

Generalized Opinion Pooling

Ashutosh Garg, T. S. Jayram, Shivakumar Vaithyanathan, Huaiyu Zhu

IBM Almaden Research Center, San Jose, CA 95120, USA

Abstract

In this paper we analyze the problem of opinion pooling. We introduce a divergence minimization framework to solve the problem of standard opinion pooling. Our results show that various existing pooling mechanisms like LinOp and LogOp are a special case of this framework. This framework is then extended to address the problem of generalized opinion pooling. We show that this framework does satisfy various desiderata and we give an EM algorithm for solving this problem. Finally we present some results on synthetic and real world data and the results obtained are encouraging.

1 Introduction

The recent explosion on the web has resulted in the availability of valuable customer feedback. Ranging from movies to various products such feedback is often available in the form of explicit user ratings. Alternatively, such ratings can also be extracted from opinions expressed in text. Several recent efforts in statistical NLP on extracting such opinions from text is available [4]. The distributed nature of the internet implies that information regarding users feedback and opinions is often available from multiple sources. Further, individual experts possessing relevant information may use different models to make predictions or to generate estimates while expressing opinions. To base inference on all available information, it is necessary to combine the information from all these different experts. In this paper we consider the problem of aggregating information from multiple experts. Typically, opinions are represented in terms of probability distribution and the aim is to arrive at a single probability distribution which represents the consensus behavior. This is accomplished using a pooling or consensus operator. Studied formally under the name of opinion pooling this problem has primarily been addressed under an axiomatic framework. In such approaches a consensus operator is chosen to satisfy a required set of axioms.

This paper tackles a problem that is more complex than the conventional opinion pooling problem. Each expert opinion is characterized by some dimensions and a consensus opinion might be desired across any subset of these dimensions. Further, various simple desiderata are defined to be required of the consensus opinions. Moving away from the traditional axiomatic approach, a model-based solution is proposed to tackle the problems of consistency and sparsity introduced by this generalization. A formal analysis of the model-based consensus results in a derivation of the conditions for which the desiderata are satisfied.

The remainder of the paper begins by first revisiting the problem of conventional opinion pooling. Section 2, motivates and formally introduces the generalized opinion pooling (GOP) problem followed by a discussion on the various desiderata (Section 2.1). In order to motivate our model-based approach, in Section 3, we first cast conventional opinion pooling as a divergence minimization problem. Here it is shown that current, popular, aggregation operators arise as solutions to special cases of this formulation. In Section 4, we extend this optimization framework to GOP problem where we propose a model-based

solution. Section 5 provides an empirical study of the proposed model using opinions collected from the Web. Section 6 describes the results of our experiments followed by discussion.

2 Preliminaries and Problem Definition

Opinions about products and services can be expressed in several different ways; as ratings on a scale, or as preferences expressed via a probability distribution - e.g., over *High* and *Low*¹. The premise of this paper is that a decision maker (DM) would be interested in aggregation of opinions from different sources. Consider, for example, the following query “*What is the opinion of Thinkpad T30 as expressed at different sources ?*”. Assuming two sources, the answer to this query is some consensus of Thinkpad T30 reviews at these sources. Note that scale over which ratings are provided often vary across sources and therefore need to be normalized somehow. Also, other opinions may not be representable over a rating scale². Due to these issues it is preferable to express opinions as a probability distribution over preference values. A detailed discussion on the conversion of scale ratings to probability distributions is beyond the scope of this paper. However, the simple model used in this paper is described in a later section.

The following notations will be used throughout the paper. Capital letters X, Y will be used to refer to random variables and the corresponding small letters x, y will denote the particular instantiation (value taken) by these. $P_X(X = x)$ will refer to the probability that a random variable X takes on value x . When the context is clear, we will denote this quantity simply by $P(X = x)$ or $P(x)$. Given a set of empirical distributions $\{\hat{P}_1, \hat{P}_2, \dots, \hat{P}_N\}$, (we will reserve the “hat” notation exclusively for empirical distributions), we will refer to \hat{P}_i , for each i , as the distribution of the opinion random variable S given by expert e_i .

The (conventional) opinion pooling problem can be stated as follows:

Definition 1 (Opinion Pooling). *Experts $\{e_1, e_2, \dots, e_N\}$ provide opinions $\{\hat{P}_1, \hat{P}_2, \dots, \hat{P}_N\}$, respectively, about a particular topic. The opinion pooling problem is to provide a consensus opinion P about that topic. In other words, we seek a pooling operator F such that $P = F(\hat{P}_1, \hat{P}_2, \dots, \hat{P}_N)$.*

We now introduce a more elaborate framework where the relationships between the experts is captured while respecting certain constraints. The experts, expressing opinions, are characterized using various *dimensions* of interest. Assume that there are m dimensions D_1, D_2, \dots, D_m of interest. Suppose there are N experts e_1, e_2, \dots, e_N who provide opinions $\hat{P}_1, \hat{P}_2, \dots, \hat{P}_N$, respectively. Each expert e_i is associated with some assignment of (legal) values to an arbitrary subset of dimensions; C_i is called the *characteristic* of expert e_i . Let T denote the topic variable about which opinions can be expressed.

Given such empirical distributions from several experts, the DM may request opinions about topics across arbitrary characteristics. Note that the desired characteristic need not agree with the characteristic of any of the experts. In addition to this reporting problem, the DM may wish to analyze the relationships across different characteristics. To address such issues while ensuring consistent answers, we propose a framework in which the consensus opinion for all characteristics is obtained via a single distribution P such that the conditional probability distribution $P(S|T = t, C)$ is well-defined for every topic t and every characteristic C . Furthermore, the DM may also wish to impose additional constraints that need to be satisfied. These constraints can be incorporated into the framework by placing suitable restrictions on P . In Section 5, such constraints are naturally expressed using a statistical model.

¹At Epinions.com reviews of products are expressed as ratings on a scale of 1 – 5.

²As an example consider a study being conducted on the likelihood of a customer coming back – and the responses are either Likely, Not Likely and Undecided. Such responses are not easily expressed on a scale.

Definition 2 (Generalized Opinion Pooling). Suppose there are N experts e_1, e_2, \dots, e_N who provide opinions $\hat{P}_1, \hat{P}_2, \dots, \hat{P}_N$, respectively. Let C_i and t_i denote the characteristic and topic of expert e_i . The generalized opinion pooling problem (GOP) is to find a distribution P that can be conditioned on every topic and every characteristic, subject to the constraints imposed by the DM. In other words, we seek a pooling operator F such that $P = F(\hat{P}_1, \dots, \hat{P}_n)$, and that $P(S|T = t, C)$ is well defined for every topic t and every characteristic C .

Note that the distribution P can potentially contain a larger set of random variables apart from the dimensions, topics and opinion, e.g. it may contain latent variables³. The solution P provides opinions for all distinct characteristics thus addressing the apparent problem of sparsity - i.e., empirical distributions may not be available for all characteristics or may need to be estimated with very little data. Moreover, the imposition of a single joint distribution ensures that reporting is consistent across all characteristics.

2.1 Desiderata for Pooling Operators

In the literature on opinion pooling, there has been a considerable study of the many properties satisfied by the various pooling operators [2]. For the generalized opinion pooling problem, particularly in the context of business and market intelligence, the opinions are distributions over some set of preference values. For this domain, we have identified three simple and natural properties that are desired of any solution—

1. **Unanimity:** If all the experts agree on the opinion of a topic, then the aggregated opinion agrees with the experts.
2. **Boundedness:** The aggregated opinion is bounded by the extremes of the expert opinions.
3. **Monotonicity:** When a certain expert changes his opinion in a particular direction with all other expert opinions remaining unchanged, the aggregated opinion changes in the direction of this expert.

3 Opinion Pooling

In order to motivate our approach to GOP, we first present a simple but powerful framework for the conventional opinion pooling problem. We will show that popular operators LinOp and LogOp arise as special cases of this formulation. Later, we extend this in a natural way to the GOP problem.

The basic intuition is that in any solution to the opinion pooling problem, we expect the aggregate distribution to be as “close” as possible to the individual experts. To formalize this, we will consider distance measures between distributions and cast conventional opinion pooling as a minimization problem. To the best of our knowledge this formulation and the associated derivations have not appeared in literature.

Let $D(P, Q)$ denote a divergence measure between probability distributions P and Q , where D satisfies (1) $D(P, Q) \geq 0$ and (2) $D(P, Q) = 0$ if and only if $P = Q$. We are given n expert distributions \hat{P}_i , and their respective non-negative weights w_i which sum to one. The goal is to obtain an aggregate distribution P via the following minimization problem:

$$P = \operatorname{argmin}_Q \sum_i w_i D(P_i, Q) \quad (1)$$

The choice of weights is governed by various criteria [3]. W.l.o.g., in the absence of any knowledge, all experts will be assumed equal. Therefore all w_i are equal and hence ignored in the remainder of the paper.

³Operators such as LinOp dramatically restrict the constraints that can be imposed on the solution[5].

Divergence $D(P, Q)$	Consensus opinion $P(s)$
$D_\gamma(P, Q) = \frac{1 - \sum_x P(s)^\gamma Q(s)^{1-\gamma}}{\gamma(1-\gamma)}$	$\frac{1}{Z} \left(\sum_i w_i [\hat{P}_i(s)]^\gamma \right)^{\frac{1}{\gamma}}$
$D_{KL}(P, Q) = \sum_x P(s) \log \frac{P(s)}{Q(s)}$	$\sum_i w_i \hat{P}_i(s)$
$D_{KL}(Q, P) = \sum_x Q(s) \log \frac{Q(s)}{P(s)}$	$\frac{1}{Z} \prod_i [\hat{P}_i(s)]^{w_i}$
$L_2(P, Q) = \sum_x (P(s) - Q(s))^2$	$\sum_i w_i \hat{P}_i(s)$
$\chi^2(P, Q) = \sum_x \frac{(P(s) - Q(s))^2}{Q(s)}$	$\frac{1}{Z} \left(\sum_i w_i / \hat{P}_i(s) \right)^{-1}$
$\chi^2(Q, P) = \sum_x \frac{(P(s) - Q(s))^2}{P(s)}$	$\frac{1}{Z} \sqrt{\sum_i w_i [\hat{P}_i(s)]^2}$

Figure 1: Different divergences and the corresponding consensus pooling operator. The quantity Z denotes the normalization constant.

Table 1 gives a summary of different divergences and the consensus distributions that arise by solving the associated minimization problems. Derivations are done using standard analytical methods and omitted in the interest of space. Two interesting cases are

1. **LinOp:** F is called *LinOp* if P can be expressed as a linear combination of the empirical distributions. Choosing either KL-distance or L^2 norm as the divergence measure in Equation 1 leads to this solution for the consensus distribution.
2. **LogOp:** F is called *LogOp* if P is the weighted geometric mean of the empirical distributions under consideration. Choice of reverse KL-distance (see Figure 1) leads to LogOp as the consensus distribution.

Having a closed form solution allows us to directly evaluate the different divergence measures via the desiderata stated earlier. First, we observe that LinOp satisfies unanimity, – for any fixed s , if $\hat{P}_i(s) = c$ for all i , then $P(S = s) = c$ and boundedness – for every s , $\min_i \hat{P}_i(s) \leq P(s) \leq \max_i \hat{P}_i(s)$ which follows easily by its definition. LinOp also satisfies a strong monotonicity property: suppose expert e_i changes his opinion \hat{P}_i to \hat{Q}_i and suppose that all other experts’ opinions are unchanged. Let P and Q be the LinOp solutions before and after e_i ’s opinion has changed. Then for every s , $Q(s) > P(s)$ (respectively, $<$, $=$) if and only if $\hat{Q}_i(s) > \hat{P}_i(s)$ (respectively, $<$, $=$).

For the pooling operators arising from other divergences, it is possible to construct easy counterexamples showing that none of them satisfy unanimity or boundedness. However, they all satisfy a weak form of monotonicity. This is shown below for the case when the divergence measure is D_γ . For other divergence measures, a similar result can be shown using the same technique.

Theorem 3. *Suppose expert e_i changes his opinion \hat{P}_i to \hat{Q}_i such that $\hat{P}_i(s) < \hat{Q}_i(s)$, for some s , while $\hat{P}_i(s') \geq \hat{Q}_i(s')$ for every $s' \neq s$.⁴ Suppose that all the other experts’ opinions are unchanged i.e. $\hat{Q}_j = \hat{P}_j$ for all $j \neq i$. If P and Q are the solutions using D_γ as a divergence before and after expert e_i ’s opinion has changed, then $Q(s) > P(s)$.*

Proof. Define $\tilde{P}(x) = (\sum_i [\hat{P}_i(x)]^\gamma)^{1/\gamma}$, and $\tilde{Q}(x) = (\sum_i [\hat{Q}_i(x)]^\gamma)^{1/\gamma}$, for all x . Since $\hat{P}_i(s) < \hat{Q}_i(s)$, we have $\tilde{Q}(s) = \tilde{P}(s) + \epsilon_s$, for some $\epsilon_s > 0$. Similarly, for every $s' \neq s$ we have $\hat{P}_i(s') \geq \hat{Q}_i(s')$ implying

⁴A dual situation is when the opinion for s decreases while the opinion for the remaining $s' \neq s$ is non-increasing, which can be handled similarly.

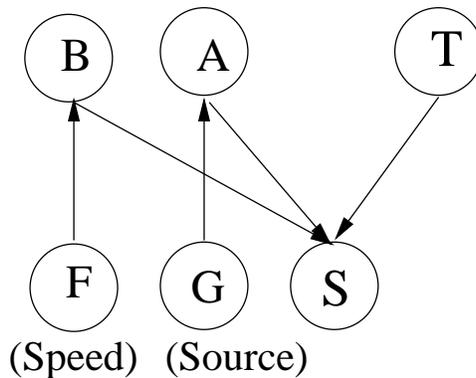


Figure 2: Bayesian network for Generalized opinion pooling.

$\tilde{Q}(s') = \tilde{P}(s') - \epsilon_{s'}$ for some $\epsilon_{s'} \geq 0$. Moreover, $\epsilon_{s'}$ is strictly greater than 0 for at least one $s' \neq s$ because $\hat{P}_i(s) < \hat{Q}_i(s)$ implies that $\hat{P}_i(s') > \hat{Q}_i(s')$ for at least one $s' \neq s$. Therefore, $\sum_{s' \neq s} \epsilon_{s'} > 0$.

From Table 1, we note that $P(x) = \tilde{P}(x)/Z$ and $Q(x) = \tilde{Q}(x)/Z'$, for all x , where $Z = \sum_x \tilde{P}(x)$ and $Z' = \sum_x \tilde{Q}(x)$ denote the normalization constants for P and Q , respectively. From the previous paragraph, it follows that $Z' = Z + \epsilon_s - \sum_{s' \neq s} \epsilon_{s'}$. Now, since $Z = \sum_x \tilde{P}(x) \geq \tilde{P}(s)$, we have

$$Q(s) = \frac{\tilde{Q}(s)}{Z'} = \frac{\tilde{P}(s) + \epsilon_s}{Z + \epsilon_s - \sum_{s' \neq s} \epsilon_{s'}} > \frac{\tilde{P}(s) + \epsilon_s}{Z + \epsilon_s} \geq \frac{\tilde{P}(s)}{Z} = P(s). \quad \square$$

4 A Model Based Solution to Generalized Opinion Pooling

The previous section presented a divergence minimization framework for opinion pooling together with an analysis of the conditions under which the various desiderata are satisfied. In this section, the framework is extended to the GOP problem. Recall that the solution to the GOP problem is a single distribution P such that the consensus opinion for every characteristic and every topic can be obtained as a conditional probability distribution, subject to the constraints imposed by the DM. Let \mathcal{P} denote the feasible set of solutions such that constraints imposed by the DM are satisfied. For each expert e_i , let P_i denote the distribution P conditioned on the topic and the characteristic of that expert. In other words, if the topic and characteristic of e_i are t_i and C_i respectively, then $P_i(s) = P(s|T = t_i, C_i)$. A natural generalization of the optimization approach considered for conventional opinion pooling (Equation 1) is that the empirical distribution \hat{P}_i of expert e_i be close to the distribution P_i , for each i :

$$\text{minimize } \sum_i w_i D(\hat{P}_i, P_i) \quad \text{such that} \quad P \in \mathcal{P} \quad (2)$$

We now address whether this solutions to the GOP problem satisfies the desiderata described in Section 2.1. We will show that under suitable conditions, indeed the minimization problem of Equation 2 satisfies unanimity, boundedness and monotonicity. First, to prove a monotonicity result for D_γ , we consider the following setup where we have two sets of empirical distributions as inputs to the minimization problem of Equation 2. We show that the difference between the two empirical distributions is positively correlated with the difference between their corresponding minima.

Lemma 4. Let $\widehat{P}_1, \dots, \widehat{P}_n$, and $\widehat{Q}_1, \dots, \widehat{Q}_n$ be two sets of empirical distributions and suppose P and Q are the corresponding distributions, obtained by solving Equation 2 using the divergence D_γ . Then,

$$\sum_i \sum_s [(\widehat{P}_i(s))^\gamma - (\widehat{Q}_i(s))^\gamma] \cdot [(P_i(s))^{1-\gamma} - (Q_i(s))^{1-\gamma}] \geq 0.$$

Proof. Let $D = D_\gamma$. Since P is a minima for $\widehat{P}_1, \dots, \widehat{P}_n$, we have $\sum_i D(\widehat{P}_i, P_i) \leq \sum_i D(\widehat{P}_i, Q_i)$ and $\sum_i D(\widehat{Q}_i, Q_i) \leq \sum_i D(\widehat{Q}_i, P_i)$. Adding the two equations gives $\sum_i D(\widehat{P}_i, P_i) + D(\widehat{Q}_i, Q_i) - D(\widehat{P}_i, Q_i) - D(\widehat{Q}_i, P_i) \leq 0$. Substituting the definition of D proves the theorem. \square

Theorem 5. Suppose expert e_i changes his opinion \widehat{P}_i to \widehat{Q}_i such that $\widehat{P}_i \neq \widehat{Q}_i$ and all other experts' opinions are unchanged i.e. $\widehat{Q}_j = \widehat{P}_j$ for $j \neq i$. Let P and Q be the solution obtained via D_γ before and after expert e_i 's opinion has changed. Let $A = \{s : \widehat{Q}_i(s) > \widehat{P}_i(s)\}$ and $B = \{s : \widehat{Q}_i(s) < \widehat{P}_i(s)\}$ Then, either for at least one $s \in A$ we have $Q_i(s) \geq P_i(s)$, or for at least one $s \in B$ we have $Q_i(s) \leq P_i(s)$.

Proof. Note that the sets A and B are nonempty because $\widehat{P}_i \neq \widehat{Q}_i$. Suppose the theorem does not hold i.e. $Q_i(s) < P_i(s)$ for all $s \in A$ and $Q_i(s) > P_i(s)$ for all $s \in B$. It follows that the LHS of the inequality of Lemma 4 is strictly negative—a contradiction. \square

For unanimity and boundedness we have the following result, for divergence D_{KL} , which assumes that class of distributions \mathcal{P} satisfies certain conditions.

Theorem 6. Let P be the distribution obtained by solving Equation 2 via D_{KL} . Then it satisfies the following conditions: (1) *Unanimity:* For any fixed s , if $\widehat{P}_i(s) = c$ for all i , then $P_i(s) = c$. (2) *Boundedness:* For every s , $\min_i \widehat{P}_i(s) \leq P_i(s) \leq \max_i \widehat{P}_i(s)$.

Proof. We provide a constructive proof for the above theorem. It shows that under certain existential conditions, unanimity condition is satisfied when the KL distance is used as the divergence measure.

Define P' such that $\forall i, P'_i(s) = c$ and $\forall s' \neq s; P'_i(s') = \frac{(1-c)}{(1-P_i(s))} P_i(s')$. We assume $P_i(s) \neq 1$ as otherwise the original KL divergence would be infinity when $P_i(s) = 1$ and $c \neq 1$.

Now if $P' \in \mathcal{P}$ (this is the existential condition) then one can show $\sum_i D_{KL}(\widehat{P}_i, P_i) > \sum_i D_{KL}(\widehat{P}_i, P'_i)$. This proves the unanimity condition. A similar argument can be used to prove the boundedness result. \square

5 Bayesian Network Aggregation

The details of the statistical model describing P_i , in equation 2, was conveniently ignored in the previous section. Recall from the previous section that the distribution of interest is the joint distribution over the random variables. Moreover, the constraints of the DM, represented as conditional dependency between the random variables, must also be modeled. A convenient representation is a Bayesian Network (BN) that captures, intuitively, the essential aspects of the problem. By varying the conditional independence relationships modeled and by the incorporation of hidden variables, the BN allows for a rich class of constraints that DM would like to impose. Once the parameters of the BN are learned it can then be queried by the DM to obtain aggregated opinions of interest. However, the complexity of the problem (learning and inference) will depend on the particular choice of network structure.

5.1 Description Of the Model

We illustrate the Bayesian network approach using a simple example. Assume that the DM is interested in opinions about laptops expressed at multiple sources. Therefore, the topic T equals laptops t , and the characteristic includes the dimension source G which takes on values $g \in \mathcal{G}$. To adequately explain the power of the BN we will assume another dimension, Speed (processor speed), F which takes on values $f \in \mathcal{F}$. User ratings can be interpreted as empirical distributions $\hat{P}(s|t, g, f)$. A BN instantiation of this example, given below, sheds more light on the learning problem. The dependency structure of the BN is assumed to be defined by a domain expert and represents the constraints of the underlying problem. Figure 3(a) shows the BN under consideration⁵. Besides the dimensions, topic T , source G and speed F there are latent variables A and B which capture the behavioral similarities exhibited by populations across the different characteristics while tackling sparsity.

Let Θ denote the set of all parameters of the network i.e. the (conditional) probability tables associated with all the nodes of the network. It is assumed that the probabilities $P(G|\Theta)$ and $P(F|\Theta)$ (the prior probabilities for the individual characteristics) are known e.g., a simple estimate is the percentage of data available for each of these variables. The remaining parameters are to be learned using available empirical distributions. These empirical distributions are over opinions for different topics conditioned on different dimensions. In particular, assume that the following empirical distributions were observed: $\hat{P}(S|t, g_i, f_i)$ for $i = 1, \dots, N$ where N being the number of experts (empirical distributions observed) and (g_i, f_i) be their corresponding characteristics. The parameter learning problem for the Bayesian network can be cast as the following optimization problem.

$$\Theta = \underset{\Theta}{\operatorname{argmin}} \sum_i D_{KL}(\hat{P}(S|t, g_i, f_i), P(S|t, g_i, f_i, \Theta))$$

$$\text{such that } P(S|t, g_i, f_i, \Theta) = \sum_{a,b} P(S|a, b, t, \Theta)P(a|g_i, \Theta)P(b|f_i, \Theta) \quad (3)$$

Simply stated this objective function attempts to minimize the divergence between the learned conditional probability distribution and the observed conditional probability distribution. Since the parameters of the BN are estimated from available empirical distributions the objective function above is different from the usual maximum likelihood (ML) learning of Bayesian networks. However, an EM algorithm [1] can still be derived to obtain the estimates of Θ .

Expanding equation 3 and ignoring the constant term the optimization problem is given as

$$\Theta = \underset{\Theta}{\operatorname{argmax}} \sum_{S,i} \hat{P}(S|t, g_i, f_i) \log P(S|t, g_i, f_i, \Theta)$$

The E-M algorithm for the above objective function can now be written. Let Θ_k denote the estimates of Θ at the k -th step of the algorithm. We have –

E Step: Compute $Q(a, b|S, t, g_i, f_i) = P(a, b|S, t, g_i, f_i, \Theta_k)$

M Step: Maximize $\sum_{S,t,i,a,b} Q(a, b|S, t, g_i, f_i) \hat{P}(S|t, g_i, f_i) \log P(S|a, b, t, \Theta)P(a|g_i, \Theta)P(b|f_i, \Theta)$

Imposing appropriate constraints leads to the following update equation

$$P(S|t, a, b, \Theta_{k+1}) = \frac{\sum_i P(a, b|S, t, g_i, f_i, \Theta_k) \hat{P}(S|t, g_i, f_i)}{\sum_S \sum_i P(a, b|S, t, g_i, f_i, \Theta_k) \hat{P}(S|t, g_i, f_i)}$$

The update equations for other parameters can be obtained in similar fashion.

⁵The analysis of model-based approaches to opinion pooling presented in Section 4 do not depend on the structure of the BN.

6 Experiments

In this section we describe some experiments to validate the approach presented in this paper. There are two main tasks that one intends to solve by opinion pooling - "reporting" and "analysis" of data. In practice the obtained opinions - in the form of empirical distributions - is often sparse (the space of possible set of characteristics is huge and we may not have data for all possible combinations) and that makes the use of model based approach both necessary and interesting. The evaluation of the presented approach is primarily divided into two categories a) robustness to data sparsity, b) ability of the model to capture behavioral similarities across dimensions. We give results of experiments on both synthetic data – as it provides a more controlled environment allowing for better study, and on real data – using opinions gathered from the Web. The model is validated against LinOp.

6.1 Synthetic Data

The BN model from which data was sampled is a joint distribution over 4 random variables $\{S, A, G, T\}$ with the factorization $P(S, A, G, T) = P(T)P(G)P(A|G)P(S|A, T)$. The synthetic data was generated for a single topic (T) from 10 geographical locations - i.e., (G) can take on one of 10 different values. The opinion of the topic is represented by random variable S which also (confusingly) can take on 10 different values (1 – 10). The data was generated to reflect three hidden behaviors optimistic, pessimistic and unbiased. For optimistic behavior there is a greater probability mass over higher values of the opinion while the converse is true for pessimistic behavior. A uniformly distributed probability mass reflects an unbiased behavior. Fig. 3(a) shows the distribution of the behaviors from which the synthetic data was sampled. A total of 10 experts were assumed (i.e., a total of 10 opinions) - one from each geographical location. Moreover, each geographical location is associated with one of the behaviors. Specifically, three geographical regions were assumed to have an optimistic behavior, 3 regions pessimistic and the remaining 4 unbiased. For each expert, a 1000 data points (i.e., a 1000 opinion values) were sampled from the appropriate behavior (based on geography). The empirical distribution of these 1000 points was taken to be the expert's opinion.

Learning was accomplished using the EM algorithm (c.f. Section 5 with all parameters initialized randomly. To test the sensitivity of the algorithm (against overfitting) the latent variable A was run with a cardinality of 4 (recall that the ground truth has cardinality 3- the distinct behaviors). Fig. 3(b) shows the learned mixture coefficients $P(a|g)$. Note that the algorithm did learn the existence of three main behaviors indicated by overlap between the class 3 and 4 on the right side. Upon examination $P(S|a = 3)$ and $P(S|a = 4)$ (not shown in interest of space) were found to be very similar.

To test robustness to sparsity, opinions (empirical distribution) were generated for 2 distinct topics (1 & 2), 10 geographical locations and identical behavior (same as the one in previous setting). The learning algorithm was allowed only a portion of the opinions for parameter estimation. Specifically, for Topic 1, opinions from all geographic locations were used, while for Topic 2, from only 5 geographical locations were used. Fig. 3(c) shows the learned distributions of opinion (averaged over all locations) for each of the two topics. Note that for Topic 1 the results of LinOp and model-based approach are identical. However, for Topic 2 there is a difference in performance between the two approaches. The model-based approach generalizes significantly better than LinOp. This is evident from the resulting distribution for the model based approach being closer to the one of Topic 1 (whose behavior is identical to Topic 2).

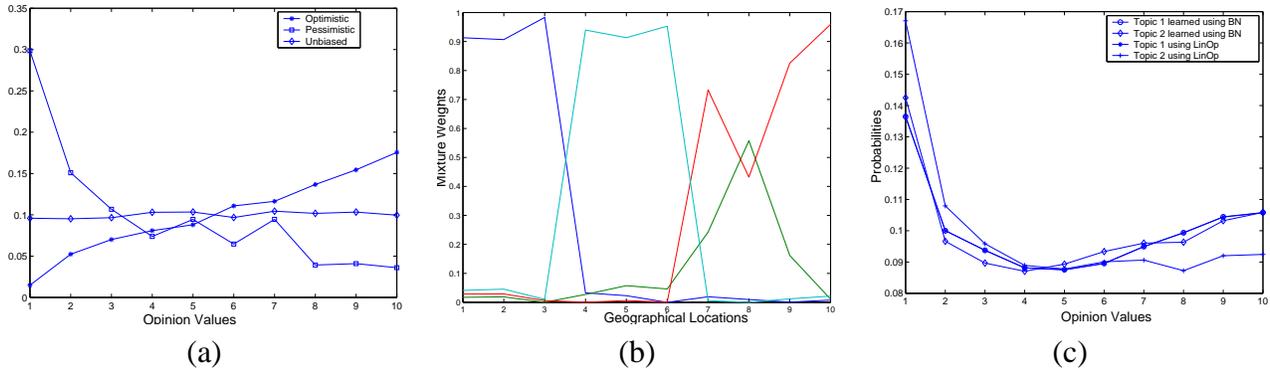


Figure 3: (a) Optimistic, pessimistic and unbiased behaviors. (b) Plot of mixture coefficients $P(a|g)$. (c) Results of the sparsity experiment

Table 1: Prediction of queries not present in the training set.

Query	Source	Brand	Model	Speed	$\hat{P}(S Query)$	$P(S Query)$
Query 1	Epinions	HP		266MHz	[0.9 0.1]	[0.89 0.11]
Query 2	ZDnet	Sony	Vaio	667MHz	[0.8 0.2]	[0.81 0.19]

6.2 Real Data

The second set of experiments was conducted on the real data consisting of opinions about different *laptops* collected from several sources on the Web namely *Epinions*, *Cnet*, *ZDnet*, and *Ciao*. Each laptop in reality is described by several dimensions (possibly tens). To make the experiments manageable only company name, model and processor speed are considered here. A total of 2180 opinions, $\hat{P}(\cdot)$, with 108 distinct characteristics, were collected from the different sources. The structure of the BN was chosen based on expert knowledge (details omitted in the interest of space).

Each opinion is expressed as a rating over a scale of either 1-5 or 1-7. These ratings were converted into a distribution over the space *High* and *Low* assuming the following simple probability model. Each rating was converted into a corresponding percentage. This is interpreted as the probability that a random reader will classify the corresponding review as *High*. Note that more complicated probability models can be used to convert ratings into more complex probability distributions.

To evaluate model robustness to data sparsity, the dataset was divided into 70-30 training/test split. For each characteristic ground truth was defined by applying LinOp over all opinions (ignoring the split) sharing this characteristic. The BN was learned using 70% of the data. For comparison a LinOp based consensus opinion was obtained for each characteristic using the appropriate opinions from the training split. The average KL distance between the ground truth and model-based approach was 0.0302 whereas the KL distance between ground truth and LinOp was 0.0439. The average was taken over all possible values of characteristics. This suggests that indeed there is information to be learnt from other opinions while providing an aggregate opinion.

Sometimes the queries of the DM may involve characteristics for which opinions may not be available in the training set. To test the predictive ability of the model the BN was tested on opinions that do not contain characteristics in the training set. Note that LinOp cannot provide an answer in such cases. Table 1 shows the results of this experiment.

Table 2: KL divergence between all pairs of $P(A|Source)$.

	Epinions	Cnet	ZDnet	Caio
Epinions	-	0.3425	0.4030	0.466
Cnet	-	-	0.111	0.3757
ZDnet	-	-	-	0.0867

Table 2 shows the symmetric version⁶ of KL-divergences between all pairs of $P(A|Source)$. The divergence between sources Caio and ZDnet is the lowest – and they are both based out of UK while the remaining two operate out of US. This interesting, albeit anecdotal, observation might be interpreted as sources exhibiting behavioral similarities.

7 Summary

In this paper we introduced a generalized opinion pooling framework for synthesizing unstructured data, with an application to business intelligence reporting. The opinion pooling problem is cast in the form of a constrained divergence minimization problem. In contrast to conventional opinion pooling where a single consensus opinion is sought from a collection of expert opinions, our framework allows the consensus opinion to take into account varying characteristics of the experts. The degree to which the differing characteristics are taken into account can be controlled by the constraints. Under reasonable conditions several desiderata are satisfied. The constraint can be implemented by some statistical models, such as Bayesian networks. We explain the training of such networks from empirical data. Finally, we presented experiments validating our approach using both synthetic data and real data.

References

- [1] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B, 1977.
- [2] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography (avec discussion). *Statistical Science*, 1:114–148, 1986.
- [3] P. Maynard-Reid II and U. Chajewska. Aggregating learned probabilistic beliefs. In *UAI*, pages 354–361, 2001.
- [4] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.
- [5] D. M. Pennock and M. P. Weliman. Graphical representations of consensus belief. In *Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence (UAI-99)*, pages 531–540, 1999.

⁶The symmetric version of the KL-divergence between two distributions p and q is given as $KL(p,q)+KL(q,p)$