Alternative Methods for Aggregation of Expert Judgments: A Preliminary Comparison

Umit Guvenc, Mitch Small, Granger Morgan
Carnegie Mellon University

CEDM Advisory Board Meeting
June 13, 2011

Questions

1. Should we aggregate expert judgments at all?
2. <u>If we do</u>, should we use a differential weighting scheme?
3. <u>If we do</u>, should we use "seed" questions?
4. <u>If we do</u>, how should we choose "appropriate" seed questions?
5. <u>If we do</u>, how do different weighting schemes perform under different circumstances?

Questions

1. Should we aggregate expert judgments at all?
2. <u>If we do</u>, should we use a differential weighting scheme?
3. <u>If we do</u>, should we use “seed” questions?
4. <u>If we do</u>, how should we choose “appropriate” seed questions?
5. <u>If we do</u>, how do different weighting schemes perform under different circumstances?

Research Question

- In aggregating expert judgments in the face of uncertainty and disagreement, how do different weighting schemes perform under different circumstances?
  - Equal weights method
  - Cooke's "classical" method
  - "Bayesian" likelihood method

1 Determine expert weights using seed questions (known)

2 Use weights to aggregate expert judgments for prediction (unknown)

Cooke’s “Classical” Method

- Normalized “Cooke scores”
  - Cooke Score = f(Calibration,Information)
- “Classical” model (vs Bayesian model)
  - “Goodness of fit” comparing consistency of experts’ uncertainty quantiles with the observed values
- Non-parametric, based on Chi-sq distribution
- “Macro” validation only
  - Based on frequencies across percentiles across all seed qs
- Not very transparent or easy to explain

“Bayesian” Likelihood Method

- Normalized likelihoods
- $L = f_x(X|M,SD)$
- Parametric, must assume and fit an error distribution for experts
  - normal distribution assumed in analysis that follows
- “Micro” validation incorporated
- More intuitive, easier to explain

Our Approach

- MC Simulation
- 10 seeds: XT(i) ~ Normal(MT(i),SDT(i)) (i=1 to 10)
- Characterize experts w.r.t. accuracy and precision
  - K1 to characterize accuracy via Mean: M(i)
  - K2 to characterize precision via SD: SD(i)
  - X5%, X50%,X95% calculated
- 2-Expert Comparison
- Multi-Expert Comparison (cross validation)
  - Simulate 10 seeds and expert estimates for each
  - Determine expert weights using 9 seeds
    - Leave one seed out at a time to predict

2-Expert Comparison

- 10 seeds, normally distributed
- Expert 1: “Ideal” Expert
- Expert 2: “Non-Ideal” Expert
  - characterized by K1 and K2

- Likelihood Method vs Cooke’s Method

- Result:
  - Do not take “monotonicity” for granted

Multi-Expert Comparison

- 10 seeds, each distributed normally
- 10 experts, characterized by K1 and K2
- “Leave-One-Seed-Out-At-a-Time” to predict
- Compare “Error”, GIVEN expert pool
- Result:
  - Performance for individual seeds:
    - Likelihood > Cooke’s > Equal weights
  - Performance across all seeds:
    - Likelihood > Cooke’s > Equal weights
      - MSE (Mean Squared Error)
        - » Equal Weights Method: 495
        - » Cooke’s “Classical” Method: 8.44
        - » “Bayesian” Likelihood Method: 4.71

Preliminary Key Points

- Cooke's method has some drawbacks compared to Likelihood method
  - Theoretical
  - Empirical
- Likelihood method is not perfect either
  - Requires a specified distribution
    - Considering alternative 3-parameter distributions that can be fit from expert assessment of 5th, 50th, and 95th percentile values

Next Steps

- Simulate with
  - Different seeds
  - Different number of experts with different profiles

- Design experiment for comparison to test with experts

- Apply to coral reef predictive model and/or other CEDM applications where expert elicitation employed