Application of the Belief Functions Theory in Problems of Information Fusion

Alexander Lepskiy

The International Laboratory of Decision Choice and Analysis, HSE, Moscow

Conference ‘Decision and Data Analysis’,
HSE, Moscow, 27-28 December, 2018

Problems of Information Fusion

Let we have several sources of information.

Examples of sources of information:

- expert evaluations (analyst recommendations);
- signals from various sensors (multiple-sensors);
- various classifiers, etc.

Then the following **problems of information fusion** can be formulated:

- how to aggregate information from different sources to obtain more reliable and accurate information;
- how to account for the different reliability of information sources;
- how to evaluate the independence of information sources;
- how to assess the consistency (conflict) of information sources, etc.

There are many approaches to solving these complex problems. The application of the **belief functions theory** is one of them.

Outline of Presentation

- The background of the theory of belief functions and combining rules
- Applications in problems of information fusion:
  - ❑Aggregation of analyst recommendations (together with E.Kutynina)
  - ❑Aggregation of technical indicators of the foreign exchange market (together with A.Suevalov)
  - ❑Aggregation of classifiers (together with K.Kuznetsov)
  - ❑Aggregation of filters in the image filtering problem (together with A.Ahmetgareeva)

The Background of the Theory of Belief Functions

[Dempster 1967, Shafer 1976]

Let X be a finite set, 2^X be a powerset of X.

The mass function is a set function $m: 2^X \to [0,1]$ that satisfies the condition $\sum_{A \subseteq X} m(A) = 1$. The value $m(A)$ characterizes the relative part of evidence that the actual alternative from X belongs to set $A \in 2^X$.

Notations and terms

- $A \in 2^X$ is called a focal element, if $m(A) > 0$;
- $\mathcal{A} = \{A\}$ be a set of all focal elements of evidence;
- $F = (\mathcal{A}, m)$ is called a body of evidence;
- $\mathcal{F}(X)$ be a set of all body of evidence on X;
- $F_A = (A,1)$ is called a categorical body of evidence;
- $F_X = (X,1)$ is called a vacuous body of evidence because it is a totally uninformative.

Convex Presentation of Evidence

If $F_j = (\mathcal{A}_j, m_j) \in \mathcal{F}(X)$ and $\sum_j \alpha_j = 1$, $0 \le \alpha_j \le 1$, then $F = (\mathcal{A}, m) \in \mathcal{F}(X)$, where $\mathcal{A} = \bigcup_j \mathcal{A}_j$, $m(A) = \sum_j \alpha_j m_j(A)$. This is denoted as $F = \sum_j \alpha_j F_j$. In particular, we have

$$F = \sum_{A \in \mathcal{A}} m(A) F_A \quad \forall F = (\mathcal{A}, m).$$

For example, if the expert predicted that the value of the shares would be in the interval $A = [40, 50]$ with a probability 0.7 or in the interval $B = [50, 55]$ with a probability 0.3, then this can be written as

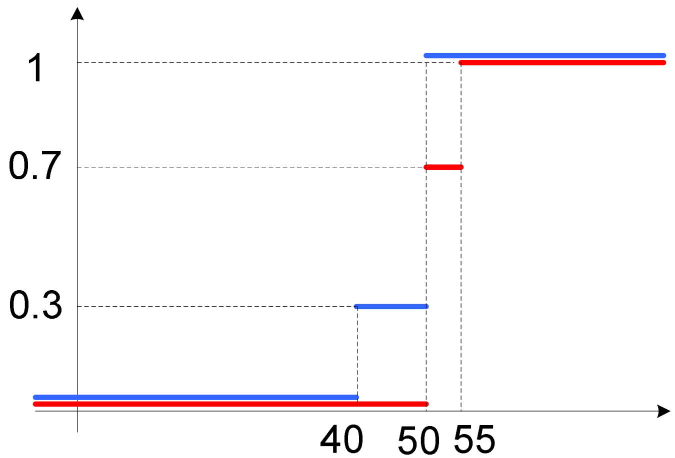
$$F = 0.7 F_{[40,50]} + 0.3 F_{[50,55]}.$$

Upper and Lower Distribution Functions

If X is a bounded set in $\mathbb{R}$, then we can calculate the upper and lower distribution functions

$$\underline{F}(x) = \begin{cases} \sum_{i:\sup A_i \le x} m(A_i), & x < \sup X, \\ 1, & x \ge \sup X \end{cases} \qquad \overline{F}(x) = \begin{cases} \sum_{i:\inf A_i \le x} m(A_i), & x > \inf X, \\ 0, & x \le \inf X \end{cases}$$

$(\underline{F}(x), \overline{F}(x))$ is a so-called ***p*-boxes**. For example, we have for previous evidence



Upper and Lower Expectations

Then we can calculate the lower and upper expectation of events.

Lower expectation

$$\underline{\mathbf{E}}[F] = \int_X s d\overline{F}(s) = \sum_i m(A_i) \inf(A_i).$$

Upper expectation

$$\overline{\mathbf{E}}[F] = \int_X s d\underline{F}(s) = \sum_i m(A_i) \sup(A_i).$$

For example, we have for previous evidence

$$\underline{\mathbf{E}}[F] = 0.7 \cdot 40 + 0.3 \cdot 50 = 43, \qquad \overline{\mathbf{E}}[F] = 0.7 \cdot 50 + 0.3 \cdot 55 = 51.5.$$

Belief and Plausibility Functions

Belief and **plausibility** functions

$$Bel(A) = \sum_{B \subseteq A} m(B), \qquad Pl(A) = \sum_{B \cap A \neq \varnothing} m(B),$$

where $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B)$ (so called **Möbius transform**).

Duality relation: $Pl(A) + Bel(\neg A) = 1$.

Bel(A), *Pl(A)* characterize the lower and upper probability estimates of the event $x \in A$:

$$Bel(A) \leq Pr(A) \leq Pl(A).$$

Example

Suppose that 10 experts give a forecast about the prospects for the development of three technologies $\{a,b,c\}$:

$$3 - \{a, b\}, \qquad 4 - \{b, c\}, \qquad 2 - \{b\}, \qquad 1 - \{c\}.$$

It is necessary to assess the probabilities of the development prospects of each technology. We have

$$m(\{a,b\}) = 0.3, \qquad m(\{b,c\}) = 0.4, \qquad m(\{b\}) = 0.2, \quad m(\{c\}) = 0.1;$$

$Bel(\{a\}) = 0,$ $\qquad$ $Pl(\{a\}) = 1 - Bel(\{b,c\}) = 1 - 0.7 = 0.3,$

$Bel(\{b\}) = 0.2,$ $\qquad$ $Pl(\{b\}) = 1 - Bel(\{a,c\}) = 1 - 0.1 = 0.9,$

$Bel(\{c\}) = 0.1,$ $\qquad$ $Pl(\{c\}) = 1 - Bel(\{a,b\}) = 1 - 0.5 = 0.5,$

$Bel(\{a,b\}) = 0.3 + 0.2 = 0.5,$ $\qquad$ $Pl(\{a,b\}) = 1 - Bel(\{c\}) = 1 - 0.1 = 0.9,$

$Bel(\{a,c\}) = 0.1,$ $\qquad$ $Pl(\{a,c\}) = 1 - Bel(\{b\}) = 1 - 0.2 = 0.8,$

$Bel(\{b,c\}) = 0.4 + 0.2 + 0.1 = 0.7,$ $\qquad$ $Pl(\{b,c\}) = 1 - Bel(\{a\}) = 1 - 0 = 1.$

Therefore $0 \le P(\{a\}) \le 0.3, \qquad 0.2 \le P(\{b\}) \le 0.9, \qquad 0.1 \le P(\{c\}) \le 0.5.$

Combining Rules

Let we have two bodies of evidence $F_1 = (\mathcal{A}_1, m_1)$ and $F_2 = (\mathcal{A}_2, m_2)$. The different combining rules $R : \mathcal{F}(X) \times \mathcal{F}(X) \to \mathcal{F}(X)$ are considered.

The **non-normalized conjunctive rule** [Smets 1990]

$$m^{D_0}(A) = \sum\nolimits_{B \cap C = A} m_1(B) m_2(C),\ A \in 2^X.$$

The value $K(F_1, F_2) = m^{D_0}(\varnothing)$ characterizes the amount of **conflict** between two sources of information.

Dempster's rule (1967)

$$m^D(A) = \frac{1}{1-K} \sum\nolimits_{B \cap C = A} m_1(B) m_2(C),\ A \neq \varnothing.$$

If $K = 1$ (absolute conflict), then Dempster's rule not applicable.

Yager rule (1987)

$$m^Y(A) = \sum_{B \cap C = A} m_1(B) m_2(C)\ \ \forall A \neq \varnothing, X,\ m^Y(X) = m_1(X) m_2(X) + K,\ m^Y(\varnothing) = 0.$$

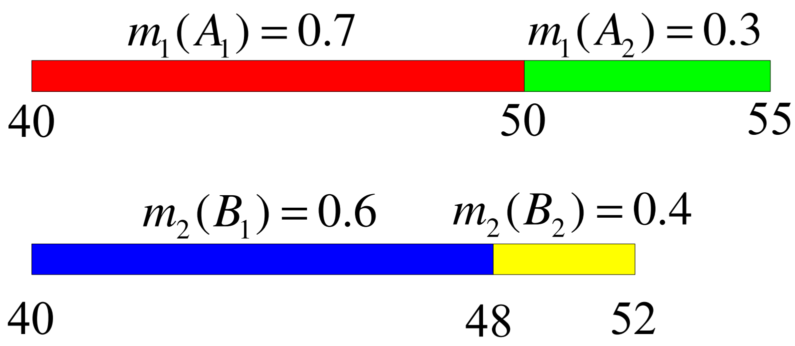
The disjunctive consensus rule [Dubois & Prade 1992]

$$m^{DP}(A) = \sum\nolimits_{B \cup C = A} m_1(B) m_2(C).$$

Example. The Dempster rule of combination

Let two experts gave the following information about the predictive value of shares

$$F_1 = 0.7F_{[40,50]} + 0.3F_{[50,55]}, \qquad F_2 = 0.6F_{[40,48]} + 0.4F_{[48,52]}.$$

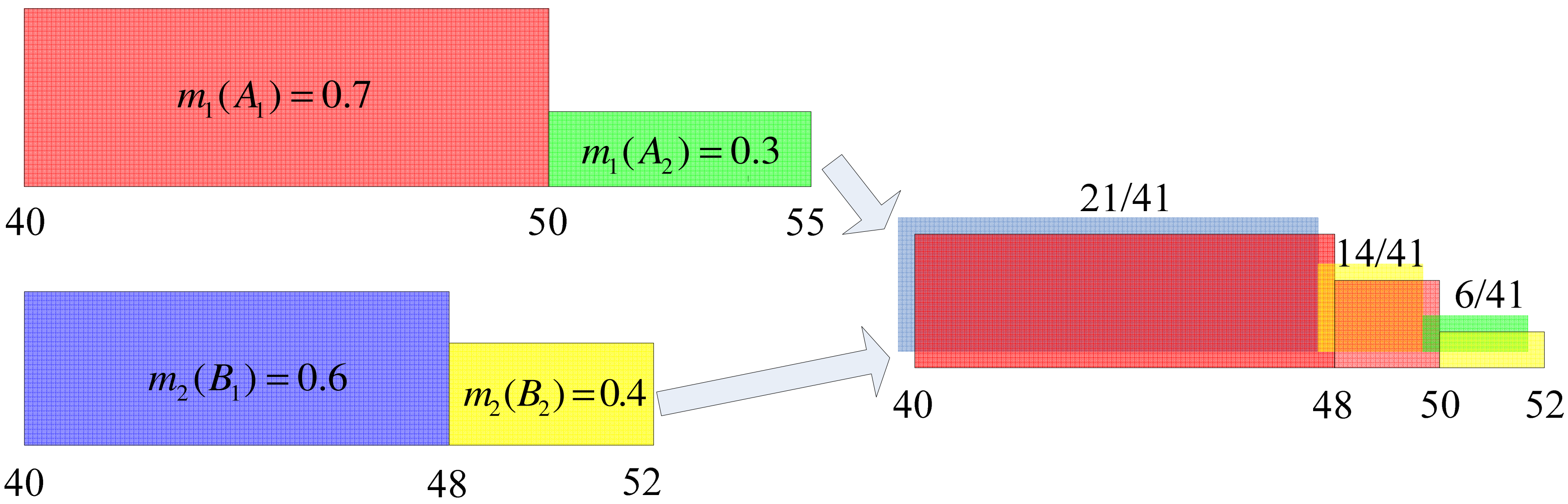


Then the conflict between these two bodies of evidence is equal $K = 0.3 \cdot 0.6 = 0.18$.

Example. The Dempster Rule of Combination

The new evidence obtained by the Dempster rule will be equal to

$$m^D([40,48)) = \tfrac{1}{1-K}\cdot 0.7\cdot 0.6 = 21/41,$$
$$m^D([48,50)) = \tfrac{1}{1-K}\cdot 0.7\cdot 0.4 = 14/41,$$
$$m^D([50,52)) = \tfrac{1}{1-K}\cdot 0.3\cdot 0.4 = 6/41.$$



Then

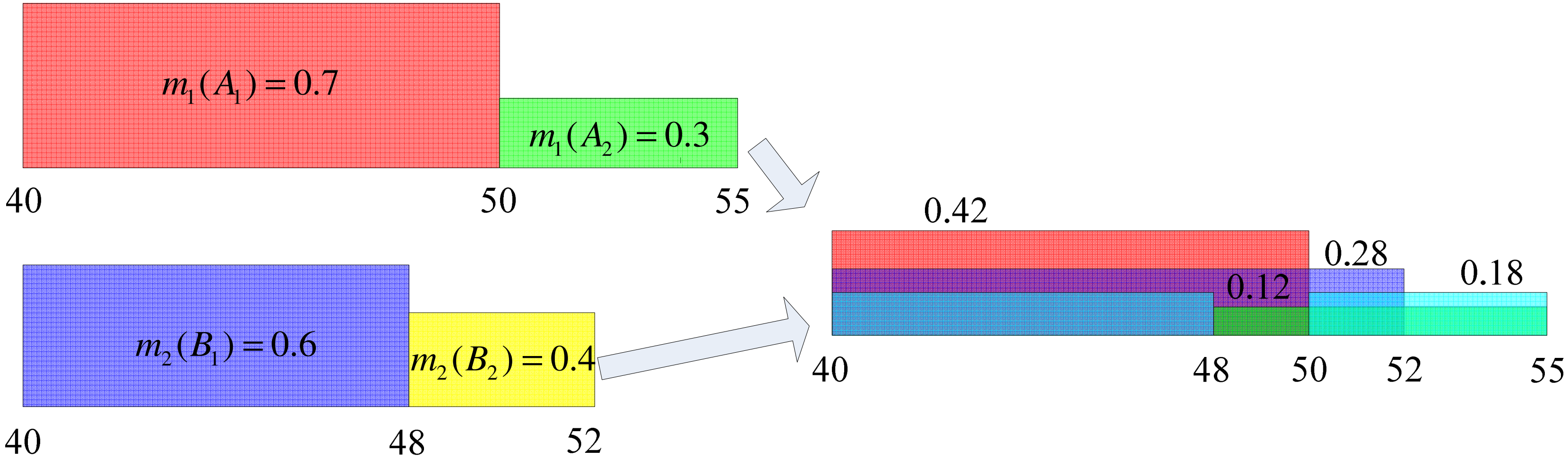
$$\underline{\mathbf{E}} = \tfrac{21}{41}\cdot 40 + \tfrac{14}{41}\cdot 48 + \tfrac{6}{41}\cdot 50 \approx 44,19, \quad \overline{\mathbf{E}} = \tfrac{21}{41}\cdot 48 + \tfrac{14}{41}\cdot 50 + \tfrac{6}{41}\cdot 52 \approx 49,27.$$

Example. The Disjunctive Consensus Rule

We have after applying the disjunctive consensus rule for the problem of predicting the value of shares

$m^{DP}([40,50)) = 0.7 \cdot 0.6 = 0.42,$ $m^{DP}([40,52)) = 0.7 \cdot 0.4 = 0.28,$

$m^{DP}([40,48) \cup [50,55)) = 0.3 \cdot 0.6 = 0.18,$ $m^{DP}([48,55)) = 0.3 \cdot 0.4 = 0.12.$



Boundaries of the expected value of shares are equal

$$\underline{\mathbf{E}} = 40.96, \qquad \overline{\mathbf{E}} = 52.6.$$

These estimates are more cautious (and more uncertain) than when combined according to the Dempster rule.

The Discounting Operation (Shafer 1976)

How to take into account the reliability of information sources?

Discounting operation. Let $\alpha \in [0,1]$ be the discount factor. Then

$$m^{\alpha}(A) = (1-\alpha)m(A),\ A \neq X\ ,\ m^{\alpha}(X) = \alpha + (1-\alpha)m(X)\,.$$

If $\alpha = 1$, then it means that information source is absolutely not reliable. If $\alpha = 0$, then it means that information source is absolutely reliable. The some combining rule applied after discounting of initial bodies of evidence.

The mass functions of all proper subsets evenly decrease with decreasing reliability (increase of α), and the mass function of the improper subset increases, which corresponds to an increase of information uncertainty.

Application Technique of the Belief Functions Theory

The main technical issue of application:

how to form the bodies of evidence (BE) $F = (\mathcal{A}, m)$ for different sources of information?

- what will be the universal set X?
- what will be a set of all focal elements of evidence $\mathcal{A} = \{A\}$?
- how to calculate mass function m?

Application 1. Aggregation of analyst recommendations (together with E.Kutynina)

Among the tasks associated with evaluating the recommendations of financial analysts, the important challenge is to aggregate recommendations and forecasts. We have:

- the target price is the share price expected by the expert at the end of the forecast period;
- recommendations of analysts can take the values "sell", "hold", "buy";
- 7 Russian banks and 3 analytical companies that provide their annual forecasts for 16 Russian companies represented on the Russian stock market during January 2010 – May 2016;
- the data on the real value of the shares of these companies in the period from January 2010 to May 2016.

How we can aggregate financial analysts' recommendations in the best way?

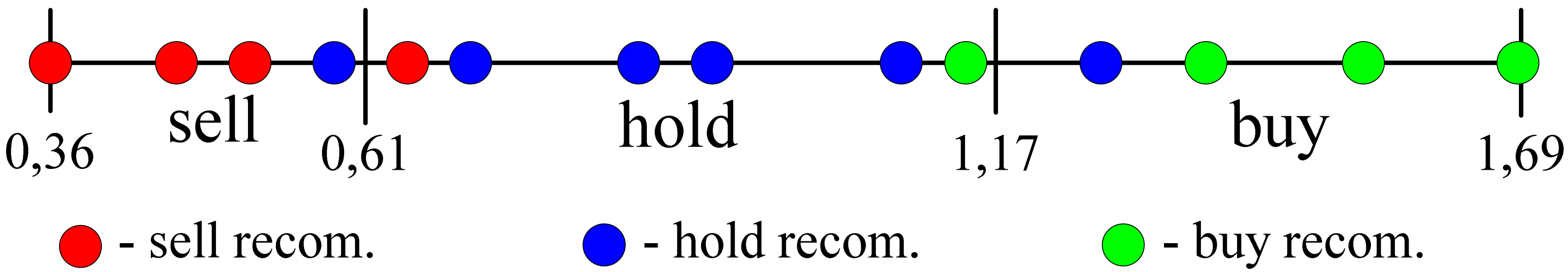
Determination of Focal Elements

1. We used the relative target price

$$Crv(stock;\ t) = \frac{\textbf{target price} \text{ of the share stock}}{\textbf{actual price} \text{ of stock on the date of the forecast } t}\ .$$

2. Boundary values of focal elements are calculated as a solution to the problem of minimization an error in the incorrect classiffcation of recommendations

Sberbank recommendations, error = 8,6%



Determination of Bodies of Analysts' Evidence

- During one year one analytical company gives several recommendations. For each analytical company *i* and each *stock* BE can be constructed $F_{i,stock} = (\mathcal{A}_{i,stock}, m_{i,stock})$.
- Each BE has not more than three focal elements $S_{i,stock}$, $H_{i,stock}$, $B_{i,stock}$, and mass functions $m_{i,stock}(A)$ equal to relative frequency of recommendation.
- The set X is added to the set focal elements in the case of discounting.

The Problem of Finding the Optimal BE

Let we have n categorical BE (recommendation of i-th source in during a year) F_{A_s} regarding the shares of a certain company that ordered by the time, where $A_s \in \{S, H, B\}$, $s = 1,...,n$, S-sell, H-hold, B-buy.

We will consider a BE

$$F(\alpha_1,...,\alpha_n) = \oplus_{s=1}^{n} F_{A_s}^{\alpha_s},\ F_{A_s}^{\alpha_s} = (1-\alpha_s)F_A + \alpha_s F_X,\ 1 \ge \alpha_1 \ge ... \ge \alpha_n \ge 0$$

for finding of recommendation of i-th source on the end a year with account of revision of forecasts.

Criteria for optimization

$$C(\alpha_1,...,\alpha_n) = \left(\mathbf{E_0}[F(\alpha_1,...,\alpha_n)] - p\right)^2 \to \min,\ 1 \ge \alpha_1 \ge ... \ge \alpha_n \ge 0,$$

where p is an actual last "pre-forecast" relative price of the share, $\mathbf{E_0}[F] = \frac{1}{2}\left(\underline{\mathbf{E}}[F] + \overline{\mathbf{E}}[F]\right)$ is the middle value of the interval of expectation of the forecast price.

Example

Let $n=4$, $F_1=F_2=F_4=F_S$ ("sell") and $F_3=F_H$ ("hold"). Then

$$F(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=F_S^{\alpha_1}\oplus F_S^{\alpha_2}\oplus F_H^{\alpha_3}\oplus F_S^{\alpha_4}=m(S)F_S+m(H)F_H+m(X)F_X.$$

The conflict of discounting BE is equal $K=K\left(F_S^{\alpha_1},F_S^{\alpha_2},F_H^{\alpha_3},F_S^{\alpha_4}\right)=(1-\alpha_3)(1-\alpha_1\alpha_2\alpha_4)$.

The values of a mass function are equal:

$$m(S)=\tfrac{1}{1-K}\alpha_3(1-\alpha_1\alpha_2\alpha_4),\ m(H)=\tfrac{1}{1-K}\alpha_1\alpha_2(1-\alpha_3)\alpha_4,\ m(X)=\tfrac{1}{1-K}\alpha_1\alpha_2\alpha_3\alpha_4.$$

Consequently, we have

$$C(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=\left(\mathbf{E_0}[F(\alpha_1,\alpha_2,\alpha_3,\alpha_4)]-p\right)^2=$$

$$\left(\frac{\alpha_3(1-\alpha_1\alpha_2\alpha_4)S_0+\alpha_1\alpha_2(1-\alpha_3)\alpha_4H_0+\alpha_1\alpha_2\alpha_3\alpha_4\Omega_0}{\alpha_3+\alpha_1\alpha_2\alpha_4-\alpha_1\alpha_2\alpha_3\alpha_4}-p\right)^2,$$

where S_0, H_0, X_0 are middles of intervals of relative prices. For example, if $S_0=0.7$, $H_0=1.1$, $X_0=0.9$ and $p=0.8$, then we obtain optimal coefficients $\alpha_1=\alpha_2=1$, $\alpha_3\approx 0.34$, $\alpha_4\approx 0.13$ and $F\approx 0.7F_S+0.2F_H+0.1F_X$.

Different Combining Strategies

It is necessary to define the rules according to which the sources for combining will be selected. Two alternative rules for selecting sources were considered:

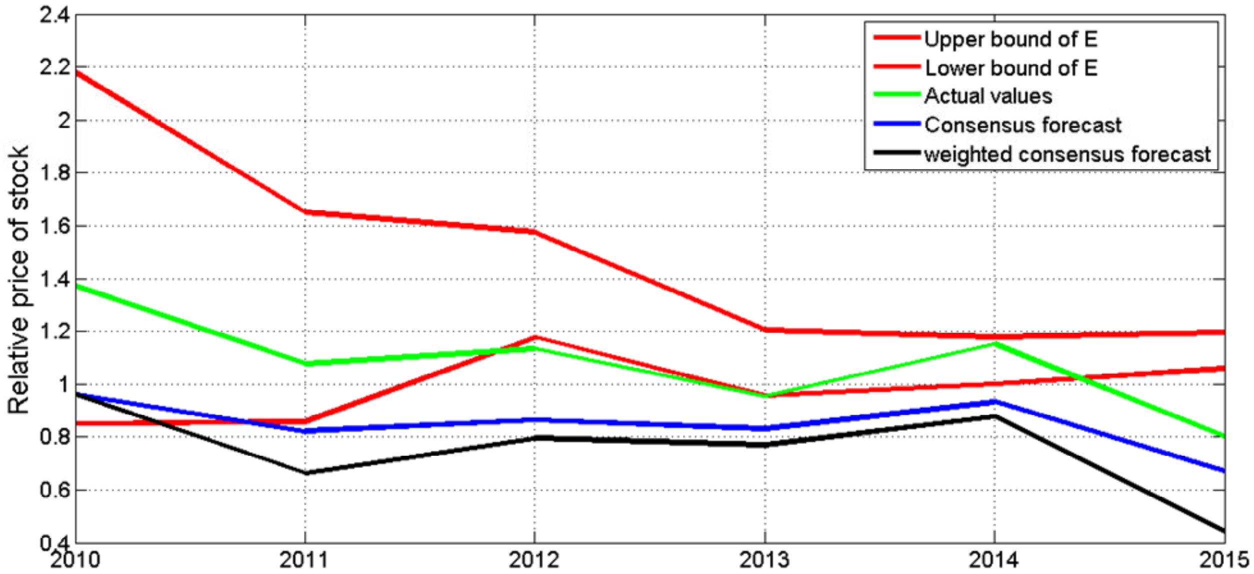
- all sources were ranked by the increase in the degree of conflict and the combination of evidence began with a pair of the least conflicting sources;
- for each data source i the degree of forecast's reliance was evaluated for each share stock basing on data of a previous period

$$\delta_{i,stock} = \frac{1}{N}\sum_{t}\frac{\left|Crv_{real}\left(stock;\ t\right) - Crv_{forecast}\left(stock;\ t\right)\right|}{\max\left\{Crv_{real}\left(stock;\ t\right), Crv_{forecast}\left(stock;\ t\right)\right\}}.$$

where N is the number of forecasts during the period, $Crv_{real}\left(stock;\ t\right)$ is the actual relative price, $Crv_{forecast}\left(stock;\ t\right)$ is the forecasted relative price. All sources were ranked by the increase $\delta_{i,stock}$ and the combination of evidence began with a pair of the most reliable and non-conflict sources.

Results

Example of forecasting for the share price of the Transneft company (TRNFP).



MAE_{stock}	**WCF**	**CF**	$\underline{E}$, **OSWC**	$\overline{E}$, **OSWC**	E_0, **OSWC**	$\underline{E}$, **NSWC**	$\overline{E}$, **NSWC**	E_0, **NSWC**	$\underline{E}$, **CLCS**	$\overline{E}$, **CLCS**	E_0, **CLCS**
GAZP	0,941	0,456	0,686	0,715	0,653	0,745	0,772	0,711	**0.275**	**0.359**	**0.308**
LKOH	0,499	0,27	**0,19**	0,99	0,59	**0,234**	0,918	0,524	**0.193**	0.309	**0.263**
ROSN	0,587	0,412	**0,153**	0,486	**0,302**	**0,194**	0,529	**0,265**	**0.110**	**0.257**	**0.191**
SBER	0,55	0,413	**0,137**	0,539	**0,33**	**0,178**	0,633	**0,349**	**0.116**	**0.319**	**0.233**
MAGN	0,298	**0,205**	0,385	0,937	0,661	0,347	1,014	0,681	0.345	0.470	0.422
SNGSP	0,485	0,332	**0,327**	0,84	0,546	**0,243**	1,037	0,598	**0.263**	0.405	**0.330**
GMKN	0,512	0,46	**0,382**	0,673	**0,447**	**0,349**	0,672	**0,345**	**0.283**	**0.314**	**0.267**
VTBR	0,637	0,249	0,451	0,376	0,344	0,525	0,273	0,286	0.269	**0.213**	**0.207**
TRNFP	0,33	0,234	**0,146**	0,356	**0,18**	**0,199**	0,416	**0,18**	**0.119**	**0.215**	**0.141**
TATN	0,568	0,3	**0,198**	0,998	0,589	**0,183**	1,027	0,596	**0.200**	0.421	0.331
MTSS	0,516	0,371	**0,268**	**0,352**	**0,249**	0,39	0,397	**0,259**	**0.214**	**0.205**	**0.210**
CHMF	0,311	0,203	**0,17**	0,369	0,227	**0,196**	0,405	0,222	**0.192**	0.240	0.225
ALRS	0,216	0,14	**0,119**	0,204	**0,116**	0,156	0,308	0,16	**0.115**	0.162	**0.110**
NVTK	0,236	0,395	0,28	0,396	**0,196**	0,28	0,495	**0,149**	0.273	**0.198**	**0.178**
AFLT	0,123	**0,033**	0,477	0,186	0,222	0,552	0,365	0,193	0.376	0.076	0.218
URKA	0,6	0,523	**0,504**	**0,216**	**0,357**	0,654	**0,285**	**0,339**	**0.338**	**0.244**	**0.271**
MAE	0,463	0,312	**0,305**	0,539	0,376	0,339	0,597	0,366	**0.230**	**0.275**	**0.244**

MAE – mean absolute error, WCF – Weighted consensus forecast; CF – consensus forecast; the Dempster's rule with discounting and (a) optimistic scenario without censorship (OSWC); (b) neutral scenario without censorship (NSWC); (c) with the choice of the least conflicting sources (CLCS).

Application 2. Aggregation of technical indicators of the foreign exchange market (together with A.Suevalov)

1) based on technical indicators evidences are built with focal elements {Sell}, {Buy}, {Sell, Hold}, {Buy, Hold};
2) each indicator defines its classifier with its reliability of prediction;
3) evidences of individual indicators (classifiers) are aggregated with the help of a particular combination rule from the theory of belief functions (for example, the Dempster rule); the aggregation parameters are the discount coefficients that take into account the effectiveness of individual indicators in the decision-making process about choosing a trading solution;
4) tuning of aggregation parameters is carried out at the stage of classifier learning.

Application 2. Determination of Bodies of Evidence

- $X = \{\text{Sell,Hold,Buy}\}$ be an universal set;
- $\mathcal{A} = \{\{\text{Sell}\},\{\text{Sell, Hold}\},\{\text{Hold, Buy}\},\{\text{Buy}\}\}$ be a set of focal elements of evidence;
- mass functions are determined by the rules of Mamdani

$$\text{IF } I_i \text{ IS } T_j \text{ THEN } m_{i,k}(A_s) = \mu_i^{T_j}(t_k^{(s)}),$$

where $\mu_i^{T_j}(t_k)$ be a value of membership function for i-th indicator with a linguistic variable $T_j \in \{\text{'Very Low'},\text{'Low'},\text{'High'},\text{ 'Very High'}\}$,
$t_k^{(s)}$ be a k-th value of the training set for $A_s \in \mathcal{A}$;

Application 2. Learning and Results

- teach the classifier, i.e. find the optimal discount coefficients $\{\alpha_i\}$ under which the discrepancy between the recommendation of the system and the real best action on the training data sample is minimized

$$\sum_k d\left(\mathcal{D}\left(\oplus_{R\,i=1}^{\,n} m_{i,k}^{\alpha_i}\right), \varphi(t_k)\right)^2 \to \min,$$

where $\oplus_R$ be a combining rule (for ex., Dempster's rule); $\mathcal{D}$ be a rule of decision making, $\mathcal{D}$: $F = (m_R, \mathcal{A}) \mapsto \{\text{Sell,Hold,Buy}\}$, $\varphi(t_k)$ be a best solution from $\{\text{Sell,Hold,Buy}\}$ for learning value t_k, d be a metric.

Results

	Without discounting		With discounting		
	Averaging	Yager rule	Averaging	Yager rule	Dempster rule
USD/RUB	86.23	22.00	148.42	84.17	78.97
EUR/USD	21.06	6.71	49.64	58.68	65.76

Application 3. Aggregation of classifiers (together with K.Kuznetsov)

Problem. Let the K be different classifiers for classes $C = \{c_j\}_{j=1}^{M}$, trained on the test sample $X = \{x\}_{i=1}^{N}$. It is required to aggregate the results of their work. The traditional aggregation scheme is as follows

$$\hat{c}(x) = \arg max_{c_j \in C} \sum_k w_k \delta\left(\widehat{c_k}(x_i), c_j\right),$$

where $\widehat{c_k}(x_i)$ be a prediction of the k-th classifier for an object x_i, $\hat{c}(x)$ be a ensemble prediction; $\delta(a, b)$ be the Kronecker symbol.

Traditional aggregators:

- Plurality vote: $w_k = \frac{1}{K}, \forall k$.
- Simple weighted vote: $w_k = \frac{a_k}{\sum_l a_l}$, where a_k be a proportion of correctly classified objects by the k-th classifier.

Application 3. Examples of traditional aggregators

- Re-scaled weighted vote. $w_k \propto a_k = \max\left\{0, 1 - \frac{Me_k}{N(M-1)}\right\}$,
  where e_k be a number of errors made by the k-th classifier;
- Best-worst weighted vote: $w_k \propto a_k = 1 - \frac{e_k - e_B}{e_W - e_B}$,

  where $e_B = min_k\{e_k\}, \quad e_W = max_k\{e_k\}$.
- Quadratic best-worst weighted vote: $w_k \propto a_k = \left(\frac{e_k - e_B}{e_W - e_B}\right)^2$.
- Weighted majority vote: $w_k \propto log\, \frac{\alpha_k}{1-\alpha_k}$,

  where α_k be a accuracy of a single classifier.

Application 3. Determination of Bodies of Evidence

Researched classifiers:

- method of *k* nearest neighbors (knn);
- logistic regression (lr);
- random forest (rfc);
- support vector machine (SVM);
- naive Bayes classifier (nb).

Determination of the body of evidence for a binary classifier:

- $X = \{C_1, C_2\}$be a universal set of two classes;
- $\mathcal{A} = \{\{C_1\}, \{C_2\}, X\}$ be a set of focal elements;
- mass function:

$$m(\{C_1\}) = ap_{C_1}(t),\ m(\{C_2\}) = ap_{C_2}(t),\ m(X) = 1 - m(\{C_1\}) - m(\{C_2\}),$$

where a be a classification accuracy, $p_{C_i}(t)$ be a probability of classifying the pattern t to the class C_i.

Application 3. Aggregation results without discounting

Database / Classification Algorithms	Dempster	NB	RFC	LR	SVM	knn	Yager
Pime	**0.837**	0.808	0.808	0.811	0.8105	0.789	0.836
Blood-trans	0.751	0.706	0.679	**0.759**	0.689	0.739	0.75
Banknote auth	**1.00**	0.935	0.991	0.992	0.995	0.995	**1.00**
Spambase	**0.994**	0.856	0.987	0.973	0.972	0.972	**0.994**
Skin segm	**0.999**	0.937	**0.999**	0.908	**0.999**	0.998	**0.999**
Phoneme	0.931	0.824	**0.959**	0.818	0.916	0.914	0.931
Ionosphere	0.935	0.851	0.941	0.839	**0.974**	0.782	0.935
Diabetes	**0.743**	0.728	0.725	0.709	0.726	0.684	0.741
Oil spil	**0.871**	0.604	0.632	0.808	0.860	0.725	**0.871**
Electricity	0.859	0.704	**0.899**	0.734	0.787	0.801	0.859

Application 4. Aggregation of filters in the image filtering problem (together with A.Ahmetgareeva)

Let $I(\mathbf{x})$ be a grayscale image,
$\tilde{I}(\mathbf{x}) = I(\mathbf{x}) + \eta(\mathbf{x})$ be a noisy image ($\eta(\mathbf{x})$ be a noise).
It is required to find such an operator (filter) φ that

$$\Phi\left(\varphi(\tilde{I}) - I\right) \to \text{extr},$$

where Φ be a filtering quality functional.

Let we have several different filters $\varphi_1,\dots,\varphi_n$. It is required to form a filter $\varphi = A\left(\varphi_1,\dots,\varphi_n\right)$, where A be some aggregation operator.

Examples of filter-evidence:

- $\varphi_1 = \left|x_5 - w\right|,\ w = \frac{1}{8}\sum_{i\neq 5} x_i$
- $\varphi_2 = \left|x_5 - \text{med}\right|,\ \text{med}=\text{median}\{x_1,\dots,x_9\}$
- $\varphi_3 = \left|x_5 - \text{mean}_k\right|,\ \text{mean}_k = \frac{1}{8+k}\left(kx_5 + \sum_{i\neq 5} x_i\right),$
  $k = 1,3,5,\dots,$ etc.

x_1	x_2	x_3
x_4	x_5	x_6
x_7	x_8	x_9

Application 4. Determination of Bodies of Evidence

- $X = \{N, F\}$ be an universal set: N be a set of **noisy** pixels; F be a set of **noise-free** pixels;
- $\mathcal{A} = \{N, F, N \cup F\}$ be a set of focal elements of evidence;
- mass function for i-th filter-evidence are determined by the rules [Lin 2008]

$$m_i(N) = g_i(\varphi_i), \quad m_i(F) = \beta(1 - m_i(N)), \quad m_i(F) + m_i(N) + m_i(N \cup F) = 1,$$

where g_i be some function, $\beta \in [0,1]$.

Application 4. Filter Aggregation

- Further we aggregate bodies of evidence filters $\varphi_1,...,\varphi_n$

$$m = \left(\oplus_R\right)_{i=1}^{n} m_i^{\alpha_i},$$

where $\oplus_R$ be a combining rule (for ex., Dempster's rule); $\alpha_i \in [0,1]$ be discount coefficients;

- and we compute $Bel(N)$ (degree of belief that the pixel is noisy);
- the filtering rule is applied for pixel x

$$y = m + \lambda(Bel(N))(x - m),$$

where m denotes some filter (for ex., the median filter),
$\lambda:[0,1] \to [0,1]$ be some monotone non-increasing function that satisfies conditions: $\lambda(1) = 0$, $\lambda(0) = 1$. If $\lambda = 0$ (pixel is corrupted), then $y = m$. If $\lambda = 1$ (pixel is not corrupted), then $y = x$.

Application 4. Results

noisy image (impulse noise, 20%)

median filter

adaptive median filter

aggregation filter

	Gauss filter	median filter	adaptive median filter	aggregation filter
MAE ↓	4.15	2.19	1.32	**1.25**
PSNR ↑	23.34	29.38	**30.88**	26.01
SSIM ↑	0.695	0.886	0.928	**0.934**

Thanks for your attention!