Aggregating Referee Scores: an Algebraic Approach

Rolf Haenni

Reasoning under **UN**certainty Group
Institute of Computer Science and Applied Mathematics
University of Berne, Switzerland

COMSOC'08
2nd International Workshop on Computational Social Choice

Liverpool, UK

3–5 September 2008

Outline

1 Introduction

2 Problem Formulation

3 The Opinion Calculus

4 Evaluating Referee Scores

5 Conclusion

Outline

1. Introduction
2. Problem Formulation
3. The Opinion Calculus
4. Evaluating Referee Scores
5. Conclusion

Outline

1. Introduction
2. Problem Formulation
3. The Opinion Calculus
4. Evaluating Referee Scores
5. Conclusion

Outline

1 Introduction

2 Problem Formulation

3 The Opinion Calculus

4 Evaluating Referee Scores

5 Conclusion

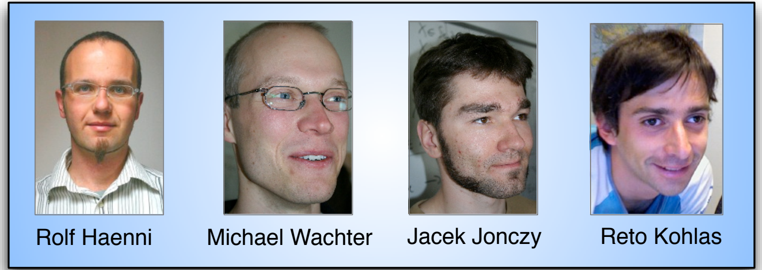
Outline

1. Introduction
2. Problem Formulation
3. The Opinion Calculus
4. Evaluating Referee Scores
5. Conclusion

Outline

1. Introduction
2. Problem Formulation
3. The Opinion Calculus
4. Evaluating Referee Scores
5. Conclusion

RUN Research Group



Supported by:

- Swiss National Science Foundation (Project PP002-102652)
- Hasler Foundation (U/Projects No. 2034 & 2042)
- Leverhulme Trust (Progicnet)


bern
capital impressions

Outline

1 Introduction

2 Problem Formulation

3 The Opinion Calculus

4 Evaluating Referee Scores

5 Conclusion

Peer Reviewing

- Peer reviewing (or refereeing) is the process of evaluating submitted documents by anonymous experts (referees)
- Widely applied by scientific journals, conferences, and funding agencies
- Submitted documents are typically reviewed by 3–4 referees
- Referee reports typically contain:
  - Scores for various criteria (e.g. originality, clarity, etc.)
  - Overall score for paper quality (e.g. 1–10)
  - Level of expertise (e.g. 1–10)
  - Detailed comments
- Papers with highest aggregated scores are accepted $\Rightarrow$ How?

Demo

A prototype implementation is available at:

http://www.iam.unibe.ch/~run/referee

Formal Setting

Input:

$\mathcal{D} = \{1, \ldots, n\}$ → submitted documents

$\mathcal{R} = \{1, \ldots, m\}$ → referees

$referees(i) \subseteq \mathcal{R}$ → referees assigned to document i

$s_{i,j} = (q_{i,j}, e_{i,j})$ → referee j's score for document i

$q_{i,j} \in [0, 1]$ → quality judgement

$e_{i,j} \in [0, 1]$ → expertise level

Formal Setting (cont.)

Output:

$$s_i = \bigotimes_{j \in referees(i)} s_{i,j} \rightarrow \text{combined score } s_i = (q_i, e_i) \text{ for document } i$$

$$q_i \in [0,1] \rightarrow \text{combined quality judgement}$$

$$e_i \in [0,1] \rightarrow \text{combined expertise level}$$

$$\mathcal{S} = \{s_1, \ldots, s_n\} \rightarrow \text{set of combined scores}$$

$$(\mathcal{D}, \preceq) \rightarrow \text{total preorder over } \mathcal{D}$$

$$r : \mathcal{D} \rightarrow \mathbb{N} \rightarrow \text{ranking function over } \mathcal{D}$$

Note that classifying the documents (e.g. accepted/rejected) is a special case of a total preorder $\preceq$

Example

Documents \ Referees	1	2	3	4	5	$\otimes$	Total Preorder: $4 \preceq \{1,3\} \preceq 2$
1	$s_{1,1}$	–	$s_{1,3}$	$s_{1,4}$	–	s_1	$r(1) = 2$
2	$s_{2,1}$	$s_{2,2}$	–	–	$s_{2,5}$	s_2	$r(2) = 1$
3	–	$s_{3,2}$	$s_{3,3}$	$s_{3,4}$	–	s_3	$r(3) = 2$
4	$s_{4,1}$	–	$s_{4,3}$	–	$s_{4,5}$	s_4	$r(4) = 4$

Problem Formulation

- Problem 1: Find an appropriate combination operator $\otimes$
- Problem 2: Find an appropriate total preorder $\preceq$

- Solution: Apply the opinion calculus
  - (i) Transform scores $s_{i,j}$ into opinions $\varphi_{i,j}$
  - (ii) Apply the combination operator $\otimes$ defined for independent opinions $\Rightarrow \varphi_i$
  - (iii) Use various probabilistic transformations $f \in \{g, h, p\}$ to turn each φ_i into a Bayesian opinion $f(\varphi_i)$
  - (iv) Use the natural total order $\preceq_0$ of Bayesian opinions to define $\preceq$

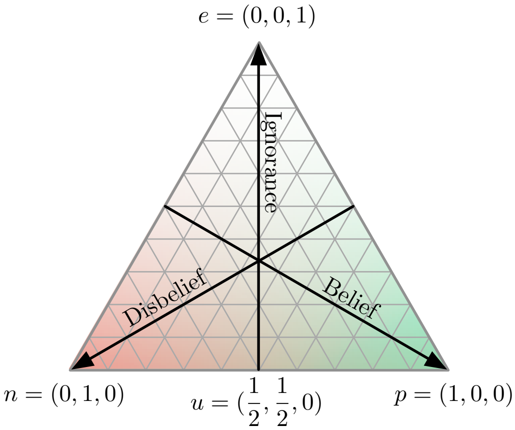
Outline

1. Introduction
2. Problem Formulation
3. The Opinion Calculus
4. Evaluating Referee Scores
5. Conclusion

Opinions

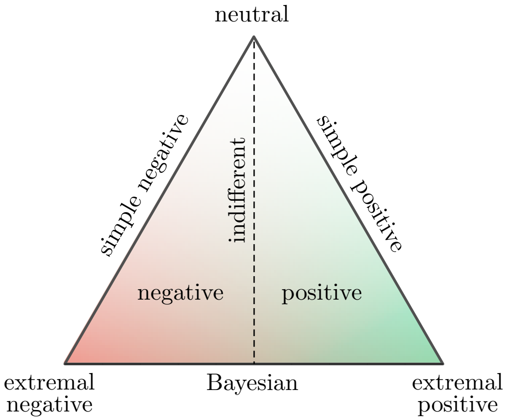
- The opinion calculus is an algebraic version of the Dempster's theory of lower and upper probabilities (Dempster, 1967) for two-valued hypotheses $H \in \{yes, no\}$
- Terminology and references:
  - (Hajek and Valdes, 1991) $\rightarrow$ Dempster pairs, dempsteroids
  - (Jøsang, 1997) $\rightarrow$ opinions, subjective logic
  - (Daniel, 2002) $\rightarrow$ d-pairs, Dempster's semigroup
- An opinion relative to H is a triple $\varphi = (b, d, i) \in [0, 1]^3$
  - $b + d + i = 1$
  - $b =$ degree of belief of H
  - $d =$ degree of disbelief H
  - $i =$ degree of ignorance relative to H
- Dempster's theory provides a probabilistic interpretation for b, d, and i

Opinion Triangle



Opinion Classes

positive:	$b > d$
negative:	$b < d$
indifferent:	$b = d$
simple:	$b = 0$ or $d = 0$
extremal:	$b = 1$ or $d = 1$
neutral:	$i = 1$
Bayesian:	$i = 0$



Combining Opinions

- Let $\varphi_1 = (b_1, d_1, i_1)$ and $\varphi_2 = (b_2, d_2, i_2)$ be independent:

$$\varphi_1 \otimes \varphi_2 = \left(\frac{b_1 b_2 + b_1 i_2 + i_1 b_2}{1 - b_1 d_2 - d_1 b_2}, \frac{d_1 d_2 + d_1 i_2 + i_1 d_2}{1 - b_1 d_2 - d_1 b_2}, \frac{i_1 i_2}{1 - b_1 d_2 - d_1 b_2} \right)$$

- Let $\varphi_i = (b_i, d_i, i_i)$, $1 \leq i \leq n$, be independent:

$$\varphi_1 \otimes \cdots \otimes \varphi_n$$
$$= \left(\frac{1}{K} \left[\prod_i (b_i + i_i) - \prod_i i_i \right], \frac{1}{K} \left[\prod_i (d_i + i_i) - \prod_i i_i \right], \frac{1}{K} \prod_i i_i \right)$$

for $K = \prod_i (b_i + i_i) + \prod_i (d_i + i_i) - \prod_i i_i > 0$

- Click here to start demo

The Opinion Monoid

- $\Phi = \{(b,d,i) : b+d+i=1\}$ is not closed under $\otimes$
  - $\otimes$ is undefined for $p=(1,0,0)$ and $n=(0,1,0)$
  - Add inconsistent opinion $z=(1,1,-1)$
  - Define $p \otimes n = n \otimes p = z$
  - Define $\varphi \otimes z = z \otimes \varphi = z$, for all $\varphi \in \Phi$
- $\Phi_z = \Phi \cup \{z\}$ is closed under $\otimes$
  - $\otimes$ is commutative
  - $\otimes$ is associative
- Therefore, $(\Phi_z, \otimes)$ is a commutative semigroup
  - $e=(0,0,1)$ is the identity element: $e \otimes \varphi = \varphi \otimes e = \varphi$
  - $z=(1,1,-1)$ is the zero element: $z \otimes \varphi = \varphi \otimes z = z$
- Therefore, $(\Phi_z, \otimes, e)$ is a commutative monoid with zero element z

Other Opinion Monoids

Name	Notation	Definition	Identity	Zero
general	Φ_z	$\Phi \cup \{z\}$	e	z
non-negative	$\Phi_\geq$	$\{(b,d,i) \in \Phi : b \geq d\}$	e	p
non-positive	$\Phi_\leq$	$\{(b,d,i) \in \Phi : b \leq d\}$	e	n
simple non-negative	Φ_+	$\{(b,d,i) \in \Phi : d = 0\}$	e	p
simple non-positive	Φ_-	$\{(b,d,i) \in \Phi : b = 0\}$	e	n
indifferent	$\Phi_=$	$\{(b,d,i) \in \Phi : b = d\}$	e	u
Bayesian	Φ_0	$\{(b,d,i) \in \Phi : i = 0\} \cup \{z\}$	u	z

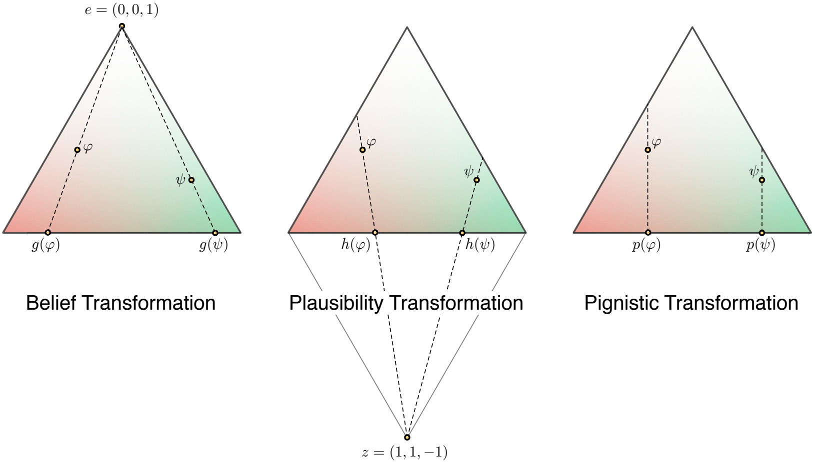
Remarks:

- Φ_+, Φ_-, $\Phi_=$, $\Phi_0 \setminus \{z\}$ possess a natural total order
- $\Phi_0 \setminus \{p, n, z\}$ forms a commutative group

Probabilistic Transformations

- A probabilistic transformation is mapping $f : \Phi_z \rightarrow \Phi_0$
  - Belief transformation:
    $$g(\varphi) = \left(\frac{b}{b+d}, \frac{d}{b+d}, 0 \right)$$
  - Plausibility transformation:
    $$h(\varphi) = \left(\frac{1-d}{1+i}, \frac{1-b}{1+i}, 0 \right) = \varphi \otimes u$$
  - Pignistic transformation:
    $$p(\varphi) = \left(b + \frac{i}{2}, d + \frac{i}{2}, 0 \right)$$
- For each $f \in \{g, h, p\}$, the total order over $\Phi_0 \setminus \{z\}$ defines a total preorder over Φ

Probabilistic Transformations (cont.)



Outline

1 Introduction

2 Problem Formulation

3 The Opinion Calculus

4 Evaluating Referee Scores

5 Conclusion

Referee Scores as Opinions

- Problem 1: Define a mapping from scores to opinions
  - Score: $s = (q, e) \in [0, 1] \times [0, 1]$
  - Opinion: $\varphi = (b, d, i) \in \Phi$
  - Mapping: $\Delta : [0, 1] \times [0, 1] \rightarrow \Phi$

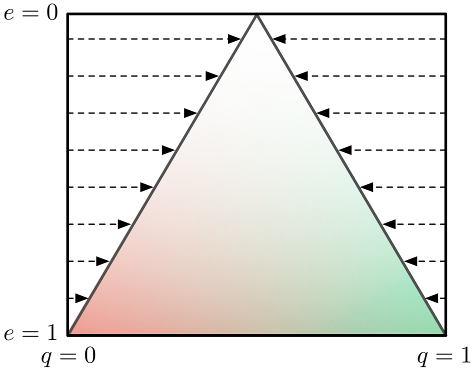
- Solution: Probabilistic interpretation of q and e
  - $e = P(E)$

    $\Rightarrow$ probability of the referee being an expert (event E)
  - $q = P(Q|E)$

    $\Rightarrow$ conditional probability of the document being a high-quality paper (event Q), given that the referee is an expert (event E)
  - If E and Q are probabilistically independent, then

$$\Delta(s) = (b, d, i) = (e{\cdot}q, e{\cdot}(1 - q), 1 - e)$$

Mapping Scores into Opinions

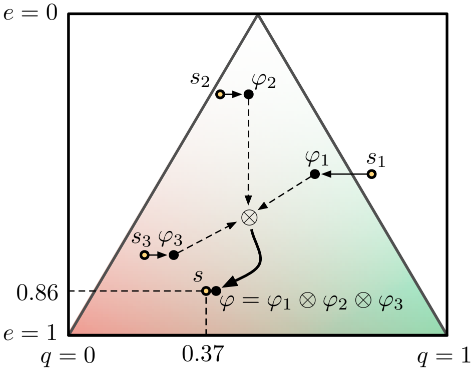


Remarks:

- Δ is invertible: $\Delta^{-1}(\varphi) = (q, e) = (\frac{b}{1-i}, 1 - i)$, for $\varphi \neq e$
- $s = \Delta^{-1}(\Delta(s_1) \otimes \cdots \otimes \Delta(s_k))$

Combining Scores

$$s_1 = (0.80, 0.50) \Rightarrow \varphi_1$$
$$s_2 = (0.40, 0.25) \Rightarrow \varphi_2$$
$$s_3 = (0.20, 0.75) \Rightarrow \varphi_3$$
$$\Downarrow$$
$$s = (0.37, 0.86) \Leftarrow \varphi$$



Document Ranking

- Problem 2: Determine document ranking
  - Documents: $\mathcal{D} = \{1, \ldots, n\}$
  - Scores: $\mathcal{S} = \{s_1, \ldots, s_n\}$
  - Opinions: $\Delta = \{\varphi_1, \ldots, \varphi_n\}$
  - Define ranking $r(i)$ for all $i \in \mathcal{D}$
- Solution: Use probabilistic transformation of φ_i
  - $f(\varphi_i) = (b_i, 1 - b_i, 0)$ for $f \in \{g, h, p\}$
  - $i \preceq j \;\Leftrightarrow\; f(\varphi_i) \preceq_0 f(\varphi_j) \;\Leftrightarrow\; b_i \leq b_j$
  - $i \prec j \;\Leftrightarrow\; i \preceq j \wedge i \not\succeq j$
  - Ranking: $r(i) = |\{j \in \mathcal{D} : i \prec j\}| + 1$

Outline

1 Introduction

2 Problem Formulation

3 The Opinion Calculus

4 Evaluating Referee Scores

5 Conclusion

Conclusion

- Peer reviewing leads to an important judgement aggregation problem
- It can be solved using the opinion calculus (Dempster-Shafer theory)
- The method can be implemented efficiently
- Future work and open problems:
  - Get into a conference management tools (CyberChair, . . .)
  - Empirical study based on data from real conferences
  - Compare/evaluate different probabilistic transformations