Advances in Knowledge-Based Technologies

Proceedings of the
Master and PhD Seminar
Winter term 2009/10, part 2

Softwarepark Hagenberg
SCCH, Room 0/2
2 February 2010

Software Competence Center Hagenberg
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 800
Fax +43 7236 3343 888
www.scch.at

Fuzzy Logic Laboratorium Linz
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 431
Fax +43 7236 3343 434
www.flll.jku.at

Program

13:00–14:00 Session 1 ***(Chair: Roland Richter)***

13:00	Bertran Steinsky: Enumeration of Graphs related to Bayesian Networks
13:30	Christoph Roschger: Bridges Between Fuzzy Logic and Linguistic Models of Vagueness
14:00	*Coffee break*

14:15-15:15 Session 2 ***(Chair: Thomas Natschläger)***

14:15	Henrike Stephani: Automatic Peak Detection for Terahertz Spectra - Dynamic Range Determination, Baseline Correction, and Hierarchical Clustering
14:45	Wolfgang Heidl: Men Take Moderately Higher Risks: Initial Results from a Visual Inspection Study

Enumeration of Graphs related to Bayesian Networks

Bertran Steinsky
Johannes Kepler University Linz
Department of Knowledge-Based Mathematical Systems
Austria
bertran.steinsky@jku.at

February 1, 2010

1 Introduction

Directed acyclic graphs (DAGs or ADGs) are used to represent conditional independencies among random variables, e.g., in Bayesian networks. Bayesian networks are used in fields like medicine, image processing, meteorology, or generally in expert systems.

A DAG is a directed graph that contains no directed cycle. A *Bayesian network* (or directed graphical model) is a DAG $G \equiv (V, E)$ together with n random variables X_i and a joint-distribution P with the property that

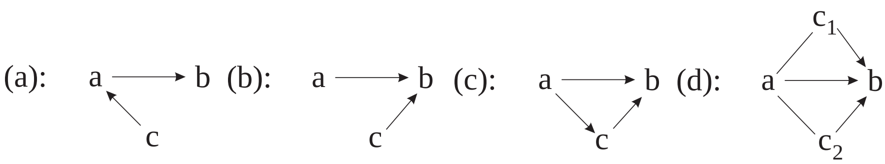
$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_{pa(i)}),$$

where we use the notation $X_A = \times_{i \in A} X_i$ for $A \subseteq V$ and, for a vertex $i \in V$, $pa(i)$ is the set of all vertices j in V such that $j \to i$. Different DAGs can represent the same conditional independence relations[4] among the random variables of the Bayesian network, i.e., the DAGs are Markov equivalent. Here, we mainly deal with counting essential graphs, which can be identified with the equivalence classes of this equivalence relation. For the definitions in this section we often follow Andersson[1], Harary[2], and Robinson[5].

If $a \to c \leftarrow b$ and a, b are not neighboured in G then we call the triplet (a, c, b) an *immorality*. (c has unrelated parents.) The *skeleton* of a graph is its underlying undirected graph. Two DAGs D_1 and D_2 are *graphically equivalent*, if they have the same skeleton and the same immoralities. The *essential graph* D^* *of a DAG* D arises by union of all DAGs, that are graphically equivalent to D. We say a graph is an *essential graph* if it is the essential graph of some DAG.

A *path* π from a to b in G is a sequence $\pi \equiv \{a = a_0, a_1, \ldots, a_n = b\} \subseteq V$ of different vertices, such that $(a_{i-1}, a_i) \in E$ for $i = 1, \ldots, n$. A graph is *strongly connected* if there is a path from every vertex to every other vertex.The *strong components* of G are the maximal strongly connected subgraphs of G. G is a *chain graph* if its strong components are undirected connected graphs. The strong components of a chain graph G are called *chain components*.

An undirected graph is *chordal (triangulated)*, if every cycle of length greater than 3 posses a *chord*, i.e., two vertices that are not neighboured within the cycle, but that are neighboured in the graph. An arrow $a \to b$ in a graph G is *strongly protected* in G, if $a \to b$ occurs in at least one of the following configurations:



2 Characterisation of Essential Graphs

Theorem 1 (Andersson *et* al.[1]) *A graph $G = (V, E)$ is an essential graph, if and only if G satisfies the following four conditions:*
(i) G is a chain graph;
(ii) for every chain component τ of G, G_τ is chordal;
(iii) the configuration $a \to b - c$ does not occur as an induced subgraph of G;
(iv) every arrow $a \to b \in G$ is strongly protected in G.

3 Enumeration of Labelled Essential Graphs

Let $e(N, K, C)$ be the number of labelled essential graphs with N vertices, K chain components, and C cliques (maximal connected undirected subgraphs) and $t(N, K, C)$ be the number of labelled chordal graphs with N vertices, K connected components, and C cliques where $N, K, C \geq 0$.

Theorem 2 (Steinsky[7]) *$e(N, K, C)$ is equal to*

$$\sum_{n=1}^{N} \binom{N}{n} \sum_{k=1}^{K} (-1)^{k+1} \sum_{c=1}^{C} t(n,k,c) e(N-n, K-k, C-c) \left(2^{N-n} - C + c\right)^k, \quad (1)$$

where $e(0, 0, 0) = 1$.

We notice that $t(N, K, C)$ can be computed by methods of Wormald[9].

4 The Number of labelled Chain Graphs

Let c_n be the number of labelled chain graphs with n vertices and

$$k_n = -\sum_{r=0}^{n-1} \binom{n}{r} k_r 2^{\binom{n-r}{2}},$$

for $k_0 = 1, n \geq 1$.

Theorem 3 (Steinsky[6]) *We have*

$$c_n = -\sum_{r=1}^{n} \binom{n}{r} k_r c_{n-r} 2^{r(n-r)},$$

where $c_0 = 1$ *and* $n \geq 1$.

5 The Asymptotic Number of labelled Chain Graphs

Let

$$\tilde{\phi}(z) = \sum_{n=0}^{\infty} \frac{k_n z^n}{n! 2^{\binom{n}{2}}}$$

Theorem 4 (Steinsky[8])

$$c_n \sim -\frac{n! 2^{\binom{n}{2}}}{\tilde{\phi}'(\tilde{z}_0) \tilde{z}_0^{n+1}}, \text{ where } \tilde{z}_0 \approx 0.9477.$$

6 Asymptotic Number of labelled Essential DAGs

An essential DAG is a DAG that is also an essential graph or in other words, an essential graph that has no undirected edge. Let a_s be the number of labelled essential DAGs with s vertices,

$$h(z) = \sum_{s=0}^{\infty} \frac{a_s z^s}{s! 2^{\binom{s}{2}}} \sum_{l=0}^{\infty} \frac{(-z)^l}{l! 2^{\binom{l}{2}}} \sum_{i=1}^{\infty} \frac{\left(\frac{zs}{2^{l+s}}\right)^i}{i! 2^{\binom{i}{2}}} \text{ and}$$

$$\phi(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n! 2^{\binom{n}{2}}}.$$

Theorem 5 (Steinsky[8]) *We have*

$$a_n \sim -\frac{(1 - h(z_0)) n! 2^{\binom{n}{2}}}{\phi'(z_0) z_0^{n+1}} = A \frac{n! 2^{\binom{n}{2}}}{z_0^{n+1}}.$$

where $z_0 \approx 1.488$ and $A \approx 0.1275$.

References

[1] S. A. Andersson, D. Madigan, and M. D. Perlman, A Characterization of Markov Equivalence Classes for Acyclic Digraphs, Ann. Stat. 25, No 2 (1997) 502–541.

[2] F. Harary, R. Z. Norman, and D. Cartwright, Structural Models: An Introduction to the theory of graphs, Wiley, New York (1965).

[3] F. Harary, E. M. Palmer, Graphical Enumeration, Academic Press, New York (1973).

[4] D. Heckerman, D. Geiger, and D. M. Chickering,"Learning Bayesian networks: the combination of knowledge and statistical data", in: Uncertainty in Artificial Intelligence, Proceedings of the Tenth Conference

[5] R. W. Robinson, Counting labelled acyclic digraphs, in: F. Harary, ed., New Directions in the Theory of Graphs, Academic Press, New York (1973) 239–279.

[6] B. Steinsky, Enumeration of Labelled Chain Graphs and Labelled Essential Directed Acyclic Graphs, Discrete Mathematics 270 (2003), 267–278

[7] B. Steinsky, Enumeration of Labelled Essential Graphs, accepted at Ars Combinatoria

[8] B. Steinsky, Asymptotic Behaviour of the Number of Labelled Essential Acyclic Digraphs and Labelled Chain Graphs. Graphs and Combinatorics 20(3): 399-411 (2004)

[9] N. C. Wormald, Counting Labelled Chordal Graphs, Graphs and Combinatorics 1 (1985), 193–200.

Bridges Between Fuzzy Logic and Linguistic Models of Vagueness

Christoph Roschger
Technische Universität Wien, Vienna, Austria

January 29, 2010

Adequate models of reasoning with vague information are not only of perennial interest to philosophers and logicians (see, e.g., [6, 5, 12, 2, 11] and references there), but also in the focus of linguistic research (see, e.g., [10, 7, 1, 8]). Of particular interest from a logical point of view are approaches to formal semantics of a natural language that can be traced back to Richard Montague's ground braking work, firmly connecting modern formal logic and linguistics (see, e.g., the handbook chapter [9] and the widely used textbook [4]). At a first glimpse, it seems that all important contemporary linguistic models of vagueness are *incompatible* with the degree based approach offered by fuzzy logic (see, e.g., [13, 14, 3]). E.g., Manfred Pinkal in his frequently cited (and translated) monograph [10] explicitly argues that many-valued, truth functional logics are inadequate for modelling central linguistic phenomena of vagueness and indeterminateness. One of the key points here is *truth functionality*: consider e.g. the following two sentences:

$$\textit{The sky is blue or the sky is not blue.} \tag{1}$$

$$\textit{The sky is blue or the sky is blue.} \tag{2}$$

If we assign the truth value 0.5 to both $blue(sky)$ and $\neg blue(sky)$, then these sentences (1) and (2) will receive the same truth value in any (truth functional) fuzzy logic. This goes, as Pinkal argues, completely against human intuition.

More specifically, contemporary linguists seem to agree that a special type of *context dependency* is the key to understand the semantics of vague predicates ('tall', 'nice', 'is a heap', 'enjoys', 'likes', ...), but also of corresponding predicate modifiers ('very', 'definitely', ...) and quantifiers ('most', 'many', 'few', ...). However, a closer look at corresponding recent papers on vagueness, in particular [1, 8, 7], reveals that contexts are primarily used to keep track of varying *standards of assertability* connected with *gradable predicates*. This observation is our starting point in exploring formal bridges concepts from t-norm based fuzzy logic and the cited linguistic models of vagueness.

I will show how *fuzzy sets* and *fuzzy relations* can be systematically extracted from a given context space endowed with a probability measure (or more generally, possibility measure) intended to model the relative salience and plausibility of different contexts (standards). Roughly speaking, the membership degree of an individual **a** (say 'Adam') in a fuzzy set modelling a predicate **T** (say 'is tall') gets identified with the probability — alternatively: degree of possibility or degree of necessity — that **a** satisfies the assertability standard associated with **T** in a randomly chosen context. In this manner t-norm and co-t-norm based operators re-emerge as semantic correlates of conjunction, disjunction, and other logical connectives, if one insists on global evaluations that ignore all dependencies between context specific standards pertaining to different predicates. In contrast, local evaluations, i.e. those referring to individual contexts, lead to an *intensional semantic framework*, also for logical connectives. While an intensional evaluation, based on a specific context space, allows to model phenomena of vague language [1, 8, 7] that escape the coarser truth functional approach of fuzzy logic, the price to be paid for the more fine grained analysis is higher computational complexity. In this respect, t-norm based truth functions can be seen as *efficient extensional approximations* to potentially very complex *intensional evaluations* with respect to context dependent assertability conditions.

References

[1] Ch. Barker: The Dynamics of Vagueness, *Linguistics and Philosophy* 25(1), 2002, 1–36.

[2] C.G. Fermüller: Theories of Vagueness Versus Fuzzy Logic – Can Logicians Learn from Philosophers? *Neural Network World Journal* 13(5) (2003), 455-466.

[3] P. Hájek: *Metamathematics of Fuzzy Logic*. Kluwer, 1998.

[4] I. Heim, A. Krätzer: *Semantics in Generative Grammar*. Blackwell, 2008 (1998).

[5] R. Keefe: *Theories of Vagueness*, Cambridge University Press, 2000.

[6] R. Keefe, P. Smith (eds.): *Vagueness – A Reader*, MIT Press, 1987.

[7] Ch. Kennedy: Vagueness and Grammar: The semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy* 30(1), 2007, 1–45.

[8] A. Kyburg, M. Morreau: Fitting words: Vague Language in Context. *Linguistics and Philosophy* 23(6), 2000, 577–597.

[9] B. Partee, H.L.W. Hendriks: Montague Grammar. In J. van Benthem, A. ter Meulen (eds), *Handbook of logic and Language*, Elsevier Science and The MIT Press, 1996, 5–91.

[10] M. Pinkal: *Logic and Lexicon*, Studies in Linguistics and Philosophy, Vol. 56, Kluwer Academic Publishers, 1995.

[11] St. Shapiro: *Vagueness in Context*, Oxford University Press, 2006.

[12] T. Williamson: *Vagueness*, Routledge, 1996.

[13] L.A. Zadeh, Fuzzy Logic, IEEE: *Computer* 21(4), 1988, 83-93.

[14] L.A. Zadeh, Fuzzy logic and approximate reasoning, *Synthese* 30(3-4), 1975, 407-428.

Automatic Peak Detection for Terahertz Spectra

Dynamic Range Determination, Baseline Correction, and Hierarchical Clustering

Henrike Stephani

henrike.stephani@itwm.fraunhofer.de

Abstract

In pharmaceutical quality control, technologies are necessary that identify different chemical compounds. This identification should be performed in a non invasive way. The Terahertz (THz) technology has proven to be a useful tool in that aim. In these wavelengths chemical compounds have characteristic absorption spectra while at the same time most packaging materials such as carton, plastics, and ceramics are not absorbing [1].

There are databases that contain the characteristic spectral expression of many chemical compounds in the infrared range [2]. As an emerging technique most THz spectra are not comparably characterized yet. There are databases such as [3] but the quality and method of acquisition declaredly differ. For most specific applications this is a problem. Particularly to address this problem of comparability, we propose a procedure to detect peaks in THz measurements of solids acquired by time-domain spectroscopy. On a set of six hyperspectral imaging measurements of chemical compounds, this procedure will be presented here.

We especially propose a method to determine the Dynamic Range (DR) of spectra based on standard peak detection. Furthermore, we propose a method for baseline correction. In spite of the normalization with a reference measurement, most transmittance spectra do not have a constant baseline which makes the classification of peaks difficult. We simulate the basic shape of the spectra and propose to use this to create such a constant baseline.

We illustrate this procedure with six measurements of chemical compounds, 36 spectra each. We use unsupervised learning, more particularly hierarchical clustering, to find a robust representation for these compounds by their peaks. This shall illustrate how such a procedure can be used to build comparable THz databases.

REFERENCES

[1] B. Ferguson and X. C. Zhang. Materials for Terahertz Science and Technology. *Nature materials*, 1(1):26–33, 2002.

[2] PM Chu, FR Guenther, GC Rhoderick, and WJ Lafferty. The NIST Quantitative Infrared Database. *Journal of Research - National Institute of Standards and Technology*, 104:59–82, 1999.

[3] NICT and RIKEN. The terahertz database.

Men Take Moderately Higher Risks: Initial Results from a Visual Inspection Study

Wolfgang Heidl WOLFGANG.HEIDL@PROFACTOR.AT
Profactor GmbH, Im Stadtgut A2, 4407 Steyr-Gleink, AUSTRIA

Abstract

Among manufacturing companies there is a widespread consensus that women are better suited to perform visual quality inspection, having higher endurance and making decisions with better reproducibility. We will analyze these gender-related differences by modeling operator decisions with machine learning classifiers. The analysis will be based on data gathered during tests with 100 subjects asked to rate synthetic images based on a predefined set of rules.

This paper presents interim results of the main study with 76 subjects tested so far. The analysis of the results is based on the decision boundary modeled by the ground truth rule set. We show that the test images have been sampled to achieve reasonable coverage of the relevant area around the decision boundary. On average, subjects rate 74% of all images correctly.

We have found two statistically significant differences between the female and male subject groups: (1) Based on the false positive and false negative rate of subject decisions we define a measure for risk propensity. In line with the pilot study, male participants show significantly higher risk propensity, with an effect size of $d = 0.52$. (2) We utilize rule-based classifiers trained on each subject to identify subject-specific decision thresholds. The most pronounced sex-difference has been found for the threshold judging the length of arc-shaped scratches with $d = 0.51$.

Acknowledgments

This work has been supported by the Austrian FEMtech programme of the Federal Ministry for Transport, Innovation and Technology under FFG grant No. 318113. It reflects only the authors' views.