \vee Q) = \neg P \wedge \neg Q \Rightarrow \varphi^F{}_{P\vee Q} = min(\varphi^F{}_P, \varphi^F{}_Q) \quad (9b)$$

Definitions 12 and 13 are similar to the conjunction and disjunction rules used in Probabilistic Argumentation Systems (Picard 2000). However You use the *min* function instead of multiply to combine belief masses in conjunctions, and You use the *max* function in disjunctions instead of add. This has three advantages:

The idea of choosing a *min* and *max* functions to combine belief masses in a conjunction and a disjunction respectively has an intuitive basis as proposed in Definitions 12 and 13.

We avoid the problem of the belief mass of a long series of conjunctions decaying to 0. Likewise, given a disjunction over many propositions P_i each with very small belief masses, we can avoid the situation of placing absolute belief $(\varphi^T{}_{Pi} = 1)$ on a proposition even though the propositions in the disjunction themselves have tiny belief masses.

Additionally, the use of *min(.)* and *max(.)* functions eases defining the relationship between the supporting and refuting belief masses. If the supporting mass is a *min(.)* function, then the refuting mass is a *max(.)* function and vice-versa, as shown in Definition 13 above. This is more intuitive than to say that if the supporting mass is a multiply function, then the refuting mass should be an add function.

Definition 14 By definition, the supporting belief mass $\varphi_P{}^T$ is a measure of how confident we are that a proposition *P* is true, while the refuting belief mass $\varphi_P{}^F$ is a measure of how confident we are that the proposition is not true. We can define the logical negation ¬P as:

$$\varphi^T_{\neg P} = \varphi^F_P \quad (10a)$$
$$\varphi^F_{\neg P} = \varphi^T_P \quad (10b)$$

Definitions 12 to 14 allow us to compute the supporting and refuting belief masses for any propositional logic statement.

There is also a series of operations that are defined on BAFs to create new frames, set relations between frames and abstract information from a set of frames. Due to space constraints these operations will be omitted from this paper.

Applying BAFs to Text Categorization

In the text classification problem text documents (typically news articles) are automatically sorted and classified into several different categories. This is useful when labeling and dealing with very large text collections such as in a library or an archive. Extensive research has been done on text classification, e.g. in (Nigam et al 2000). In this section we will compare the performance of a BAF-Logic based classifier against a Naïve Bayes classifier and a Probabilistic Argumentation Systems classifer.

Formulating the Text Classification Problem

Keyword Selection A stop-list is used to remove spurious words from the list of words extracted from a documents. The words are then stemmed using a Porter Stemmer. Stemmed words occurring fewer than τ times (τ nominally set to 3) are removed to produced the final set of keyword terms t_{ijk}, which is the *kth* keyword term of the *jth* document of the *ith* class.

Naïve Bayes Classifier A Naïve Bayes classifier is used to form the baseline performance against which the BAF-Logic classifier is measured against.

BAF-Logic In BAF-Logic the *jth* document D_{ij} in the document class c_i is taken to be a conjunction of terms t_k:

$$D_{ij} = t_{ij0} \wedge t_{ij1} \wedge \ldots \wedge t_{ij(n-1)} \quad (11)$$

Each term and document is related by a set of relation

$$R_{ijk} = \{(D_{ij}, term, t_k, \varphi^T_{ijk}, \varphi^F_{ijk}) \mid t_k \text{ is a term in } D_{ij}\}. \quad (12)$$

In addition, given a set of documents D in class c_i, we apply an *abstract* operation, which extracts keywords occurring in a minimum number of documents in a class, on all R_{ijk} to produce a set of relations characterizing the class c_i. We call this set of vectors the *characteristic vector* v_i of class i.

$$v_i = (S_{i0}, S_{i1}, S_{i2}, \ldots, S_{i(m-1)}) \quad (13)$$

Here each S_{ik} is the relation

$$S_{ik} = \{(c_i, term, t_k, \varphi^T_{ik}, \varphi^F_{ik}) \mid t_k \text{ occurs in at least } \alpha\% \text{ of documents } D_i \text{ in class } c_i\} \quad (14)$$
$$\varphi^T_{ik} = min_j\, \varphi^T_{ijk} \quad (15)$$
$$\varphi^F_{ik} = max_l max_j\, \varphi^T_{ljk},\ l \neq i \quad (16)$$

φ^T_{ik} is the mass that supports the fact that term t_k implies class c_i, while φ^F_{ik} supports the fact that term t_k implies some other class c_l. I.e. it refutes the fact that the term t_k implies the class c_i.

S_{ik} then represents the belief that term t_k implies the class c_i. To classify an unseen document D_u, we derive the keyword terms $t_{unk,\,k}$. We can derive the following masses that support and refute D_{unk} belonging to class c_i:

$$\varphi^T_{i,\,unk} = min(\varphi^T_{i0}, \varphi^T_{i1}, \varphi^T_{i2}, \ldots, max(\varphi^F_{i0}, \varphi^F_{i0}, \ldots)) \quad (17)$$
$$\varphi^F_{i,unk} = max(\varphi^F_{i0}, \varphi^F_{i1}, \varphi^F_{i2}, \ldots, min(\varphi^T_{i0}, \varphi^T_{i0}, \ldots)) \quad (18)$$

The degree of inclination $D_{i,unk}$ is given by its definition:

$$DI_{i,\,unk} = \varphi^T_{i,\,unk} - \varphi^F_{i,\,unk} \quad (19)$$

A document is classified in class *win* by maximizing $DI_{i,unk}$:

$$win = argmax_i\, DI_{i,\,unk} \quad (20)$$

Probabilistic Argumentation Systems In PAS we again represent a document D_{ij} in class i with a conjunction of terms.

$$D_{ij} = \wedge_k\, t_{ijk} \quad (21)$$

Where the term t_{ijk} is the kth keyword term in document D_{ij} An identical abstraction operation is performed on a set of documents to produce a characteristic vector v_i for every class c_i.

To classify a document D_{unk} consisting of keyword terms $t_{unk,k}$, we derive the following argument:

> A term t_k supports that the document D_{unk} belongs to class c_i if it occurs in abstraction vector v_i and not in v_j, $j \neq i$.

To put it in Prepositional Argumentation System form:

$$qs(D_k \text{ in class } i) = t_{i0} \wedge t_{i1} \wedge \ldots t_{i,n-1} \wedge \neg (t_{j0} \wedge t_{j1} \wedge \ldots t_{j,,n-1}),\ l \neq i \quad (22)$$

Deriving the degree of support:

$$p(qs(D_k \text{ in class } i)) = \Pi_k\, p(t_{ik}) \Sigma_k\, (1.0 - p(t_{lk})) \quad k \neq i \quad (23)$$

Since it isn't contradictory for a term to appear in several classes, the most sensible value for $p(qs(\perp))$ would be 0. Using this we obtain $dsp(qs(D_k \text{ in class } i)$ as:

$$\frac{p(qs(D_k inclassi)) - p(qs(\perp))}{1 - p(qs(\perp))} \quad (24)$$

$$= p(qs(D_k \text{ in class } i))$$

The document belongs in the class with the largest degree of support.

The percentage α in equation 14 is called the *abstraction degree* and represents how strictly we would want to determine whether a relation characterizes the class the document occurs in. It also controls the size of the reference lexicon. Larger α implies that the term must occur in more files, and hence there will be fewer terms in the lexicon.

Having modeled the classification problem as a Naïve Bayesian Classifier, BAF-Logic and PAS model, we proceeded to evaluate the classification performance under each model.

Experiment Results

The performance of the Naïve Bayes, BAF-Logic and PAS classifiers were evaluated on a large classification task consisting of 20,000 Newsnet articles from 20 News Groups. The classifiers were trained on 19,800 Newsnet articles divided gleaned from 20 News Groups. The classifiers were then evaluated based on the 19,800 training articles (*inside test)* and the 200 unseen articles (*outside test)*. Details of these tasks are presented in the following sections.

In addition we evaluate the performance of the classifiers under various degrees of abstraction and with Jeffreys-Perks (JPerks) smoothing.

Jeffreys-Perks smoothing is similar to Add One smoothing, but gives slightly better performance for the Naïve Bayes Classifier. However both Jeffreys-Perks and Add One give identical results under BAF-Logic and PAS.

The articles in each newsgroup are typically (though not always) unmoderated, possibly carrying irrelevant materials, and materials written with a wide range of English proficiency skills, vocabulary and writing styles. This makes it difficult to classify a document correctly within the 20 news groups.

In this task we consider abstraction degrees of up to just 20%. I.e. a term must occur at least τ times in at least *20%* of all documents to be considered. At higher levels of abstraction too few terms are left in the lexicon to produce good classification decisions. This supports our view of the irregular nature of Newsnet articles as stated earlier.

Figure 1 below compares the performance of the three classifiers on this task:

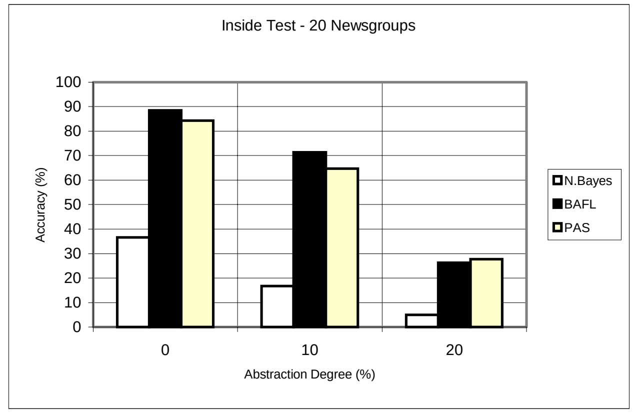

Figure 1 Inside Test Results – 20 Newsgroups

The BAF-Logic Classifier produces the best results at *0%* and *10%* degrees of abstraction, with PAS performing slightly worse. In all three cases the Naïve Bayes Classifier performs poorly.

Figure 2 shows the Outside Test results for the 20 Newsgroups task:

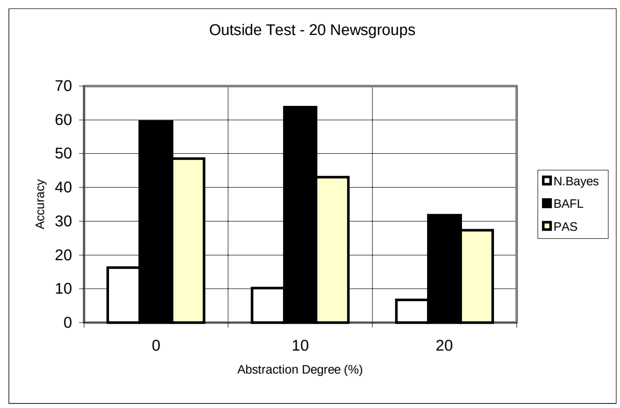

Figure 2 Outside Test Results – 20 Newsgroups

Here the BAF-Logic Classifier performs significantly better than either the Naïve Bayes or the PAS Classifiers. The BAF-Logic Classifier actually performed slightly better when more terms were taken away. PAS and Naïve Bayes however both suffered when more terms were removed from the lexicon through a higher abstraction degree.

Analysis

In both tests the performance of the BAF-Logic classifier was similar to that of the PAS classifier, but much better than the Naïve Bayes classifier. Its likely that both the PAS and BAF-Logic classifiers utilized arguments both for and against classifying a document in a particular class to provide for better classification decisions, whereas the Naïve Bayes classifier only utilized scores supporting that a document belongs to a particular class.

It is interesting to note that the BAF-Logic classifier works particularly well with unseen data, consistently outperforming both the Naïve Bayes and PAS classifiers. In addition the BAF-Logic classifier's accuracy remains stable even as the lexicon shrinks under higher degrees of abstraction. In terms of formulation the PAS and BAF-Logic classifiers are almost identical except for the way the classes were scored, and we feel that the good results relative to PAS demonstrates the usefulness and power of explicitly separating the supporting and refuting (φ^T and φ^F) masses and allowing them full independence. This allows us to set φ^T_{Sik} and φ^F_{Sik} to the probability that the term t_k belongs to class c_i and the largest probability that it belongs to another class c_j $(i \neq j)$ respectively. The results suggest that this is a useful strategy especially for classifying unseen data.

Conclusion

In this paper we introduced the concept of a Belief Augmented Frame or BAF together with a reasoning system called BAF-Logic. We then studied the application of BAF-Logic to the text classification problem.

Our experiment results support the view that considering both evidences for and against a document belonging to a particular class gives better differentiation between classes and thus better classification accuracy.

BAF-Logic goes one step further than PAS by declaring that the masses that support a fact and the masses that refute it are completely independent and may be drawn from different sources. Our experiment results suggest that this strategy provides for even better classification results, in particular for previously unseen documents.

Our experiment results in this paper are very promising. More work should be done to compare BAF-Logic with other approaches like clustering, neural network classifiers, expectation maximization etc. In addition a detailed study of why the BAF-Logic classifier should be particularly good at classifying previously unseen documents should be carried out, especially in relation to the fully independent masses supporting and refuting the presence of a term belonging to the class.

References

E.H. Shortliffe E.G., Buchanan B.G., 1985, "A Model of Inexact Reasoning in Medicine", *Rule Based Expert Systems*, 232-262.

Dubois D., Prade H., Smets P., 1996, "Representing Partial Ignorance", *IEEE Systems, Man and Cybernetics*.

Dempster A. P., 1967, "Upper and Lower Probabilities induced by a Multivalued Mapping", *Annals of Mathematical Statistics*, 38:325-339.

Shafer G., 1976, *A Mathematical Theory of Evidence*, Princeton University Press.

Smets P., 2000, "Belief Functions and the Transferable Belief Model", http://ippserv.rug.ac.be/documentation/belief/belief.pdf,.

Boeva V., Tsiporkova E., De Baets B., 1999, "Plausibility and Belief Measures Induced by Kripke Models", *1st International Symposium on Imprecise Probabilities and their Applications*, Ghent, Belgium, 1999.

Koller D., Halpern J. Y., 1992, "A Logic for Approximate Reasoning", *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, pp. 153-164.

Parsons S., Kubat M., 1994, "A First Order Logic for Reasoning under Uncertainty using Rough Sets", Journal of Intelligent Manufacturing, 5:211-223.

Pawlak Z., 1984, "Rough Classification", *International Journal of Man-Machine Studies*, 20:469-483.

Pawlak Z., Slowinski K., Slowinski R., "Rough Classification of Patients After Highly Selective Vagotomy for Duodenal Ulcer", *International Journal of Man-Machine Studies*, 29:81-95.

Haenni R., Kohlas J., Lehmann N., 1999, "Probabilistic Argumentation Systems".

Picard J., 2000, "Probabilistic Argumentation Systems Applied to Information Retrieval", Doctoral Thesis, Universite de Neuchatel, Suisse,.

Nigam K., McCallum A., Thrun S., Mitchell T., 2000, "Text Classification from Labeled and Unlabeled Documents using EM", *Machine Learning*, 39(2/3):103-134.