

# Course aux avantages FNC

Enzo Poggio

11 avril 2018

## Résumé

La course est une méthode d'évaluation des systèmes entre eux. Elle permet d'avoir le détail des erreurs de chaque système. Ainsi, elle nous montre les lacunes chez chacun d'eux, et par ailleurs, les avantages que certains systèmes ont sur les autres. De plus, elle nous permet de formuler des hypothèses pour des améliorations ultérieures. La course, comme l'analyse des résultats, se passe seulement sur 80 % de l'ensemble de test.

## 1 La course.

### 1.1 Départ au point de rencontre.

Stance	unrelated	agree	disagree	discuss	Somme
Points	3560	406	1	2204	6171
Différence	105.5	1162	549	1346	3162.5
Pourcentage	97.12%	25.89%	0.18%	62.08%	66.11%

TABLE 1 – Résumé de l'avancement commun des participants (FNC score)

Le point de rencontre est l'avancement commun des participants. C'est-à-dire tous ce que les participants ont labélisés de la même manière, et qui est juste. C'est le minimum qu'obtient l'ensemble des systèmes combinés. Dans le tableau ci-dessus, on se réfère à plusieurs éléments. La ligne « points » indique le nombre d'élément bien labélisé par label. La ligne « différence » est la soustraction du nombre d'éléments total du label, moins ceux trouvés. Enfin, la ligne « pourcentage » est l'exactitude des éléments communs aux systèmes par

label. Par extension, il nous indique combien de pourcentage d'amélioration il nous reste à faire.

On peut alors initialiser un compteur pour chacun des participants :

	Points	Solat	Athene	UCLMR
Départ	6171	6171	6171	6171

TABLE 2 – Initailisation des compteurs pour la course.

## 1.2 Progression par label.

Ainsi dans cette section nous allons voir comment les différents systèmes se départagent sur chacun des labels. Nous allons ajouter une dimension que nous nommerons « avantage ». Un système a un avantage quand il n'a pas fait d'erreur là où un ou les deux autres systèmes en ont commis une. Nous parlerons de « désavantage commun » si les trois systèmes ont fait une erreur de classification. Nous appellerons « avantage fort » : un avantage sur deux systèmes. Et nous appellerons « avantage faible » : un avantage sur un seul système. Dans chacun des tableaux suivants nous allons commenter le nombre de points gagnés par chaque système. Chaque tableau est organisé dans l'ordre décroissant des nombres à gagner. Chaque ligne représente ce que chaque système a répondu. Si la réponse est verte c'est que le système a attribué le bon label. Si la réponse est noire c'est que le système a attribué un autre label *Related* incorrect (pour la classe *unrelated* les bonnes réponses seront toujours des réponses en noir). Si la réponse est rouge c'est que le système a classé les éléments en *unrelated* alors qu'ils étaient *Related* (ou vice-versa). Le nombre de points représente le nombre d'élément classé (sauf pour la classe *unrelated* où il faut appliquer un facteur 4).

## 1.2.1 Agree

### 1.2.1.1 Avantage : une erreur

n°	points	Solat	Athene	UCLMR
1	117	117	117	29.25 discuss
2	87	87	21.75 discuss	87
3	73	18.25 discuss	73	73
4	30	0 unrelated	30	30
5	12	12	3.0 disagree	12
6	4	4	4	0 unrelated
7	3	3	0 unrelated	3
8	2	0.5 disagree	2	2
9	1	1	1	0.25 disagree
	329	242.75	251.75	236.5

TABLE 3 – Tableau pour le label agree pour 1 erreurs dans les participants.

La plus grosse source d'erreur pour la classe agree est de les associer à des discuss. Les lignes 1, 2 et 3 en témoignent. Chaque système pêche sur cette distinction apparemment à plus ou moins grand échelle. Nous remarquons de manière mineur la présence de la classe disagree, qui apporte des erreurs peu significatives.

### 1.2.1.2 Avantage : 2 erreurs

n°	points	Solat	Athene	UCLMR
1	262	262	65.5 discuss	65.5 discuss
2	48	12.0 discuss	48	12.0 discuss
3	47	11.75 discuss	11.75 discuss	47
4	18	0 unrelated	18	4.5 discuss
5	13	13	3.25 disagree	3.25 discuss
6	12	3.0 discuss	3.0 disagree	12
7	8	8	0 unrelated	2.0 discuss
8	7	7	0 unrelated	0 unrelated
9	6	0 unrelated	6	0 unrelated
10	6	1.5 discuss	0 unrelated	6
11	5	0 unrelated	0 unrelated	5
12	4	0 unrelated	1.0 discuss	4
13	4	1.0 disagree	1.0 disagree	4
14	3	0 unrelated	0.75 disagree	3
15	3	0.75 discuss	3	0 unrelated
16	3	0.75 disagree	0.75 discuss	3
17	3	3	0.75 discuss	0 unrelated
18	2	2	0.5 discuss	0.5 disagree
19	1	0.25 disagree	1	0.25 discuss
20	1	1	0 unrelated	0.25 disagree
21	1	1	0.25 disagree	0.25 disagree
	457	328.0	164.5	172.5

TABLE 4 – Tableau pour le label agree pour 2 erreurs dans les participants.

Solat prend un fort avantage par rapport aux autres (ligne 1). Pour autant, chacun des systèmes est spécialisé pour retrouver des éléments labélisés agree parmi les discuss. Encore une fois, la classe disagree est présente, mais sur des erreurs qui ne rapportent que peu de points.

### 1.2.1.3 Désavantage commun : 3 erreurs

n°	points	Solat	Athene	UCLMR
1	236	59.0 discuss	59.0 discuss	59.0 discuss
2	48	0 unrelated	0 unrelated	0 unrelated
3	22	5.5 discuss	0 unrelated	5.5 discuss
4	18	0 unrelated	4.5 discuss	4.5 discuss
5	18	4.5 discuss	0 unrelated	0 unrelated
6	15	0 unrelated	0 unrelated	3.75 discuss
7	8	2.0 discuss	2.0 disagree	2.0 discuss
8	3	0 unrelated	0.75 disagree	0.75 discuss
9	3	0.75 discuss	0.75 discuss	0 unrelated
10	2	0.5 discuss	0.5 discuss	0.5 disagree
11	1	0 unrelated	0.25 discuss	0 unrelated
12	1	0 unrelated	0.25 disagree	0 unrelated
13	1	0.25 disagree	0.25 discuss	0.25 discuss
	376	72.5	68.25	76.25

TABLE 5 – Tableau pour le label agree pour 3 erreurs dans les participants.

Ici, nous avons la preuve que la distinction entre discuss et la classe agree est primordial pour améliorer les résultats. Plus de 300 points sont perdus dans ce type d’erreur de classification.

### 1.2.2 Disagree

#### 1.2.2.1 Avantage : une erreur

n°	points	Solat	Athene	UCLMR
1	7	1.75 agree	7	7
2	2	0.5 discuss	2	2
3	2	2	2	0.5 agree
4	1	1	0.25 discuss	1
5	1	1	1	0.25 discuss
	13	6.25	12.25	10.75

TABLE 6 – Tableau pour le label disagree pour 1 erreurs dans les participants.

Il n’y a que très peu de points où plusieurs systèmes s’accordent à labéliser disagree. Ceci montre bien que cette classe sous-représentée n’a pas de traits

significatifs suffisants.

#### 1.2.2.2 Avantage : 2 erreurs

n°	points	Solat	Athene	UCLMR
1	11	2.75 discuss	11	2.75 discuss
2	11	2.75 agree	11	2.75 discuss
3	7	1.75 discuss	1.75 agree	7
4	6	1.5 discuss	1.5 discuss	6
5	6	1.5 agree	6	1.5 agree
6	6	1.5 agree	1.5 agree	6
7	4	0 unrelated	4	1.0 discuss
8	4	1.0 discuss	4	1.0 agree
9	4	1.0 agree	1.0 discuss	4
10	3	0.75 agree	3	0 unrelated
11	2	0 unrelated	2	0.5 agree
12	2	2	0.5 agree	0.5 discuss
13	2	0.5 agree	0 unrelated	2
14	1	0 unrelated	0 unrelated	1
15	1	0 unrelated	1	0 unrelated
16	1	0.25 discuss	1	0 unrelated
17	1	1	0.25 discuss	0.25 discuss
	72	18.25	49.5	36.25

TABLE 7 – Tableau pour le label disagree pour 2 erreurs dans les participants.

Nous voyons aussi ici qu'il est rare que deux systèmes soient de concert même dans l'erreur.

### 1.2.2.3 Désavantage commun : 3 erreurs

n°	points	Solat	Athene	UCLMR
1	60	15.0 agree	15.0 discuss	15.0 discuss
2	59	14.75 discuss	14.75 discuss	14.75 discuss
3	46	11.5 discuss	11.5 agree	11.5 discuss
4	45	11.25 agree	11.25 agree	11.25 agree
5	38	0 unrelated	0 unrelated	0 unrelated
6	35	8.75 agree	8.75 agree	8.75 discuss
7	32	8.0 discuss	8.0 agree	8.0 agree
8	22	5.5 discuss	0 unrelated	5.5 discuss
9	17	4.25 discuss	4.25 discuss	4.25 agree
10	16	4.0 agree	4.0 discuss	4.0 agree
11	14	3.5 discuss	0 unrelated	0 unrelated
12	11	0 unrelated	2.75 discuss	0 unrelated
13	9	0 unrelated	2.25 discuss	2.25 discuss
14	7	1.75 discuss	1.75 discuss	0 unrelated
15	7	1.75 agree	0 unrelated	1.75 agree
16	6	1.5 agree	0 unrelated	0 unrelated
17	6	1.5 agree	0 unrelated	1.5 discuss
18	5	0 unrelated	0 unrelated	1.25 discuss
19	5	0 unrelated	1.25 agree	1.25 discuss
20	5	0 unrelated	1.25 agree	1.25 agree
21	4	0 unrelated	0 unrelated	1.0 agree
22	4	0 unrelated	1.0 agree	0 unrelated
23	4	1.0 discuss	1.0 agree	0 unrelated
24	3	0.75 discuss	0 unrelated	0.75 agree
25	2	0.5 agree	0.5 agree	0 unrelated
26	1	0 unrelated	0.25 discuss	0.25 agree
27	1	0.25 agree	0.25 discuss	0 unrelated
	464	95.5	89.75	94.25

TABLE 8 – Tableau pour le label disagree pour 3 erreurs dans les participants.

L'absence d'entrainement suffisant, et la prédominance des autres classes, fait que les points de la classe sont hors d'atteinte de tous les systèmes.

### 1.2.3 Discuss

#### 1.2.3.1 Avantage : une erreur

n°	points	Solat	Athene	UCLMR
1	334	83.5 agree	334	334
2	113	113	113	28.25 agree
3	93	93	0 unrelated	93
4	91	91	22.75 agree	91
5	34	34	34	0 unrelated
6	25	0 unrelated	25	25
7	13	13	3.25 disagree	13
8	11	11	11	2.75 disagree
9	2	0.5 disagree	2	2
	716	439.0	545.0	589.0

TABLE 9 – Tableau pour le label discuss pour 1 erreurs dans les participants.

De manière similaire, la classe discuss souffre du même mal que la classe agree. On voit que Solat labélise plus souvent agree que discuss. Le facteur de connivence entre le titre et le corps du texte est peut-être à revoir pour Solat.



### 1.2.3.2 Avantage : 2 erreurs

n°	points	Solat	Athene	UCLMR
1	83	83	20.75 agree	20.75 agree
2	78	19.5 agree	19.5 agree	78
3	76	19.0 agree	76	19.0 agree
4	30	30	0 unrelated	0 unrelated
5	17	0 unrelated	17	0 unrelated
6	15	0 unrelated	0 unrelated	15
7	14	3.5 agree	14	0 unrelated
8	13	13	0 unrelated	3.25 agree
9	10	0 unrelated	10	2.5 agree
10	10	2.5 agree	0 unrelated	10
11	8	0 unrelated	2.0 agree	8
12	8	2.0 agree	8	2.0 disagree
13	7	1.75 agree	1.75 disagree	7
14	6	6	1.5 agree	1.5 disagree
15	5	5	1.25 disagree	1.25 agree
16	3	0 unrelated	3	0.75 disagree
17	2	2	0.5 disagree	0 unrelated
18	1	0 unrelated	0.25 disagree	1
19	1	1	0.25 disagree	0.25 disagree
20	1	1	0.25 agree	0 unrelated
21	1	0.25 disagree	1	0.25 disagree
22	1	0.25 disagree	1	0.25 agree
23	1	0.25 disagree	0.25 agree	1
	391	190.0	178.25	171.75

TABLE 10 – Tableau pour le label discuss pour 2 erreurs dans les participants.

En somme les différentes erreurs de chaque participant ne participant pas à favoriser un système ici.

### 1.2.3.3 Désavantage commun : 3 erreurs

n°	points	Solat	Athene	UCLMR
1	100	25.0 agree	25.0 agree	25.0 agree
2	84	0 unrelated	0 unrelated	0 unrelated
3	19	4.75 agree	0 unrelated	0 unrelated
4	10	0 unrelated	0 unrelated	2.5 agree
5	8	0 unrelated	2.0 agree	2.0 agree
6	5	1.25 agree	0 unrelated	1.25 agree
7	5	1.25 agree	1.25 agree	1.25 disagree
8	4	1.0 agree	1.0 agree	0 unrelated
9	3	0 unrelated	0.75 agree	0 unrelated
10	1	0.25 agree	0.25 disagree	0 unrelated
	239	33.5	30.25	32.0

TABLE 11 – Tableau pour le label discuss pour 3 erreurs dans les participants.

### 1.2.4 Unrelated

#### 1.2.4.1 Avantage : une erreur

n°	points	Solat	Athene	UCLMR
1	36.75	36.75 unrelated	36.75 unrelated	0 discuss
2	13.25	0 discuss	13.25 unrelated	13.25 unrelated
3	7.25	7.25 unrelated	7.25 unrelated	0 agree
4	6.25	6.25 unrelated	0 discuss	6.25 unrelated
5	2.75	2.75 unrelated	0 agree	2.75 unrelated
6	2.0	0 agree	2.0 unrelated	2.0 unrelated
7	0.5	0.5 unrelated	0 disagree	0.5 unrelated
8	0.25	0.25 unrelated	0.25 unrelated	0 disagree
	69.0	53.75	59.5	24.75

TABLE 12 – Tableau pour le label unrelated pour 1 erreurs dans les participants.

Nous remarquons que UCLMr a plus de mal à classer les unrelated par rapport aux autres. Mais cela ne reste qu'un petit désavantage par rapport aux autres classes.

#### 1.2.4.2 Avantage : 2 erreurs

n°	points	Solat	Athene	UCLMR
1	15.5	0 discuss	15.5 unrelated	0 discuss
2	3.75	0 discuss	0 discuss	3.75 unrelated
3	2.0	0 agree	2.0 unrelated	0 discuss
4	1.75	1.75 unrelated	0 discuss	0 discuss
5	1.5	0 agree	1.5 unrelated	0 agree
6	0.5	0 discuss	0.5 unrelated	0 agree
7	0.25	0.25 unrelated	0 discuss	0 agree
8	0.25	0.25 unrelated	0 agree	0 agree
9	0.25	0 agree	0.25 unrelated	0 disagree
10	0.25	0 agree	0 discuss	0.25 unrelated
	26.0	2.25	19.75	4.0

TABLE 13 – Tableau pour le label unrelated pour 2 erreurs dans les participants.

Le biais vers les classes agree et discuss est bien souligné.

#### 1.2.4.3 Désavantage commun : 3 erreurs

n°	points	Solat	Athene	UCLMR
1	9.25	0 discuss	0 discuss	0 discuss
2	0.5	0 discuss	0 agree	0 agree
3	0.5	0 agree	0 agree	0 agree
4	0.25	0 discuss	0 disagree	0 agree
	10.5	0	0	0

TABLE 14 – Tableau pour le label unrelated pour 3 erreurs dans les participants.

Au final, seuls peu de points sont perdus pour la classe unrelated par rapport aux classes related.

### 1.3 Fin de courses et analyse des arrivées.

Stance	points	Solat	Athene	UCLMR
Départ	6171	6171	6171	6171
agree				
1 erreur	329	242.75	251.75	236.5
2 erreurs	457	328.0	164.5	172.5
3 erreurs	376	72.5	68.25	76.25
disagree				
1 erreur	13	6.25	12.25	10.75
2 erreurs	72	18.25	49.5	36.25
3 erreurs	464	95.5	89.75	94.25
discuss				
1 erreur	716	439.0	545.0	589.0
2 erreurs	391	190.0	178.25	171.75
3 erreurs	239	33.5	30.25	32.0
unrelated				
1 erreur	69.0	53.75	59.5	24.75
2 erreurs	26.0	2.25	19.75	4.0
3 erreurs	10.5	0	0	0
Somme	9333.5	7652.75	7639.75	7619.0

TABLE 15 – Tableau récapitulatif des sommes des avantages de chaque système en fonction des labels et du nombres d’erreurs.

En faisant la somme de tous les avantages à une erreur, ceux à deux et le départ<sup>1</sup> on obtient 8244 points. C’est-à-dire un pourcentage FNC de 88.32%. Donc avec la combinaison adéquate des modèles actuels nous pouvons beaucoup améliorer le score général.

De plus, si nous prenons en compte les biais inhérents aux données comme nous l’avons vu précédemment il est peut-être possible d’améliorer d’avantage ce résultat. À savoir qu’il faut mieux distinguer la classe discuss de la classe agree. Se concentrer sur la classe disagree pourrait être rentable<sup>2</sup> seulement si nous trouvons des traits plus distinctif pour cette classe. Sinon c’est une classe qui ne vaut pas le coup. Améliorer les résultats pour la classe unrelated paraît difficile. Les résultats pour cette classe sont beaucoup trop bons. Essayer de

1. C’est-à-dire toutes les entrées où au moins un système ne s’est pas trompé.

2. Rentabilité relative à la proportion de données dans les corpus.

les améliorer serait futile.