Subjective Probabilistic Knowledge Grading and Comprehension

A.Suresh Babu asureshjntu@gmail.com
Asst. Professor
Department of Computer
Science Engineering JNTUA,
Anantapur

Dr P.Premchand p.premchand@uceou.edu
Professor, Department of Computer
Science Engineering Osmania
University Hyderabad

Dr A.Govardhan govardhan_cse@yahoo.co.in
Professor, Department of Computer
Science Engineering
JNTUHCEJ, Jagityala

Abstract

Probabilistic Comprehension and Modeling is one of the newest areas in information extraction, text linguistics. Though much of the research vested in linguistics and information extraction is probabilistic, the importance is disappeared in 80's. This is just because of the input language is noisy, ambiguous and segmented. Probability theory is certainly normative for solving the problems related to uncertainty. Perhaps human language processing is simply non-optimal, non-rational process. Subjective Probabilistic approach fixes this problem, through scenario, evidence and hypothesis.

Keywords: Probability & Statistics, Probabilistic Comprehension, Subjective Grading, Subjective Probability.

1. INTRODUCTION

It is not possible to stop grading which is an objective standardized measure for valuating the texts. Subjective and objective measures declare the values in the texts. Subjective grading is percussive than objective grading, as the previous researches prove to be mechanistic towards grading the knowledge when objective measure are used.

One of the central problems in the field of knowledge extraction or discovery is the availability and the development methods to determine good measures of interestingness of discovered patterns. Such measures of interestingness are divided into objective measures - those that depend only on the structure of a pattern and the underlying data used in the discovery process, and the subjective measures - those that also depend on the class of users who examine the pattern. Objectiveness describes absolute grading of knowledge, whereas subjectiveness describes relative grading. Probabilistic belief is the most important concern that paves out the clarity in selection of measures.

1.1 Important Definitions

Objective – is a statement that is completely unbiased. It is not touched by the speaker's previous experiences or tastes. It is verifiable by looking up facts or performing mathematical calculations.

Subjective – is a statement that has been colored by the character of the speaker or writer. It often has a basis in reality, but reflects the perspective through with the speaker views reality. It cannot be verified using concrete facts and figures.

1.2 When to Be Objective and Subjective

Objective – it is important to be objective when you are making any kind of a rational decision. It might involve purchasing something or deciding which job offer to take. You should also be objective when you are reading, especially news sources. Being objective when you are meeting and having discussions with new people helps you to keep your concentration focused on your goal, rather than on any emotions your meeting might trigger. In essence, accomplishing the goal without any distractions is objective oriented.

Subjective – can be used when nothing tangible is at stake. When you are watching a movie or reading a book for pleasure, being subjective and getting caught up in the world of the characters makes your experience more enjoyable. If you are discussing any type of art, you have to keep in mind that everyone's opinions on a particular piece are subjective. In essence, accomplishing the goal with the complete knowledge and the supplements are not distractions rather support enriching the knowledge to fulfill the goal.

Natural language degree expressions denote degrees or relations between degrees and norms of expectation that lie on a scale [3]. What has not been clarified yet is what kind of relations is expressed by degree expressions. Degree expressions are not only a means of denoting graded properties, but they also allow for the adaptation to varying precision requirements as well as for very efficient communication by referring to entities that are only available implicitly or derivable from the context of the phrase.

1.3 Basis

The Dempster-Shafer theory, also known as the theory of belief functions, is a generalization of the Bayesian theory of subjective probability. Whereas the Bayesian theory requires probabilities for each question of interest, belief functions allow us to base degrees of belief for one question on probabilities for a related question. These degrees of belief may or may not have the mathematical properties of probabilities; how much they differ from probabilities will depend on how closely the two questions are related [4].

The Dempster-Shafer theory is based on two ideas: the idea of obtaining degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence.

Implementing the Dempster-Shafer theory in a specific problem generally involves solving two related problems. First, we must sort the uncertainties in the problem into a priori independent items of evidence. Second, we must carry out Dempster's rule computationally. These two problems and their solutions are closely related. Sorting the uncertainties into independent items leads to a structure involving items of evidence that bear on different but related questions, and this structure can be used to make computations feasible.

2. RELATED WORK

2.1 Dempster-Shafer theory

The method of reasoning with uncertain information known as Dempster-Shafer theory arose from the reinterpretation and development of work of Arthur Dempster and by Glenn Shafer in his book *a mathematical theory of evidence*, and further publications. Dempster-Shafer theory is a belief system that deals with the evidence available for a hypothesis on uncertainty. The uncertainty is represented as $Bel(\mathbf{U})$, the belief which is an evidence for hypothesis, $Plaus(\mathbf{U})$, the plausibility which is an evidence that does not contradict the hypothesis.

Suppose, for example, that Betty and Sally testify independently that they heard a burglar enter my house. They might both have mistaken the noise of a dog for that of a burglar, and because of this common uncertainty, it is not possible to combine degrees of belief based on their evidence directly by the Dempster's rule. But if to consider explicitly the possibility of a dog's presence, then three independent items of evidence can be identified: evidence for or against the presence of a dog, evidence for Betty's reliability and the evidence for Sally's reliability. These items of evidence can be combined by Dempster's rule and the computations are facilitated by the structure that relates the different questions to arise.

Belief and Plausibility can be viewed as providing a lower and upper bounds respectively on the likelihood of $\mathbf{U}$. Over all the possible states and worlds the Dempster and Shafer belief and plausibility functions are defined.

$$Bel(x) : 2^W \to [0..1]$$
$$Plaus(x) : 2^W \to [0..1]$$

Where $W = \{w_1, w_2, ..., w_n\}$

Since belief and plausibility encode evidence, they cannot be defined solely on individual states

$$\sum_i Bel(w_i) \le 1$$
$$\sum_i Plaus(w_i) \ge 1$$

2.2 Grading Knowledge

Information Extraction is akin to "Knowledge Extraction" in text maps natural language onto a formal representation of the facts contained in the texts. Common text knowledge extraction methods show a severe lack of methods for understanding natural language "degree expressions", which describe gradable properties like price and quality, respectively [3]. However, without an adequate understanding of such degree expressions it is often impossible to grasp the central meaning of a text. Degree expressions describe gradable attributes of objects, events or other ontological entities. Complex Lexical semantics is to be instrumented when describing what remains constant when a word is put into different contexts. Fuzzy Logic is even implemented for defining various grades and scales for measuring quantitative text. The special calculi for the text are available to derive valid conclusions from a representation that sticks near to the surface of the utterance and can also handle non-gradable text. Ontological research degree expressions are denoted as their own entities, the upper ontology consisting of the categories such as, Physical-Object and Action augmented by the primitive concept Degree. Some researchers believe that ontological entities are too unparsimonious; the lexical approach only seems to be the one that overcomes the problems. Adjectives of the text propose much easier method of grading. The gradability of (most) adjectives is more obvious than the gradability of members of the other word classes. Besides the fine frequency of gradable adjectives another reason to focus on them is that any advancement of more general mechanisms for degree expressions can easily be contrasted with other authors' research on adjectives, while — but as a fact — degree expressions in general have not been considered in any approach towards deep understanding of language.

A well accepted and approved classification of adjectives is much more important for classifying minor and major groups. Metonymous collection of adjectives plays a fair role in describing the adjective classes. Multiple word senses, even disregarding the metonymous figurative interpretations many relative adjectives still cannot be attributed a single meaning. A common example for this observation is the dichotomy between "physical — mental" and "physical — logical". Depending on the context "physical" is put into, the appropriate conceptual scale must be chosen. Word sense disambiguation, hypallage, deserves far more attention to be developed technologically.

3. PROPOSED WORK

3.1 Probabilistic Grading

Degree expressions, gradability on comparisons, sub-categorizations, ontological significances, metonymy, word senses and sense disambiguation are some of the concrete and qualified methods for grading text. These methods are employed with priori information about the characteristics of the text corpus in the experiments related to determining the knowledge. The preponderance on the methods of grading has evidence in assigning or grading text or knowledge to some degree by means of comparisons and with probabilistic distinctions is the quintessential area of the research that has been working around by IE personnel waiting with eagerness. Current models for this problem have been studied mostly from a linguistic perspective and less so from one of real-world text understanding. The semantic significance is bought into the picture of research only to a small extent that promises the development of rich cohesive frameworks as a future work. A large number of implausible readings that deal with realistic outputs on the literature generate a negative impact on the natural language analysts. Syntax oriented approaches methodically fail to account for interpretations that depend entirely on semantic or conceptual criteria.

In the present work, probabilistic grading employs statistical and probabilistic methods for grading knowledge. There are several works that have been implemented for measuring correlation to express the relationship between two or more variables. Canonical correlation is an additional procedure for assessing the relationship between variables. As there are computational issues in canonical correlation which are viewed as limitations for determining the degree by comparison for grading the text, the subjective probability is applied for the successive discovery of grading the text and knowledge from the large corpus. The method proposed is feasible to implement in the theories of "discourse and corpus" that endorses the quality results in grading.

3.2 Dempster-Shafer Algorithm

Dempster-Shafer evidence theory and subjective probability provides a useful computational scheme for integrating uncertainty information from multiple sources [4]. The algorithm initiates with; the frame of discernment of the problem domain that is, to determine the grade of text. Assume U be the set of mutually and exhaustive hypotheses i.e., all possibilities of expressing the grade of text. The degree of evidence is used to determine the degree of the text. In general, the evidence is calculated for the individual of the hypotheses, but in the context the evidence determines the grade probabilistically for the valued text. That the more of its value the text is graded as good; evidence holds the fact of good grade for the text. The evidence is the evidence is observed using a function m(x) that provides the following Basic Probability Assignments on U.

Given a certain piece of evidence about the value of the text, the belief that one is willing to commit for the specified value for the text exactly to A is represented by m(A), this holds for any $A \subseteq U$. The subset A of frame U is called the focus element of m, if m(A)>0. That is; to determine grade of a particular text is within the scope of a set of specified values.

$$\begin{cases} m(\phi)=0 \\ \sum_{A\subset U} m(A)=1 \end{cases}$$

The DS theory starts by the assumption of Universe of Discourse θ, also called a Frame of Discernment. This is a set of mutually exclusive alternatives. Considering the current case study of determining the grade of a text, θ would be the set consisting of all possible grades.

Elements of 2^{θ}, i.e., subsets are the class of general propositions in the domain. For example, the proposition "The grade is highly starred" corresponds to the set of the elements of θ which are high graded stars.

A function m:$2^{\theta}\to[0,1]$ is called a basic probability assignment if it satisfies m(ϕ)=0 and

$$\sum_{A\subseteq\theta} m(A) = 1$$

The quantity m(A) is defined as A's basic probability number. It represents the strength of some evidence; our exact belief in the proposition represented by A.

A function m:$2^{\theta}\to[0,1]$ is called a belief function if it satisfies Bel(ϕ)=0, Bel(ϕ)=1, and for any collection $A_1,\ldots A_n$ of subsets of θ.

$$Bel(A_1 \cup,\ldots \cup A_n) \geq \sum_{\substack{1\subseteq\{1\ldots n\} \\ I\neq\phi}} (-1)^{|I|+1} Bel\left(\bigcap_{i\in 1} A_i\right)$$

A belief function assigns to each subset of θ a measure of our total belief in the proposition represented by the subset.

There corresponds to each belief function one and only one basic probability assignment. Conversely, there corresponds to each basic probability assignment one and only one belief function. They are related by the following two formulae:

$$Bel(A) = \sum_{B\subseteq A} m(B)$$, for all A ⊆θ

m(A) = $$\sum_{B\subseteq A} (-1)^{|A-B|} Bel(B)$$

Thus a belief function and a basic probability assignment convey exactly the same information. Corresponding to each belief function are three other commonly used quantities that convey the same information:

A function Q:$2^{\theta}\to[0,1]$ is called a commonality function if there is a basic probability assignment, m, such that

$$Q(A) = \sum_{A\subseteq B} m(B)$$ for all A⊆θ

The doubt function is given by

$$Dou(A) = Bel(\sim A)$$

And the upper probability function is given by

$P^*(A) = 1 - Dou(A)$
This expresses how much we should belief in A if all currently unknown facts were to support A.

This the true belief in A will be somewhere in the interval $\lfloor Bel(A), P^*(A) \rfloor$

3.3 Probabilistic Comprehensive Grading
It ought to be certainly counted that probabilistic approaches focus a paradox in process modeling and data comprehension. And it is simultaneously one of the oldest and one of the newest research areas in linguistics, where already much of the research that was done is of statistical and probabilistic in nature [1][2]. Actually, the probabilistic comprehension and modeling is drawn from early Bayesian precursors.

Probability theory is certainly the best normative model for solving problems of decision making under uncertainty [1]. But perhaps it is a good normative model, but a bad descriptive one. Despite the fact that probability theory was originally invented as a cognitive model of human reasoning under uncertainty, perhaps people do not use probabilistic reasoning in cognitive tasks like language production and comprehension.

Probabilistic modeling provides evidence for knowledge comprehension and as well as for knowledge grading. Since, they are not definitely descriptive, a fuzzy set of graded values are attributed to the knowledge for subject grading.

Since one of the oldest and most robust effects of the linguistic texts is the word frequency effect, word frequency plays an important role in auditory, comprehensive, productive and visual modalities [2]. Since, variations of word frequency induce noise, grading of knowledge which includes word frequency effect requires probabilistic comprehension to determine the suitable importance of the knowledge. Grammatical subjects make more impact in frequency identification. In an instance of research experiments, “When subjects made recognition errors, they responded with words that were higher in frequency than the words that were presented”. In an another instance of research experiment “gating paradigm”, in which subjects hear iteratively more and more of the waveform of a spoken word, to show that high-frequency words were recognized earlier (i.e. given less of the speech waveform) than low frequency words [2].

Using the Dempster-Shafer algorithm, it is perfectly possible to perform an experiment for subjective grading of knowledge. The different components that are involved in the experimentation are scenario, evidence, hypothesis and their relationships. Various probabilistic factors are assigned to the attributes that belong to the knowledge. In different hypothesis various grades are attributed to the knowledge, in a selected scenario and an instance of hypothesis the knowledge is graded with select set of qualities. While working with Dempster-Shafer Engine, it is evidently found that a graphical assessment of grading knowledge is possible. In a corpus of linguistic text where a select set of nuggets can be applied with scenario, evidence and hypotheses may be performed to determine the suitable grade.

4. CONCLUSIONS

In this paper the work is aimed at comparison of the probabilistic approaches. The probability theory works on crispy results, where grading requires a scalable demarcation to qualify the knowledge units. The scalable demaracative units are described as evidences and various hypotheses are prepared. The Dempster-Shafer algorithm using the Dempster-Shafer Engine (DSE) has become more practicable to implement the concept of evaluating the grade of the knowledge units using subjective probability.

5. REFERENCES

[1] Rieko Matsuda,Yuzuru Hayashi,Chikako Yomota,Yoko Tagashira,Mari Katsumine and Kazuo Iwaki, "Statistical and probabilistic approaches to confidence intervals of linear calibration in liquid chromatography", The Analyst, www.rsc.org/analyst, (C) 2001.

[2] Dan Jurafsky, "Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production", appeared in ‘Probabilistic Linguistics’, edited by Rens Bod, Jennifer Hay, and Stefanie Jannedy, MIT Press, (C) 2002.

[3] Steffen Staab, "Grading Knowledge: Extracting Degree Information from Texts", ISBN 3-540-66934-5 (C) Springer-Verlag Berlin Heidelberg 1999.

[4] Nic Wilson, "Algorithms for Dempster-Shafer Theory", School of Computing and Mathematical Sciences, Oxford Brookes University.