

Agreeing to Disagree: Leveraging Consensus and Divergence in Bayesian Belief Aggregation

Kshanti A. Greene and George F. Luger

University of New Mexico
Department of Computer Science
{kshanti, luger}@cs.unm.edu

Abstract

We present a new approach for combining the beliefs of many individuals using graphical models. Existing Bayesian belief aggregation methods break several theoretical assumptions for Bayesian reasoning. More practically, existing opinion pool functions that compute a single value to represent the belief of all contributors do not represent reality well, especially in cases where there are many diverse opinions. Divergence is a natural result of combining opinions from individuals with different beliefs, backgrounds and experiences. Instead of forming a single consensus value that will *average out* this diversity, we find clusters of agreement for each probability distribution and propagate the cluster means throughout the network during inference. We utilize a social network that tracks the agreement between individuals and the normalized graph cut algorithm to find emerging groups of consensus in the agreement network. We leverage the agreement that occurs across multiple belief estimates to help reduce the complexity that may arise as the means are propagated through-out a belief network. By monitoring agreement over time we may also expose the variety of backgrounds that will help explain divergence in belief. This paper discusses the approach, background and our motives for ongoing research.

Introduction

Many fields have a need to build predictive models from a number of different individuals who each can contribute their experience and beliefs to the whole. The motivations for collecting beliefs from multiple individuals fit into two different categories. First, we may be interested in building the most *accurate* model of a domain or future of interest from a set of experts or sensors, each of which brings a different background or specialization to the table. This situation would be typical of an expert system or fusing output from a sensor network to improve situation understanding. Second, we may simply be interested in the opinions of the individuals, and desire to build a concise democratic model that is *most representative* of the beliefs of the individual contributors. This situation most closely models polling, and thus far, has not been discussed in Bayesian belief aggregation literature. Our approach addresses both motivations, and therefore presents a generalized technique for

belief aggregation that can be used in many real-world situations. However, we constrain our approach to situations in which enough disagreement occurs to render existing belief aggregation approaches imprecise. We suspect that many realistic situations fit into this category.

Consider an example in which we would like to build an *accurate* predictive model from a number of experts. Several law enforcement units working autonomously have discovered different suspicious activities relating to the purchase of restricted biochemicals, information retrieved about a city's water system, and unusual travel activity into the state. Now imagine that there is a network of very observant individuals connecting all law enforcement units, government agencies and military intelligence units. These individuals are able to monitor information entered by the different groups and make connections between them. Each group may enter its own piece of information. Group *A* enters "Suspect purchases water soluble chemicals in Canada and smuggles it into Buffalo." Group *B* enters "Frequent downloads of information such as maps of the Hudson River water basin from an IP address in Pakistan." Group *C* enters "Several single men with Passports from Pakistan arrived at airports across New York within a two week period in August."

One of the very astute individuals monitoring all this information may notice that all the activity is in New York, and there is overlap between group *A*'s and *B*'s observations referring to water and an overlap between *B* and *C*'s observations regarding Pakistan. This astute (or paranoid) person may conjecture that there could be a relationship between these pieces of information and raise a flag to the appropriate authorities to look into this. However, perhaps another agent sees the same activity and concludes that some of these coincidences can be explained away by additional information. For example, perhaps the Pakistani men happened to be graduate students at state universities and it was the start of the fall semester.

Thanks to imperfect recall, bias, and varying backgrounds, even if multiple individuals have observed the same activities, they will often have differing opinions on the implications of their observations. One of them may believe that an attack is imminent while another has intuition that indicates a less threatening scenario. Whenever we have multiple human (and even autonomous) contributors to a model we must consider that there will be disagreement.

However, in order to make decisions based on the observations, we must be able to form a concise conclusion from their opinions.

We are aggregating the partial probabilistic models of individuals with diverse and sometimes conflicting knowledge in order to build a representative model. Existing belief aggregation approaches are unable to realistically represent diverse beliefs because they attempt to form a single *consensus* model which averages away any conflict. As a consequence of this approach, they also break many theoretical assumptions that are central to Bayesian logic. Our research leverages agreement between individuals to allow models to be built that capture both consensus and diversity among opinions. This also enables us to provide approximate solutions to the issues that have limited the progress of belief aggregation and topological fusion. Capturing belief and structure from diverse information sources is a significant bottleneck for probabilistic systems and is one that needs to be addressed before any theoretical system can be used in practice. Our solution offers a practical and theoretically sound approach to this challenge, enabling researchers and knowledge engineers to build holistic models from diverse and potentially conflicting sources.

This paper is organized as follows. We first discuss related work in Bayesian belief aggregation and introduce problems with existing approaches that we are addressing. We then discuss the details of our approach, including the belief aggregation technique we base our approach on, using clusters of consensus to represent divergent beliefs, and using social networks to capture agreement for clustering and inference. We then discuss how our approach can be used to develop both accurate and representative models, and finally describe some initial results from an experiment which elicited beliefs on political outcomes. We conclude by raising some research questions that we address during the ongoing research.

Background

Belief aggregation is the process of combining probability estimates on the same distribution from multiple human or software agents. Early researchers developed various *opinion pool* functions whose output was a numeric result of the combination of a number of inputs. Matzkevich and Abramson (1992) cited two different approaches to belief aggregation that were discussed at the time. The first was called *posterior compromise*, which combines the beliefs after the network and probabilities have been defined and a query has been made. In other words, one would query separate networks and then combine the result. The authors introduced their alternative approach called *prior compromise* that instead found a consensus network *before* inference was done to determine the result of a query. This approach would involve fusing together networks that may also have different structure. Once networks were fused, they combined the beliefs on local relationships using an approach called *family aggregation* (Pennock and Wellman 1999).

An opinion pool function is a mathematical function to form a single aggregate value from multiple beliefs. Mathematically, $P_0 = f(P_1, P_2, \dots, P_n)$ where each P_i is the

probability estimation from the i^{th} contributor given N contributors. P_0 is the *consensus* estimation. The two most commonly used opinion pools are the linear opinion pool (LinOP) and the logarithmic opinion pool (LogOP). If the world is composed of m possible binary events LinOP is a weighted arithmetic mean with the following formula:

$$P_0(x) = \sum_{i=1}^N \alpha_i P_i(x) \quad (1)$$

where α_i is a non-negative weight assigned to each of the N contributors and $\sum_{i=1}^N \alpha_i = 1.0$. LogOP is a weighted geometric mean with the following formula:

$$P_0(w_j) = \frac{\prod_{i=1}^N [P_i(w_j)]^{\alpha_i}}{\sum_{k=1}^{2^m} \prod_{i=1}^N [P_i(w_k)]^{\alpha_i}} \quad (2)$$

where w_j and w_k are each one of 2^m possible events given m states of the world (Pennock and Wellman 1999).

Given the potential for opinion pool functions to form models by aggregating multiple beliefs, Pennock and Wellman investigated whether combined belief yields enough structure to form a graphical representation. In (Pennock and Wellman 1999) the authors show that even when agents are in agreement on the structure of a model, existing aggregation methods do not yield the same structure. They prove that it is not possible to maintain consistent structures using an opinion pool function unless Markov independencies are preserved.

Pennock and Wellman also introduced the market-based belief elicitation and aggregation approach. This approach requires that individuals back up their beliefs by buying and selling *stocks* that indicate their confidence in an event occurring (Pennock and Wellman 1997; 2005). The consensus value is determined by the resulting stock price. While this approach may improve accuracy when all agents have the same risk tolerance, this case is highly unlikely. In general a market based approach increases the subjectivity of the result as each individual has unequal desire to make a bet on their beliefs. While problematic, a number of other researchers have followed in the competitive market-based approach to belief aggregation (Ottaviani and Sorensen 2006; Napp and Jouini 2006; Maynard-Reid and Chajewska 2001).

Belief aggregation raises a more philosophical issue that has thus far not been discussed in the literature. The logic behind averaging to find one *consensus* based on many possibly divergent opinions is flawed. The more divergent the opinions, the more unrealistic and misrepresentative the average is. Consider the following situation; Joe believes that Democrats winning the election is very unlikely (10%). Susan believes that Democrats winning the election is almost certain (90%). The result of averaging these opinions implies that people believe the election is a tossup, while the individual opinions clearly are quite polarized. A second situation has three opinions; one at 55%, one at 45% and one at 50%. The average of these also calls the election a tossup, but the opinions more closely reflect this conclusion. To maintain a realistic representation of belief from many individuals, the resulting consensus model should distinguish between these two situations.

Our Approach

We now discuss our aggregation approach that merges family aggregation, clustering and social networks to form and propagate *consensus belief clusters* for Bayesian inference.

Family Aggregation

Family aggregation (Pennock and Wellman 1999; Matzkevich and Abramson 1992) is an aggregation approach in which LinOP is applied within each conditional probability table (CPT) between the parents and child. Family aggregation is in the form:

$$P_0(X|Pa_x) = f(P_1(X|Pa_x), \dots, P_m(X|Pa_x)) \quad (3)$$

Where P_0 is the consensus probability and $P_i(X|Pa_x)$ is each of m individual's probability estimate of X given its parents Pa_x and $P_0(X|Pa_x)$ is the consensus. Using LinOP, the elements in the multiple supplied CPTs are averaged to form a *consensus* CPT. We will utilize independence assumptions such that each table being considered will represent a single parent \rightarrow child relationship. This will reduce the overall complexity and enable fusion of differing sub-structures (Matzkevich and Abramson 1992).

The problem is that family aggregation also fails to uphold Bayesian properties, specifically the Bayes rule:

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)} \quad (4)$$

An example in (Pennock and Wellman 1999) illustrates this. In summary, multiple agents supply beliefs on the conditional probabilities of a parent child relationship, specifically $P(X)$, $P(Y|X)$, and $P(Y|\neg X)$. The estimates are then averaged to find a consensus CPT and then the results are used to find the joint distribution $P(Y, X)$. They then utilize the original agent estimates to reverse the edges between parent and child to find $P(X|Y)$ using (4) and again find the consensus CPT by averaging these values. When they utilize the consensus CPT to find the joint distribution $P(X, Y)$, these values are not equal to the values $P(Y, X)$, found using the initial consensus CPT, which means that the Bayes rule properties are broken using family aggregation. We observe that the variance (or error) between the joint distributions $P(Y, X)$ and $P(X, Y)$ is much greater when the original beliefs are more divergent, for example, agent A states that $P(Y|X) = 0.2$ and agent B states that $P(Y|X) = 0.8$. This observation is a foundation of our approach.

Consensus Belief Clusters

Our approach leverages agreement and disagreement between individuals to reduce the error that occurs in situations similar to the previous example, as well as form a more realistic consensus model. Instead of computing a single consensus value (or average) to represent the beliefs of many potentially divergent opinions, we cluster similar probability estimates to form *consensus belief clusters* and apply an opinion pool function to each of the clusters. The result is a set of k means, as well as a weight ω for each cluster indicating the relative number of estimates that fit into the cluster. The variance σ within each cluster is partially determined

by a similarity measure s , that indicates the maximum distance between probability estimates that we will allow. A lower value of s will result in decreased error, but a greater number of clusters.

Two key variables determine the appearance of the consensus model. The first is the degree of *disagreement* between belief estimates, or δ . If all estimates are within s of each other, then we can utilize the traditional opinion pool functions to form a single consensus value. If not, then clustering will split the estimates into multiple groups of consensus. The second important variable is m , or the number of individual belief estimates. In situations where there are only a few estimates, we find the mean of each cluster using LinOP. However, if we have a significant number of belief estimates we can fit Gaussians to the clusters, resulting in a *Gaussian mixture model* (Dasgupta and Schulman 2007; Sanjeev and Kannan 2001) containing k components. The mixture model provides us with a compact, yet informative representation to visualize the distribution of belief.

Utilizing consensus belief clusters allows us to represent both consensus and divergence across a probability distribution. The positive implications of our approach include:

- Although we cannot eliminate the inconsistencies that cause existing opinion pool functions to break Bayesian formalisms, utilizing consensus belief clusters will reduce the error that arises (for example when computing the joint distribution using family aggregation).
- We can represent both converging and diverging opinions where appropriate, building a more realistic model. This has benefits in many applications including polling and prediction. In polling, we would like to capture a concise representation of significant clusters of opinions. In prediction, maintaining an awareness of outliers (beliefs that are not well supported with consensus) is important in situations that may involve rare and unexpected events.
- The outliers may also indicate noise and bias that can be isolated from the dominant consensus. The clusters containing stronger consensus will likely be more accurate than a model that aggregates all beliefs into one consensus using traditional aggregation methods because any noise and bias that appears as outliers will not be included.

Social Networks of Agreement

To determine the extent of agreement and extract the clusters of consensus we utilize a graph theoretic approach that has been used by many researchers across various computational and statistical fields (Sanjeev and Kannan 2001; Shi and Malik 2000; Clauset, Newman, and Moore 2004). We build a *social network* in which nodes represent individuals and edges between nodes indicate that two individuals agree (in other words their estimates are within a similarity measure, s of each other). Each edge also has a weight that indicates the frequency that two individuals agree across multiple probability estimates. More formally, we define the graph G to be composed of (V, E, s) , where V is a set of vertices (each representing a unique individual), E is a set of edges, and s is a similarity measure. An edge exists between two nodes v_i and v_j , $i \neq j$ iff the individuals represented by

the nodes have supplied belief estimates P_i and P_j such that $|P_i - P_j| \leq s$. Each edge is also associated with a weight w that is a function of the frequency of agreement between two individuals i and j .

Utilizing a graph to track similarity as described is a common approach to capture similarity between many individual features or agents. (Sanjeev and Kannan 2001) describe an algorithm to find mixture models by creating edges between nodes that share close neighbors, and then finding the cliques in the graph. (Shi and Malik 2000) utilize a similarity graph to describe the similarity between pixels in an image. Some researchers have used social networks to describe agreement or shared beliefs (Robins and Lusher 2006).

The primary purpose of forming a graph in which edges represent similarity or agreement is to create a structure from which algorithms can then be used to extract or distinguish clusters of nodes that tend to group together due to their higher degrees of similarity. In our case, we utilize the graph to extract the clusters of belief for individual probability distributions *in addition to* detecting agreement across *multiple* probability distributions which highlights more consistent consensus between individuals and groups.

We utilize the *normalized graph cut* algorithm defined by (Shi and Malik 2000) to extract the consensus clusters from the agreement network. The graph cut algorithm works by separating partitions of a network by removing edges between dissimilar partitions. As in our agreement network, edge weight w_{ij} represents the similarity between nodes i and j . Given a graph G , it can be divided into two partitions A and B . The *cut* is a value that indicates the amount of dissimilarity between A and B . It is defined as follows:

$$cut(A, B) = \sum_{u \in A, v \in B} w_{uv} \quad (5)$$

An optimal partitioning is one that minimizes this value. While finding the minimum cut is an NP hard problem, many efficient approximations exist (Shi and Malik 2000). The *normalized cut* algorithm is an extension of the minimum cut that normalizes the cut value by the overall association between groups. This helps to reduce the incidence of cuts occurring that partition small, isolated cuts but ignore more interesting partitions between larger, less distinguished partitions.

To determine the consensus clusters for a single probability distribution, we utilize an edge weight of 1 for all edges between nodes that have a similarity $< s$. Once clusters of consensus have been found in our agreement network, we determine the means of these clusters. The next step is to propagate these means throughout the network for Bayesian inference.

Leveraging Agreement for Inference

Bayesian inference is the process of propagating probability distributions across network nodes to compute the overall joint probability distribution of the variables in a network. The problem is that now instead of one consensus value for each probability distribution, we have a set of k means. If we have k_A values that represent the probability distribution

for variable A , then we will potentially have to propagate each of those values to A 's children. If B is A 's child and B has k_B means, we will have to combine k_A means with k_B means as well as the means from all of B 's other parents. This explosion of values could easily make inference using our consensus models intractable.

One obvious approach would be to find the largest consensus cluster from each probability distribution and only propagate its mean throughout the network. However, this would make our goal of realistically representing divergence unachievable since we would be dropping many valid consensus clusters.

Consider instead the case in which every person that agreed with another person on one belief also agreed with that person on all other beliefs. If this were truly the case, then we would not need to combine one group's beliefs with any other group's beliefs since there would be no overlap between groups. Instead we could have k agents with each agent representing a consensus group, each with a model of the network and a single consensus value for each probability distribution. Then the problem is simply a typical propagation problem in which we propagate one value throughout the network. This would result in k results for each query in which those k results would accurately represent the beliefs of each group.

Of course it is not likely that individuals will *always* agree on beliefs, therefore this example is much too simplistic. However, it is certainly possible that many of the individuals who agree will agree a significant proportion of the time. We return to our agreement network and edge weights. Each time two individuals i and j agree, we increment a count on the edge between v_i and v_j . The edge weight w_{ij} is this count, normalized by the total number of times individuals have agreed. In other words, the weight represents the amount of agreement between individuals *over time*. When we run the graph cut algorithm on the network using these weights, we will find the consensus clusters (or groups) that emerge across all probability distributions. We then assign an agent to represent each group that will maintain the network for the group, including propagating values within each network. While we will still have agreement between groups we can greatly reduce the explosion of the number of values being propagated throughout the network by constraining propagation based on the amount of agreement.

Accurate versus Representative Models

In many domains of interest the goal is to combine (or fuse) a number of sources to create a single representative model (Steinberg and Bowman 2004). In typical applications, such as medical, military and technical, the knowledge engineers are interested in building the most accurate model possible. In this case, the performance of each expert may be a consideration when contributions are combined. Better performing experts (or agents) may be given more weight while contributions from unreliable sources are discounted. In some situations, perhaps one expert may simply have more experience in a particular area than another. Our approach is to consider the amount of *support* that an individual brings to

the model. First of all, an expert may be supported by other experts that agree with her on the current belief, or which have agreed with her on other beliefs. This *social support* is a function of the degree of the node that represents this individual in the agreement network and the weights of the node’s edges.

An expert may also be supported by the amount of information that she supplies to back up her claim, or *informational support*. In a causal model, this property can be inferred from the set of variables and causal relationships for which the expert supplies belief estimates. In the counter-errorism example used in the Introduction, one agent had an intuition that the university semester was starting, and therefore he might add a variable to the network representing this possibility, with a conditional probability that would decrease the likelihood of an attack on New York because of visiting Pakistani men. If we look for the differences across multiple probability estimates, we may be able to infer the backgrounds that explain differences in expert opinions.

Informational support is also interesting when aggregating belief in models in which all individual’s opinions are considered equally valid. Some domains, such as polling or surveys, may be more interested in models that provide the best *representation* of the contributors’ beliefs. In this case, we may not use informational support to validate the individuals, but we can still use it to discover variations in individuals’ backgrounds. In the *representative model*, we are particularly interested in generating a realistic distribution of the opinions. A higher quality representative model will be one in which the results of querying the network more closely match the results of querying a reasonable sampling of the contributors.

Results

The research discussed in this paper is in progress, therefore we only have partial results available to demonstrate our approach. We ran an experiment using Mechanical Turk (Amazon 2008), an Amazon.com sponsored system that allows one to hire many individuals to undertake simple on-line tasks. We asked Mechanical Turk workers to supply likelihoods for several political outcomes. In Table 1 we show the results of using a spectral normalized graph cut algorithm (Shi and Malik 2000) to split an agreement network into two sub-groups based on answers from 186 individuals. In the agreement network an edge exists between two nodes representing individuals whose probability estimates are within $s = 0.26$ of each other. The weight on the edge indicates the number of questions the individuals agreed upon. The columns labeled $G1$ and $G2$ show the means of the two groups.

The table shows considerable divergence across many questions that are typical *Democrat* versus *Republican* issues. For example, the group labeled $G1$ believed that it was 73% likely that Obama will be elected president and also overwhelmingly believed that climate change is effected by humans. The group labeled $G2$ believed it was less likely that Obama will win and also believed the likelihoods of finding bin Laden and independence on foreign oil were higher if McCain wins. These results demonstrate that it

is possible to find groups of consensus even across several issues. The results also demonstrate the potential of our approach for use in polling. Instead of just asking people their belief on a single event, we can ask people to back up their beliefs with reasoning, or cause them to consider the consequences of certain events. We also note that the larger group’s consensus was a more accurate prediction of the outcome of the election than the smaller and combined group’s consensus.

Prediction	G1	G2
McCain will be elected president	27.0	47.9
Obama will be elected president	73.0	52.1
US economy continues to decline if McCain wins	91.3	48.3
US economy continues to decline if Obama wins	50.7	58.6
Increased taxes if bailout passes	78.6	69.6
US economy continues to decline if bailout passes	82.9	53.8
US exits Iraq by 2010 if McCain wins	10.0	46.5
US exits Iraq by 2010 if Obama wins	78.6	65.1
Iraq is stabilized if the US leaves by 2010	51.4	37.8
Find bin Laden if McCain wins	9.4	41.6
Find bin Laden if Obama wins	42.1	29.9
Independence on foreign oil in 10 years if McCain wins	10.7	54.4
Independence on foreign oil in 10 years if Obama wins	48.6	47.1
Independence on foreign oil due to alternative fuels	90.0	53.5
Independence on foreign oil due to expanded drilling	10.0	46.5
Climate change effected by humans	94.8	68.6
Number of instances	107	79
Percent	57.6	42.4

Table 1: Results of normalized graph cut on predictions elicited using Mechanical Turk. Columns $G1$ and $G2$ show the means of sub-graphs split by the cut. The predictions were in the range 0 . . . 100.

Research Questions

Aside from validating our approach, we have a number of questions that we will be addressing during our research.

- At which point do traditional belief aggregation methods fail to capture the divergence that occurs and when is our approach most appropriate?
- What are the differences in behavior depending on the number of contributors we have? Are we able to fit Gaussians to the clusters when there are sufficient statistics and does this benefit our approach?
- For which types of problems is our approach most appropriate? We demonstrated its potential in polling— an untouched area for Bayesian reasoning. Can we also leverage divergence to improve accuracy and infer the varied backgrounds of the individual contributors?

Our research aims to reduce the error that arises in existing belief aggregation approaches, improve richness and realism of the consensus model formed from many diverse opinions, and reduce noise and isolate bias. We also anticipate advancement in inferring backgrounds of individuals and encouraging individuals to support their claims with reasoning. This will help build both accurate and representative models across domains and will add to the advancement of Bayesian reasoning as a practical modeling tool.

Acknowledgements

We would like to thank Carl Stern of Management Sciences, Inc. and Joe M. Kniss at UNM for their valuable contributions to this research.

References

- Amazon. 2008. Mechanical turk. <http://www.mturk.com>.
- Clauset, A.; Newman, M. E. J.; and Moore, C. 2004. Finding community structure in very large networks. *Physical Review* 70.
- Dasgupta, S., and Schulman, L. 2007. A probabilistic analysis of em for mixtures of separated, spherical gaussians. *The Journal of Machine Learning Research* 8:203–226.
- Matzkevich, I., and Abramson, B. 1992. The topological fusion of bayes nets. In *Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence*.
- Maynard-Reid, P. I., and Chajewska, U. 2001. Aggregating learned probabilistic beliefs. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*.
- Napp, C., and Jouini, E. 2006. Aggregation of heterogeneous beliefs. *Journal of Mathematical Economics* 42(6):752–770.
- Ottaviani, M., and Sorensen, P. N. 2006. Aggregation of information and beliefs in prediction markets. Technical report, University of Copenhagen. <http://econpapers.repec.org/paper/kudkuiefr/200701.htm>.
- Pennock, D. M., and Wellman, M. P. 1997. Representing aggregate belief through the competitive equilibrium of a securities market. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*.
- Pennock, D. M., and Wellman, M. P. 1999. Graphical representations of consensus belief. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*.
- Pennock, D. M., and Wellman, M. P. 2005. Graphical models for groups: belief aggregation and risk sharing. *Decision Analysis* 2(3):148–164.
- Robins, G., and Lusher, D. 2006. Statistical models for social network structures, individual beliefs and perceived social norms. Presented at NetSci 2006.
- Sanjeev, A., and Kannan, R. 2001. Learning mixtures of arbitrary gaussians. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, 247 – 257.
- Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8).
- Steinberg, A., and Bowman, C. 2004. Rethinking the jdl data fusion levels. NSSDF JHAPL.