Агрегирование информации в рамках теории свидетельств

А.Е. Лепский (НИУ ВШЭ, Москва)

IV Российский экономический конгресс
круглый стол «Индивидуальный выбор и задачи агрегирования – современное состояние и приложения»

21 декабря 2020г

Агрегирование информации в экономическом анализе

Агрегирование информации, полученной из разных источников – одна из центральных проблем при решении задач принятия решений, в том числе, в сфере экономики и финансов, в бизнес-аналитике и т.д.

Агрегирование позволяет:

- получить более полную информацию об объекте интереса (бизнес-процессе, клиенте и т.д.);
- учесть разные, может быть «не очевидные», особенности объекта;
- компенсировать недостаток данных одних источников – другими и т.д.

Требование к агрегированию

Математический аппарат агрегирования должен учитывать:

- неопределенность информации;
- противоречивость информации, полученной из разных источников (а может и из одного источника);
- разную степень надежности информации и т.п.

Дополнительные трудности: число источников информации и объем этой информации зачастую является небольшими (например, при агрегировании экспертных оценок) для построения надежных статистических оценок или машинного обучения.

Теория свидетельств Демпстера-Шейфера (the Dempster-Shafer theory of evidence)

= теория функций доверия (theory of belief function),
= теория случайных множеств (theory of random set).

Dempster A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping// The Ann. of Math. Stat. 1967, 38(2), 325-339.

Shafer G. A Mathematical Theory of Evidence. – Princeton, N.J.: Princeton University Press, 1976.

Базовые понятия

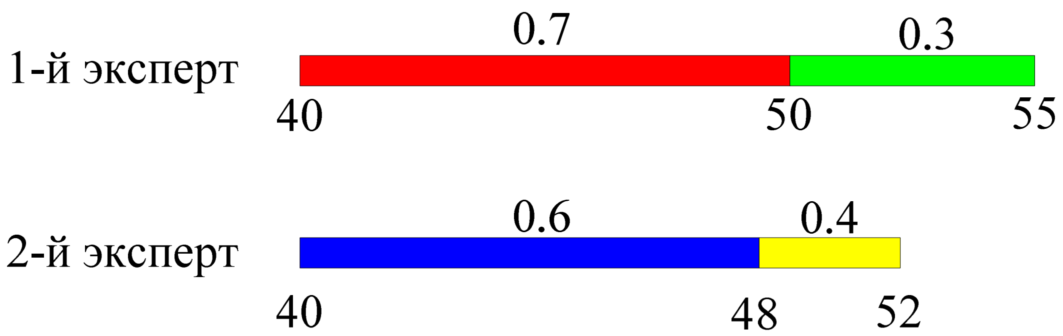
- X – универсальное множество всех возможных значений результатов эксперимента (наблюдений, альтернатив и т.д.);
- 2^X – множество всех подмножеств из X;
- **функция масс (базовая вероятность)** $m\colon 2^X \to [0,1]$, $m(\varnothing)=0$ и $\sum_{A\in 2^X} m(A)=1$ (конечное число ненулевых слагаемых).

Н/р, $m(A)=c/N$, где c – количество наблюдений $x\in A$, N – общее количество наблюдений.

- **фокальный элемент** $A\subseteq X$, если $m(A)>0$;
  $\mathcal{A}=\{A\}$ – множества всех фокальных элементов;
- **тело свидетельства** $F=(\mathcal{A},m)$, где $m(A)$, $A\in\mathcal{A}$;
  $\mathcal{F}(X)$ – множество всех тел свидетельств на X.
- **категоричное** (categorical) тело свидетельств $F_A=(A,1)$;
  н/р, стоимость акций будет в интервале $A=[40,50]$;
- **бессодержательное** (vacuous) тело свидетельств $F_X=(X,1)$;
  н/р, стоимость акций будет в интервале $X=[0,+\infty)$.

Пример

Пусть два эксперта независимо дают следующую информацию о прогностической стоимости акций некоторой компании через месяц



Тогда на универсальном множестве $X=[0,+\infty)$ имеем два свидетельства $F_1=(\mathcal{A}_1,m_1)$ и $F_2=(\mathcal{A}_2,m_2)$:

$$\mathcal{A}_1=\{A_1,A_2\},\ A_1=[40,50),\ A_2=[50,55],\ m_1(A_1)=0.7,\ m_1(A_2)=0.3;$$

$$\mathcal{A}_2=\{B_1,B_2\},\ B_1=[40,48),\ B_2=[48,52],\ m_2(B_1)=0.6,\ m_2(B_2)=0.4.$$

Или

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

Функции доверия и правдоподобия

Функция доверия (belief function): $Bel(A) = \sum_{B \subseteq A} m(B)$,

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B)$$

(преобразование Мёбиуса для конечного множества X)

Функция правдоподобия (plausibility function): $Pl(A) = \sum_{B \cap A \neq \varnothing} m(B)$.

Отношение двойственности: $Pl(A) + Bel(\neg A) = 1$.

J. Halpern & R. Fagin (1992): функцию доверия можно определить, как функцию, удовлетворяющую ослабленным вариантам аксиом Колмогорова, характеризующих вероятность.

$\Rightarrow$ *Bel(A)*, *Pl(A)* – нижняя и верхняя оценки вероятности события *A*:

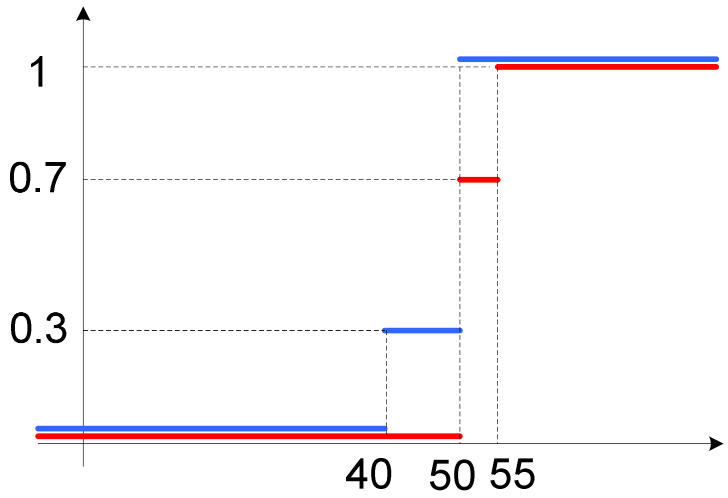
$$Bel(A) \leq Pr(A) \leq Pl(A).$$

Верхняя и нижняя функции распределения

Если экспертные оценки – ограниченные числовые множества, то можно вычислить верхнюю и нижнюю функции распределения

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

$(\underline{F}(x),\overline{F}(x))$ – p-блоки (p-boxes). Например, для задачи о прогнозировании стоимости акций



Верхнее и нижнее математические ожидания

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

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

Для задачи о прогнозировании цены акций

$$\underline{\mathbf{E}}[F_1] = 0.7 \cdot 40 + 0.3 \cdot 50 = 43, \qquad \overline{\mathbf{E}}[F_1] = 0.7 \cdot 50 + 0.3 \cdot 55 = 51.5$$

и

$$\underline{\mathbf{E}}[F_2] = 0.6 \cdot 40 + 0.4 \cdot 48 = 43.2, \qquad \overline{\mathbf{E}}[F_2] = 0.6 \cdot 48 + 0.4 \cdot 52 = 49.6.$$

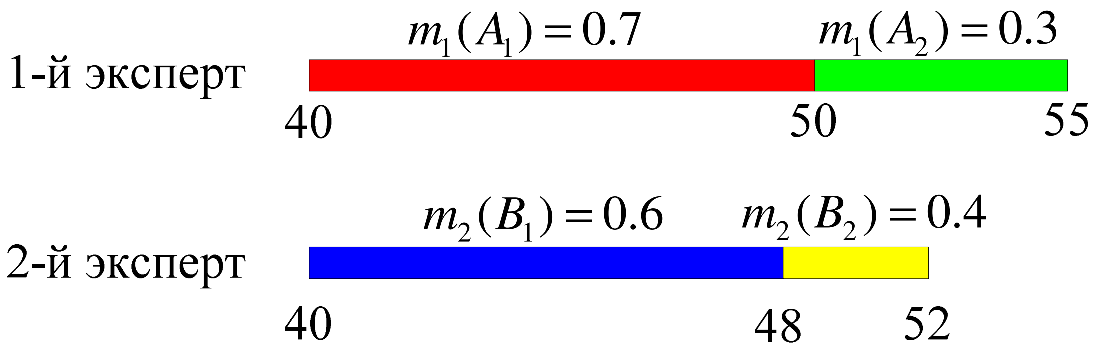
Конфликтность источников информации

Исторически первой (и наиболее популярной) оценкой конфликтности является величина, вычисляемая в правиле Демпстера:

$$K_0(F_1, F_2) = \sum_{A \cap B = \varnothing} m_1(A) m_2(B).$$

$K_0 \in [0,1]$ и $K_0 = 1$ соответствует абсолютной конфликтности свидетельств: $B \cap C = \varnothing$ для всех $B \in \mathcal{A}_1$, $C \in \mathcal{A}_2$,

Для примера двух экспертов



конфликт равен $K_0 = 0.3 \cdot 0.6 = 0.18$.

Агрегирование свидетельств

Если есть два независимых источника информации, заданных свидетельствами $F_1 = (\mathcal{A}_1, m_1)$ и $F_2 = (\mathcal{A}_2, m_2)$ на X, то возникает вопрос, как агрегировать эти данные?

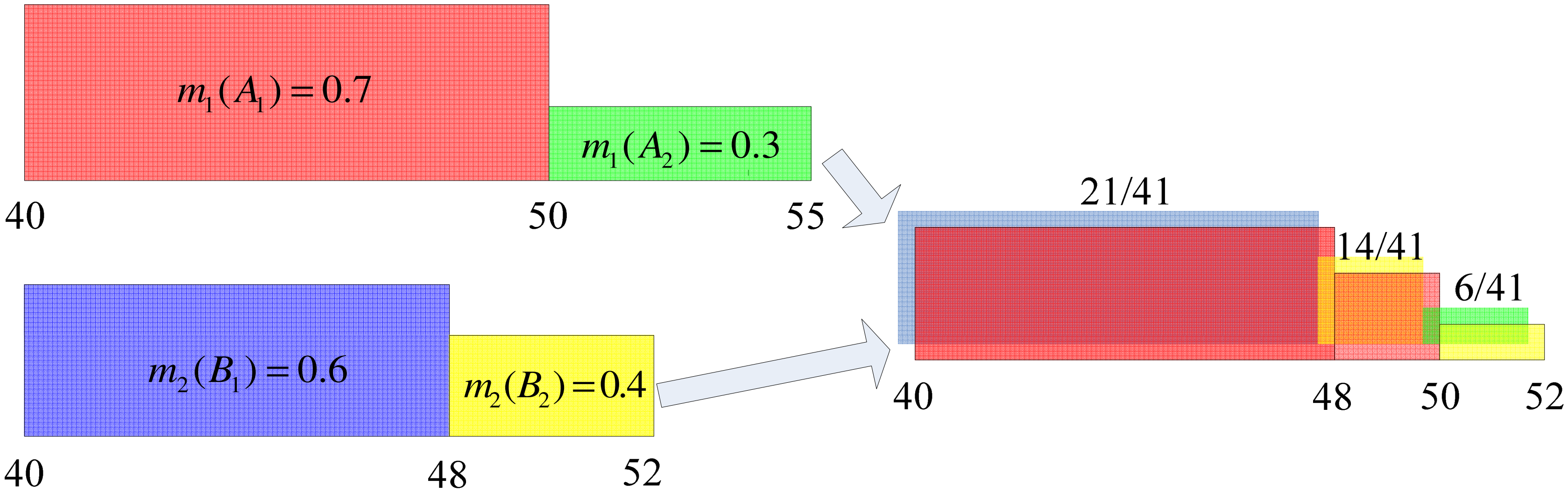
Правило Демпстера (1967). Базовая вероятность $m^D = m_1 \oplus_D m_2$ нового свидетельства вычисляется по формуле

$$m^D(A) = \frac{1}{1 - K_0} \sum_{B \cap C = A} m_1(B) m_2(C), \; A \neq \varnothing,$$

где $K_0 = K_0(F_1, F_2) = m^D(\varnothing) = \sum_{B \cap C = \varnothing} m_1\left(B\right) m_2\left(C\right)$. В случае $K_0 = 1$ (абсолютная конфликтность свидетельств) правило Демпстера неприменимо.

Правило Демпстера

Для задачи о прогнозировании стоимости акций



конфликт равен $K_0 = 0.3\cdot 0.6 = 0.18$, а новое свидетельство, полученное с помощью правила Демпстера, будет равно

$m^D([40,48)) = \frac{1}{1-K_0}\cdot 0.7\cdot 0.6 = 21/41,\qquad m^D([48,50)) = \frac{1}{1-K_0}\cdot 0.7\cdot 0.4 = 14/41,$

$m^D([50,52)) = \frac{1}{1-K_0}\cdot 0.3\cdot 0.4 = 6/41.$

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

Дизъюнктивное правило консенсуса

Правило Демпстера – **оптимистично**: если в соответствии с одним источником информации истинная альтернатива $x \in A$, а в соответствии с другим – $x \in B$, то после применения правила Демпстера – $x \in A \cap B$.

Другим «крайним» случаем агрегирования является **дизъюнктивное правило консенсуса** [Dubois & Prade 1992]

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

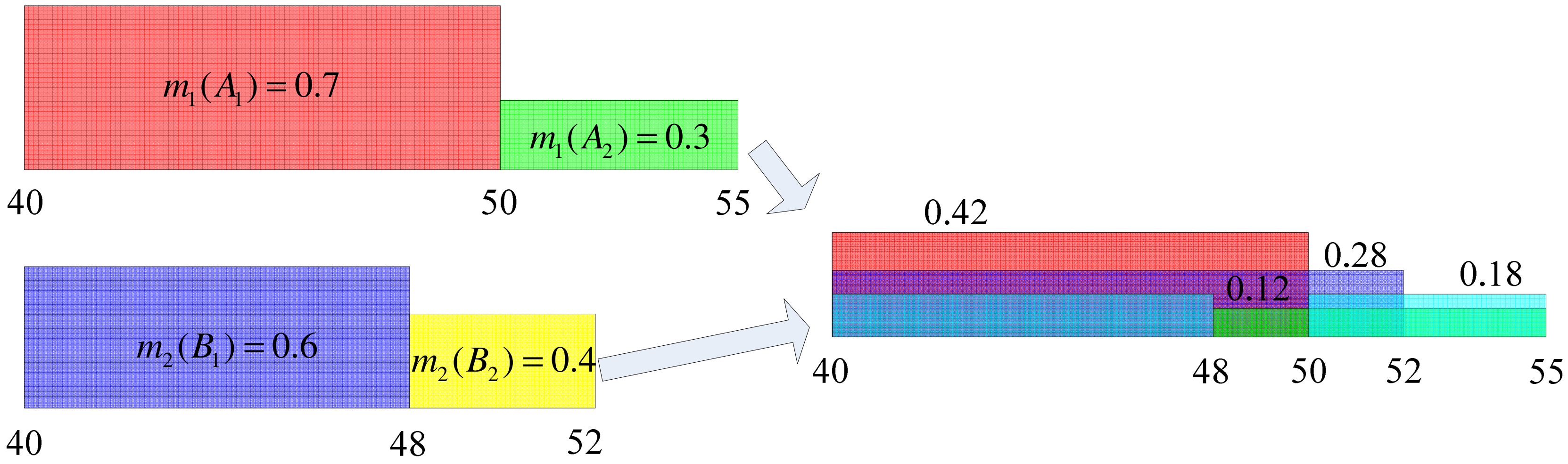
Это правило пессимистично: если один источник информации говорит, что истинная альтернатива $x \in A$, а другой – $x \in B$, то после комбинирования по **дизъюнктивному правилу** – $x \in A \cup B$.

Дизъюнктивное правило консенсуса. Пример

Для задачи о прогнозировании стоимости акций:

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

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



Оценки математического ожидания стоимости акции равны:

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

Оценки являются более осторожными (и более неопределенными), чем в случае комбинирования по правилу Демпстера.

Некоторые другие правила

Правило Ягера [Yager 1987]:

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

$$m^Y(X) = m_1(X) m_2(X) + K_0, \ m^Y(\varnothing) = 0.$$

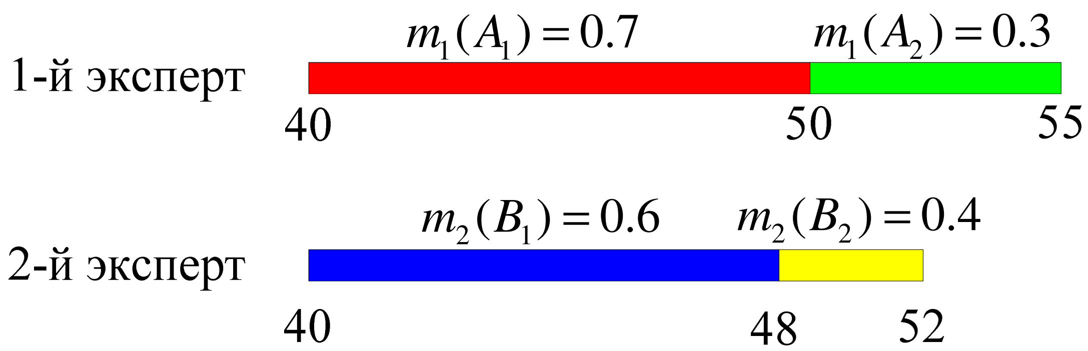
Конъюнктивные правила Демпстера и Ягера – **оптимистичны**: если в соответствии с одним источником информации истинная альтернатива $x \in A$, а в соответствии с другим – $x \in B$, то после применения правил Демпстера или Ягера – $x \in A \cap B$.

Неопределенность свидетельств

Степень информативной неопределенности оценивают с помощью, так называемых индексов неточности. Примером такого индекса является так называемая **обобщенная мера Хартли**:

$$H(\mathbf{m}) = \sum_{A \neq \varnothing} m(A) \ln\left(|A|\right).$$

Для задачи о прогнозировании стоимости акций двумя экспертами (считая, что $|A| = l(A) + 1$, $l(A)$ – длина интервала)



имеем $H(\mathbf{m}_1) \approx 2.216$ и $H(\mathbf{m}_2) \approx 1.962$, т.е. неопределенность первого свидетельства выше неопределенности второго.

Дисконтирование свидетельств

Предположим, что источники информации имеют разную надежность.

Вопрос: как учесть надежность источников информации?

Правило дисконтирования G. Shafer (1976). Функции масс дисконтируются с коэффициентом $\alpha \in [0,1]$ по правилам:

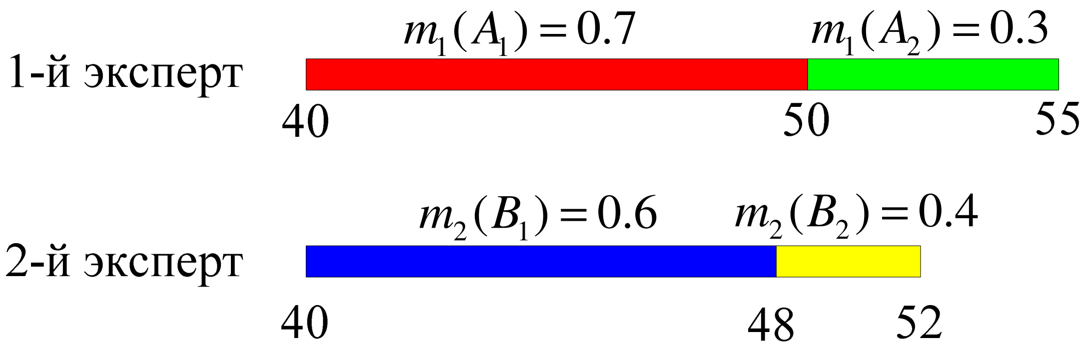
$$m^{\alpha}(A) = (1-\alpha)m(A),\ A \neq \Omega,\ m^{\alpha}(\Omega) = \alpha + (1-\alpha)m(\Omega).$$

Если $\alpha = 0$, то источник информации считается абсолютно надежным. Если же $\alpha = 1$, то источник информации абсолютно ненадежен.

Смысл правила: при уменьшении надежности (увеличении α) функции масс всех собственных подмножеств равномерно уменьшаются, а для несобственного подмножества – увеличивается, что соответствует увеличению информационной неопределенности.

Пример дисконтирования

Предположим, что в задаче о прогностической стоимости акций



надежность прогноза первого эксперта равна 100% ($\alpha = 0$), а второго – 70% ($\alpha = 0.3$). Тогда после дисконтирования второго источника, получим:

$$m_2^{0.3}(B_1) = 0.7 \cdot 0.6 = 0.42,\ m_2^{0.3}(B_2) = 0.7 \cdot 0.4 = 0.42,\ m_2^{0.3}(X) = 0.3.$$

После этого конфликтность источников информации уменьшится с $K_0 = 0.3 \cdot 0.6 = 0.18$ до $K_0^{0.3} = 0.3 \cdot 0.42 = 0.126$.

Принятие решений в торговых системах

- $X = \{\text{Sell}, \text{Hold}, \text{Buy}\}$ – универсальное множество;
- $\mathcal{A} = \{\{\text{Sell}\}, \{\text{Sell, Hold}\}, \{\text{Hold, Buy}\}, \{\text{Buy}\}\}$ – множество фокальных элементов;
- функции масс $m_{i,k}(A)$, $A \in \mathcal{A}$ для технического индикатора I_i и k-го значения обучающей выборки определяется по аналогии с функциями принадлежности лингвистической переменной;
- обучение системы: найти коэффициенты дисконтирования $\{\alpha_i\}$, которые минимизировали бы функционал

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

где $\oplus_R$ – правило комбинирования; $\mathcal{D}$ – правило принятия решений, $\mathcal{D}$: $F = (m_R, \mathcal{A}) \mapsto \{\text{Sell}, \text{Hold}, \text{Buy}\}$, $\varphi(t_k)$ – наилучшее решение из $\{\text{Sell}, \text{Hold}, \text{Buy}\}$ для обучающего значения t_k, d – метрика.

	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

Агрегирование классификаторов

Даны K классификаторов для классов $\mathcal{C} = \{C_j\}$, обученных на выборке $\{x\}_{i=1}^{N}$. Требуется агрегировать результаты их работы.
Правило отнесения объекта классу после агрегирования:

$$\widehat{C}(x) = \arg\min_{C_j \in \mathcal{C}} \sum_k w_k \delta(\widehat{C}_k(x), C_j),$$

где $\widehat{C}_k(x)$ – предсказание k-го классификатора для объекта x, $\widehat{C}(x)$ – предсказание ансамбля; $\delta(a, b)$ – символ Кронекера.

Некоторые традиционные агрегаторы:

- Plurality vote: $w_k = \frac{1}{K}, \forall k$;
- Simple weighted vote: $w_k = \frac{a_k}{\sum_l a_l}$, где a_k – доля правильно классифицированных объектов k-м классификатором;
- Re-scaled weighted vote. $w_k \propto a_k = \max\left\{0, 1 - \frac{Me_k}{N(M-1)}\right\}$, где e_k – количество ошибок допускаемых k-м классификатором.

Агрегирование классификаторов

Построение тела свидетельств для бинарного классификатора:

- $X = \{C_1, C_2\}$ – универсальное множество двух классов;
- $\mathcal{A} = \{\{C_1\}, \{C_2\}, X\}$– множество фокальных элементов;
- функции масс:

$m(\{C_1\}) = ap_{C_1}(t)$, $m(\{C_2\}) = ap_{C_2}(t)$, $m(X) = 1 - m(\{C_1\}) - m(\{C_2\})$, где a – точность классификации, $p_{C_i}(t)$ – вероятность отнесения образа t классу C_i.

БД/Алгоритмы классификации	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

Агрегирование рекомендаций

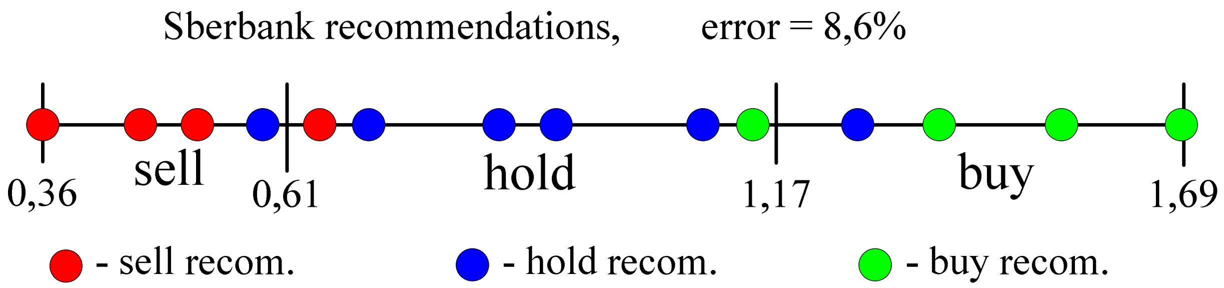
- множество рекомендаций финансовых аналитиков (7 российских банков и 3 аналитические компании) относительно акций российских компаний («голубые фишки») вида {"sell", "hold", "buy"} + целевая цена по акции к концу периода;
- данные по стоимости ценных бумаг за предыдущий период;
- история рекомендаций финансовых аналитиков.

Тело свидетельств j-го аналитика:

$X = [0, t_{\max}]$, $t \in X$, где $t = \dfrac{\text{целевая цена бумаги}}{\text{реальная цена бумаги на момент прогноза}}$;

$F_j = (\mathcal{A}_j, m_j)$, где $m_j(A)$ – относительное число рекомендаций,

$A \in \mathcal{A}_j = \{S_j, H_j, B_j\}$,

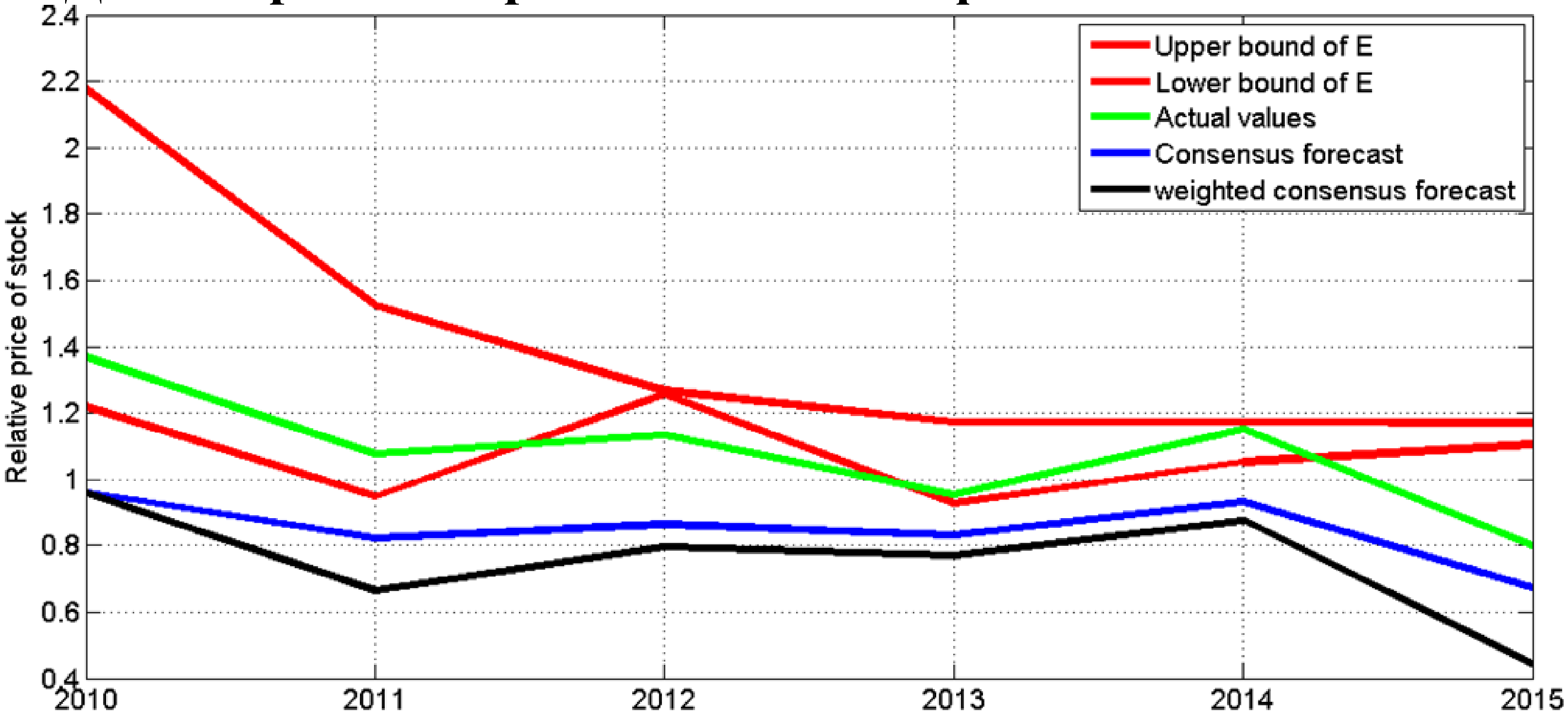


Агрегирование рекомендаций

Стратегии выбора аналитиков для агрегирования

1. Выбираются аналитики с наименьшим конфликтом.
2. Выбираются аналитики с наибольшей надежностью, не конфликтующие между собой.

Прогноз целевой цены акций Транснефти. Агрегирование – правило Демпстера с выбором наименее конфликтных аналитиков



Агрегирование рекомендаций

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	**0.19**3	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**

Remark. 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).

Список литературы

1. Bronevich A., Lepskiy A., Penikas H. The Application of Conflict Measure to Estimating Incoherence of Analyst's Forecasts about the Cost of Shares of Russian Companies// Procedia Computer Science, 2015, 55, pp.1113-1122.
2. Dempster A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping// The Ann. of Math. Stat., 1967, 38(2), 325–339.
3. Dubois D., Prade H. On the combination of evidence in various mathematical frameworks// In: Flamm, J., Luisi, T. (eds.) Reliability Data Collection and Analysis, Kluwer Acad. Publ., Dordrecht, 1992, 213–241.
4. Shafer G. A Mathematical Theory of Evidence. – Princeton, N.J.: Princeton University Press, 1976.
5. Lepskiy A., Suevalov A. Application of the Belief Function Theory to the Development of Trading Strategies// Procedia Computer Science, 2019, 162, 235–242.
6. Kutynina E., Lepskiy A. Aggregation of Forecasts and Recommendations of Financial Analysts in the Framework of Evidence Theory// in: Kacprzyk J. et al. (eds). Advances in Intelligent Systems and Computing, 2018, 642, Springer, Cham, 370–381.
7. Yager R. On the Dempster–Shafer framework and new combinations rules// Inform. Sc., 1987, 41, 93–137.

Спасибо за внимание!