# ====-- 一种新的基于证据理论的合成公式

孙全 叶秀清 顾伟康

## 主要内容

* 较为详细的分析了与合成公式的优缺点
* 引入证据可信度的概念
* 提出新的合成公式

## 内容探讨

### 合成公式定义

设和分别是同一识别框架下的基本置信度指派函数，焦元分别为和，假若存在映射满足

当：



当：



此处为基本置信度指派函数，，其大小反应两个证据源的冲突程度。

### 合成公式的缺点

对于合成公式，当为1的时候该合成公式便不可用，且当非常接近1的时候得出的结果将是有悖于常理的。

例1：存在一个辨识框架，：，，；：，，；

根据合成公式可得，







注意：文中计算得到的，，，。计算有误，得出的结果是尽管对的支持非常低，但合成结果却认为正确的，该示例无法直接说明合成公式的不足点，需要另寻实例。

### 的合成公式







合成公式直接将冲突部分放入未知领域，而未分配给命题中的任何部分。可以看出，冲突的部分在该合成公式中完全被抛弃，这使得原来信任度低的命题在合成后得到的命题仍然非常低，避免了因值的过大导致得到合成结果产生悖论。

### 合成公式的不足

合成公式的最大一个不足之处即当存在多个证据源时，虽然绝大多数证据源对某一命题的支持度都非常高，但若存在一个证据源对该命题的支持度极低时将使得合成的结果对该命题的支持度变得极低。

### 新的合成公式

首先定义证据可信度，假设对应的证据集为，设证据集和之间的冲突大小为，可得



此时，证据可信度可为，而，其中，为证据源的个数，从一定程度上代表了证据的总体冲突程度，是的减函数。可知，证据源越多的值将越小，证据的可信度将越大。

新的合成公式定义如下：







其中：

，

对合成公式进行变形有：



该合成公式其实是一个加权和的形式，和为加权系数。当较小时该式的前一项占主导地位，当较大时后一项占主导地位。

## 本文中存在的问题

### 举例问题

在指出合成公式在冲突非常大的情况下得到的合成结果可能是错误的中，所提出的案例结果计算有误，该案例不足以证明出该中结论。

### 证据可信度定义问题

在证据可信度的定义中，本文定义为随着证据源个数的增加，证据的可信度将增大。实际上当证据源数量有限时，证据的可信度不一定会增大，本文并未给出较为详细的说明与推导。

## 问题改进

### 针对举例问题的改进

存在一个识别框架，其中；，，；

根据合成公式可得，，。从所给的案例中可以看出虽然这两个证据源对命题的信任度都非常小，但是根据合成公式得到的信任度却为1，这显然是不对的。

### 针对证据可信度定义问题的改进

## 总结与反思

# Belief function combination and conflict management

E.Lefevre,O.Colot,P.Vannoorenberghe

## 主要内容

* 简单的介绍证据理论
* 提出一个普遍的框架来统一诸多证据融合规律
* 提出其他的融合法则来处理冲突的分配问题
* 提出冲突分配到各命题的权值确定的方案

## 内容探讨

### 证据理论基本概念

定义一个集合为识别框架如下：



由该集合的子集构成的集合用表示：



证据理论的一个关键点即为置信度赋值(basic belief assignment)，类似于可能性赋值(probability assignment)，我们需要将权值分配到集合的每个部分，但是不同的是置信度赋值不仅要考虑单元素子集，还要考虑多元素子集。

将信任度分配到一个证据源定义如下：







信任度函数(belief function)定义如下：



似真度函数(plausibility function)定义如下：



### Dempster’s rule of combination

定义为融合了条证据源后得到的信任度函数，如下所示：



当融合两个证据源和时，融合规则如下：







产生的原因有如下几点：

1. 证据源本身数据有误，如传感器在接受到信息的时候收集或者处理产生错误。
2. 绝大多数信任度分配函数都是依照以距离为参照的相邻信息(neighbourhood information according to a distance)或者可能性函数(likelihood functions)，而对这两个方法的不恰当或者不准确应用将导致冲突的产生。
3. 随着证据源个数的增加，的值也将增大。

针对证据融合过程中产生的冲突，存在两种解决方案。其一为假设所提供的证据源都是合理的没有误差，冲突产生的最主要问题即信任度函数定义问题。Dempster和Smets对此提出了D-S合成公式，Smets保留下但是在归一化中并没有用到该值。其中Smets定义的证据融合公式如下：





其二为针对不完整或者不清晰的证据源，Yager、Dubois和Prade认为识别框架是一个开放的世界体系，Yager将分配到整个集合上面，即冲突产生的原因是识别框架不完整。定义的融合公式如下：





Dubois和Prade定义的融合公式如下:



另一种处理方法即是对不完整或者不清晰的证据源加入衰减因子，以此代表证据源的信任程度，定义如下：





## Generic framework

### 整体框架

合成公式的核心即是将分配到命题集合中，而每个子集所分配到的冲突按照权值因子来分配，其中.因此，最终合成后的信任度分配如下所示：



其中，表示冲突分配到的值，定义如下：







### Smets和Yager组合公式重构

Smets并不将分配给证据合成过程中的任何命题，即完全摒弃冲突带来的影响，故可理解为将分配给空集。



故将冲突分配给空集的权值定为1：



Yager认为冲突产生的原因是由于识别框架的不完整导致的，故是将分配到命题的全集上，即。可表示如下：



同时，冲突分配给全集的权值为1：



文章原话摘抄：

This method involves the separation of the whole confliction mass and furthermore, implies that it participates in the decision process for distinguishing the hypotheses. This rule of combination id commutative but not associative. It is therefore necessary to define an order for the fusion process.

### Dempster证据合成公式重构

Dempster将冲突分配到全集上去（即），故有：



同时，冲突分配的权值表示如下：



因此，Dempster合成公式有如下形式：



### Dubois和Prade合成公式重构

原文摘抄：Each information source , with , gives a degree of belief to each focal element belonging to . When the focal elements are compatible, that is to say when the intersections between these subsets in  are not empty, the mass product assigned to these sets is assigned to the intersection. If the propositions are incompatible, that is to say when their intersection is equal to the empty set, a partial conflict denoted  appears.

定义为部分冲突(a partial conflict),定义如下：





假使证据源支持的一个命题，分配的信任度为；证据源支持的一个命题，分配的信任度为。假设此两个命题没有交集，即，虽然无法断定哪个证据源是正确的，但对于两条证据源的合成，该两个命题中必有一个是正确的。

故部分冲突可如下表示：





而将冲突分配给是以来确定的，其中，该权值的确定如下：



## 权重自学习模型

Pignistic概率距离是以识别框架U下各子集的最大距离作为证据距离，Pignistic概率定义如下：



其中，是的子集的基数；是的子集的基数；

根据Pignistic概率距离和成员指示器之间的均方误差（反应估计量与被估计量之间的关系）为评价因子，定义如下：



其中，表示训练向量与Pignistic参数之间的隶属情况，若为的子集则为1，否则为0.

理解：is the number of elements，表示为整个识别框架构成的集合内的元素的个数。

表示该识别框架中命题的个数，即单子集的个数。

## 本文中可能存在的问题

### Dempster合成公式的重构问题

文章中说明Dempster’s rule将冲突分配给全集，原文内容为：Dempster’s rule of combination is defined as follows. The set on which the conflicting mass is distributed is , so:



但是本人认为按照Dempster’s rule 所说明的，其是将冲突分配给除了之外的的集合，即：



## 本人存在的疑问

### 训练集向量的均方误差定义问题

文章中定义的训练集向量的均方误差如下：



其中，表示的pignistic可能性的实际含义不清楚，以至对该均方误差公式的定义不理解。

# Are alternatives to Dempster’s rule of combination real alternatives? Comments on “About the belief function combination and the conflict management problem”---Lefevre et al

Rolf Haenmi

## 主要内容

主要是都对文章《Belief function combination and conflict management》中若干论点进行分析并提出质疑。

## 内容探讨

### 质疑一——权重因子在存在大规模的信任函数的情况下计算将变得很难

根据文章中所提供的权重因子自学习法可以看出，当辨识框架里面的命题较少时对权重因子的计算量并不是十分大。但是，辨识框架的扩展将导致该计算的计算量呈指数状态增长。

### 质疑二——权重因子在辨识框架的子集较多时计算也将变得很难

貌似与质疑一重复，都是辨识框架如果较大时计算量将非常大。

### 质疑三——没有对多变量辨识框架相关情况进行分析

原文：it is furthermore common to have multi-variate frames of discernment, where there are not only exponentially many possible subsets, but where the sizes of the frames themselves depend exponentially on the number of variables involved.

不理解多变量辨识框架的含义。

### 质疑四——统一操作算子不满足结合性

对于多变量集合(multi-variate setting)，该信任函数是完全满足Shenoy’s axiomatic framework of valuation networks的，而结合性是满足局部计算(local computations)的一个特性。在存在大规模的信任函数的算例中，局部计算将变得格外重要，而文章提出的统一结合算子不满足结合性对这种情况是矛盾的。

### 质疑五——作者主观上不赞同引进其他的证据融合算法

作者认为Dempster’s rule of combination 已经是足够简洁和经典了，对于会出现反直观的结果应从辨识框架等数据源部分找问题等等之类的。

## 本人存在的疑问

### 多变量辨识框架

文章中透露多变量辨识框架的大小取决于这些变量，这些变量影响机理不清楚。

### Shenoy’s axiomatic framework of valuation networks and the local comutations

尚未接触到相关内容，对此两者不太了解。

## 总结与反思

# Combining belief functions when evidence conflicts

Catherine K.Murphy

## 主要内容

* 提出基于证据可信度的合成方法在多证据源的情况下存在的问题，并介绍目前存在的几种解决方案以及存在的局限性。
* 为了满足收敛性的条件，本文在合成理论中引进平均可信度(average belief)。

## 内容探讨

### D-S合成公式存在的问题

* D-S合成有可能因为某个证据源的否定而使得信任度原本很高的命题变成低信任命题
* 当某证据源指向单一命题是时，将导致“ignorance”区间彻底消失
* 当要处理的识别框架基数非常大时，信任度分配出错的概率也将非常大

### 存在的解决方案

* 去除归一化，即允许将信任度分配到空集中
* 将空集的信任度再分配到全集中
* 定义平均信任度分配方案，即对合成公式进行重定义，公式如下：



* 定义信任度区间来对证据合成后的信任度进行模糊确定，其中





### 本文中的解决方案

首先排除将空集中的信任度分配到全集中的方法，因为其操作是不满足交换律的。在多条证据源的组合中随意交换任一组合顺序都将导致最终结果的不同，同时这种方法将不断使得结果尽早趋向于稳定，导致后续的组合效果衰减。

可知，当存在某个证据源对信任度大的命题的信任程度非常小的时候，组合出的该命题将是低信任度命题。对于这个问题，定义信任度区间的方法则完全不合适，因为其下限信任度是趋于0的。与此相反，使用平均信任度的方法将变得更加理智。

使用D-S理论对多条证据源进行合成时，在各证据源信任度的偏重不同时将导致合成后的结果很难趋于稳定，抑或是冲突所获得的信任度将变得非常大，但使用平均信任度将使得合成后得到的结果更加收敛。

问题：两个证据源对的信任度都为0.7合成后得到的值是小于信任度分别为0.5和0.9。

分析一个合成后将所有信任度分配到原证据源中信任度非常小的命题中的实例，可知未归一化之前其冲突分配的信任度非常大，归一化之后将整个信任度分配到了原两个证据源中信任度非常小的命题。同时，平均信任度方案得到的分配结果比较中肯，即合成后的信任度分配并无较强偏移。

分析存在独立信任度(individually assigned mass)和共享信任度(shared masses)造成最终的合成结果存在不合理处的案例。

问题：Averaging followed by D-S combination gives:







是如何计算得到的？

## 总结与反思

本文主要是针对D-S合成理论中所存在的三个主要问题进行了一系列的分析，通过实际的案例分析现在已经存在的解决方案，诸如移除归一化，允许将信任度分配到空集中、引入平均信任度到D-S合成里面、使用上下限信任度等等。经过分析，作者认为引入平均信任度可以较好的解决现D-S合成理论所存在的若干问题，同时循环使用平均信任度可以更好的使分配的信任度更加集中且收敛更快。

# A new distance between two bodies of evidence

Anne-Laure Jousselme,Dominic Grenier,Eloi Bosse

## 主要内容

* 引入principled distance来表示两个BPAs或者证据体之间的关系。
* 给出BPA的几何解释以及证明principled distance可满足度量的所有要求
* 说明该距离与George and Pal提出的Dempster冲突量化的关系
* 与Fixsen和Mahler提出的MOP进行比较

## 内容探讨

### 两BPAs之间的距离

定义两BPAs之间的距离为



其中为一个的矩阵（的具体含义），利用Jaccard相似指数可定义单个（矩阵是否与和存在对应关系，如何对应？）

文中再次定义两BPAs之间的距离如下：



同时，又可写成如下形式：（如何推导得？ ）



其中



### 距离公式的收敛性

对于某个辨识框架，是存在特定的解的，我们假设这个特定的解为（注意次解只为单子集），其中可为单子集亦可为多子集，此解也构成了一个BPAs。

当利用合成算法得到的结果向正确结果收敛时是，存在



给出具体距离表达式，



疑问：该操作算子是否具有收敛性，即，这对整体理论体系是否存在影响。

## 总结与反思

本文最大的贡献即是引进并定义了两个BPAs之间的差距程度，并用“距离”来量化这种差距程度。根据这定义可以很好的体现出合成算法计算后结果的收敛情况，同时可进一步比较精确体现合成过程中冲突的变化情况。（本文中许多公式推导及应用还不是十分理解，还需进一步阅读）

# Combining belief functions based on distance of evidence

Deng Yong,Shi WenKang,Zhu ZhenFu,Liu Qi

## 主要内容

* 对average method进行了改进，依照证据源之间的距离来构造证据体的权重系数。从而替代原average method中平均权重的分配，该改进模型可以更好的体现不同证据体对结果的不同影响。

## 内容探讨

### 权重系数的确定

定义两个证据体的相似度为



存在个证据源，得到的每个证据源相对的权重为



其中



以此系数代替原average method中的平均值，即可得到更加合理的合成结果。

## 存在的疑问

### 证据体的相似度是否具有可加性

本文中证据体的相似度是按照归一化后的证据源之间的距离定义而成，该定义得到的相似度是否具有完成四则运算的能力。

### 如何理解单一证据体对整体合成过程的贡献

在考虑证据体之间的冲突会对合成结果产生不利的影响，本文依照证据体之间的距离对不同证据体进行加权。但是加权系数是在合成之前处理出来的，但是考虑到合成过程中其冲突是会变化的，尤其是合成后产生新的冲突，因此是不是可以考虑将这个思想作用到多证据合成过程当中。

# Analyzing the degree of conflict among belief functions

Weiru Liu

## 主要内容

* 提出同时利用和构成的冲突评价因子对证据合成中将产生的冲突进行分析

## 内容探讨

定义两个分别组成的距离中子集之间的最大差距为the distance between betting commitments of the two bbas.



而为



以此构成了冲突评价因子，由于单个评价因子在某些特定的合成中会产生错误的冲突预测。即当非常小的时候表明合成产生的冲突也非常小，但实际的冲突是非常大的，从而导致错误使用合成公式得到不正确的结果。但是如果使用和的组合进行冲突预测，其结果的准确的相比使用单一标准是高很多的，当两者都比较小时表明合成冲突小，当两者都非常大时，表明合成冲突大。

## 存在的疑问

### 证据源不独立情况下的合成

文章中曾提起到证据源获取到的证据体独立性的概念，但当确知获得的两个证据体是不独立的情况下该如何对这两个证据体进行处理，当这两个证据体是不完全独立的情况时又该如何处理。

### 证据合成的外拓

文章中提起到社会选择理论，其主要是研究分析个人偏好和集体选择之间的关系，同时其中某些定理所描述的过程（如阿罗不可能性定理）与证据合成较为合适，两者是否具有通性及借鉴可能。

# 广义证据理论及应用

## 概念建立

### 广义基本概率指派定义

在原基本概率指派的基础上，去掉了这一条件，将未知领域（识别框架不完整下未知命题）全部规划到上去.

定义：设为开放世界的辨识框架，的幂集构成命题集合，，如果函数满足：



定义：给定一个GBPA，命题A的GBF可以表示为：







定义：给定一个GBPA，命题A的GPF可以表示为







### 存在的疑问

#### 另一种广义基本概率指派的定义

假设存在一个不完整的辨识框架，其对应必存在一些命题是不包括在这个辨识框架里面的，我们将这些命题（包括其扩展幂集）用表示，即可得到对应的广义基本概率指派函数，





#### 对GBF定义中和GPF定义中疑问

按照经典证据理论定义Bel和Pl如下：





现将其扩展到广义识别框架上，假设不包括在这个辨识框架里面的这些命题（包括其扩展幂集）用表示，可知为扩展幂集后的集合，其中是包含辨识框架内的信息的。设内任意一个子集用或者表示，可以如下定义等式：





明显可以得知，



文章中将当成一个整体，认为是不完整识别框架外的其他部分。但是更细致的分析可以看出在证据理论定义里面，由于是将识别框架里面的命题扩展到幂集上，从而导致里面的子集是与识别框架里面的存在交叠的。

#### 广义组合规则对的处理

若两个不同信息源在广义识别框架下，存在不同的识别框架外的命题，即是不同的，该如何进行组合处理，或者如何进行预处理。

## 广义基本概率指派生成方法

### 强约束GBAP生成方法

基于模糊数学中多子集命题表示方法，利用识别框架中各命题相同属性构建三角模糊数。对于特定部分的识别框架，可以利用最大值、最小值、均值进行三角模糊数的构建。

其中，各命题所构成的三角模糊数在同一个坐标系里面几何交叠处即为各命题的交集表示。

### 强约束GBAP生成方法中的若干疑问

#### 识别框架中的命题存在多个三角模糊数的情况

文章所提到的Iris data数据，每个样本是有四个属性的。文中仅讨论了一个属性（即花萼长度），若加入其他三个属性该如何分析。由于仅讨论一个属性即可获得一种GBAP的分配方案，加入其他属性必会生成不同的GBAP，该如何处理两着或者多者的关系，如果两种属性产生的GBAP存在较大冲突时该如何取舍。

#### 强约束GBAP生成方法中冲突的归一化问题

文中提起在GBPA生成过程中对于特定的样本可能生成的GBPA将大于1，而如果累积之和超过1，则将它们归一化且设的值为0。如果累积之和未超过1，则生成。

但若存在两个样本，生成的GBPA累计值都大于1。但是这两者的累计是不同的，如果简单的直接自行归一化，是否对不同样本的GBPA分配不公平。因为存在一下情况，有的接近于1，没有进行归一化；有的远超过1，进行了归一化；这将导致得到的GBPA是没有统一标准的。

#### 三角形模糊数生成问题

文中使用50个样本中的30各样本进行三角形模糊数的生成，那么如何进行选择。即如何保证选择的这30个样本能体现出该辨识框架的性质，如何保证所选择的样本不是冲突样本。

### 弱约束GBPA生成方法

没懂。

## 新的冲突表示模型

### 在广义识别框架下组合规则使用策略

当时说明识别框架是完整的，以此仅考虑证据距离dis。如果证据距离小时则说明冲突小，可以使用Dempster组合规则；反之说明冲突大，应先修正数据模型再使用Dempster组合规则。

当时说明识别框架是不完整的，则不考虑证据距离dis，一概不使用Dempster组合规则。

### 广义识别框架下组合规则定义的若干疑问

#### 广义证据组合中的定义

本文中在考虑广义证据组合冲突的时候沿用了传统证据冲突定义方法，是否需要考虑识别框架外的命题对整体冲突的影响？（明显当存在额外的焦元，得到的冲突系数必然不相等）

#### 广义证据组合下冲突表示

广义识别框架中表示的是目标命题是属于已知识别框架内可能性，当为不为0时文中认为不能使用Dempster组合规则进行处理，但并没有给出具体的处理办法。

### 总结

本文所定义的新的二元组合来表示证据之间的冲突，与Liu在不考虑识别框架不完整的情况下所定义的所不同的是：本文使用的是广义组合规则中产生的“冲突”。但是可以看到两者组合冲突的定义是一样的，即。因此，是否会存在这样一种情况：但就组合规则中产生的冲突而言，存在两个在已知识别框架中的信任度分配相似的样本（即对于相同命题在该样本中信任度占比相近）。但是一个确定为是完整的识别框架，而另一个是不完整的。由此必定有部分信任度会分配到中去，从而导致。这说明不完整识别框架计算的“冲突”是小于完整识别框架的，但是对于现实工程而言，明显完整识别框架所带来的信息更有效，冲突应更小。

## 不完整辨识框架的识别方法

### 辨识框架存在的不同情况

辨识框架主要存在以下四种情况：在无干扰的情况下辨识框架完整、在无干扰的情况下辨识框架不完整、在有微小干扰的情况下辨识框架完整、在有较大干扰的情况下辨识框架完整。

### 辨识框架的识别策略

本文主要是通过结合三个特征参数来判断辨识框架的完整与否和受到的干扰程度，这三个特征参数分别为：在GBPA分配的过程中产生的、一个样本内证据间、所有样本间的。对于不同的辨识框架，这三个特征参数是不一样的，故根据这三个特征参数可以很好的区分出辨识框架的状态。

### 存在的疑问

#### 另一种辨识框架的考虑

倘若存在一种辨识框架，即该辨识框架是不完整的而且存在较大的冲突（传感器获得的数据有错误）。根据本文的辨识框架的识别策略，其大致的数据特征应可能与辨识框架不完整无干扰的相似，故如何区分二者。同时还需考虑另一个问题，即在识别框架不完整的情况下定量描述此识别框架存在的冲突（干扰）程度。

#### 辨识框架与证据属性的考虑

以文中的iris为例，首先考虑GBPA的分配过程。此完整的识别框架为三种植物，即{Setosa,Versicolour,Virginica}，而此三种植物存在四种属性。那么一个完整的识别样本应为

，我们需要根据四个属性判断出这个样本的Plant是什么。 现我们构造了针对这四个属性的三角形模糊数，对于给的的样本可以生成四个GBPA分配方案，即在GBPA分配过程中即可产生与证据属性相同的。因此认为在GBPA分配的过程中产生的和一个样本内证据间理论上是相同的。

## 封闭世界中的冲突处理

### 最优折扣系数融合算法

* 计算加权证据距离

在原证据距离的基础上对各BPAs进行加权，计算出加权证据距离，表达式如下：



* 根据全局证据距离规划出加权系数

定义全局证据距离，当全局证据达到最小时说明对BPAs的预处理达到最佳状态。



目标函数为





* 根据加权系数得到折扣系数

将各加权系数除以加权系数中最大值便可得各加权系数对应的折扣系数。

